



HAL
open science

Codage vidéo flexible par association d'un décodeur intelligent et d'un encodeur basé optimisation débit-distorsion

Jean-Marc Thiesse

► **To cite this version:**

Jean-Marc Thiesse. Codage vidéo flexible par association d'un décodeur intelligent et d'un encodeur basé optimisation débit-distorsion. Traitement du signal et de l'image [eess.SP]. Université Nice Sophia Antipolis, 2012. Français. NNT: . tel-00719058

HAL Id: tel-00719058

<https://theses.hal.science/tel-00719058>

Submitted on 19 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE DE NICE-SOPHIA ANTIPOLIS
ECOLE DOCTORALE STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA
COMMUNICATION

T H E S E

pour l'obtention du grade de

Docteur en Sciences

de l'Université de Nice-Sophia Antipolis

Mention : Automatique, Traitement du Signal et des Images

présentée par

Jean-Marc THIESSE

**Codage vidéo flexible par association d'un décodeur
intelligent et d'un encodeur basé optimisation
débit-distorsion**

Thèse dirigée par

Marc ANTONINI, Directeur de Recherche, CNRS

Equipe d'accueil : CReATIVE, laboratoire I3S, UNSA-CNRS

et

Joël JUNG, Ingénieur de Recherche, Orange Labs

Equipe d'accueil : CVA, Laboratoire OPERA, Orange Labs - France Télécom R&D

soutenue le 18 avril 2012

Jury :

James FOWLER	Mississippi State University, USA	Rapporteur
Pascal FROSSARD	EPFL, Suisse	Rapporteur
Philippe SALEMBIER	University of Catalonia, Espagne	Rapporteur
Marc ANTONINI	CNRS-Université de Nice-Sophia Antipolis	Directeur de thèse
Joël JUNG	Orange Labs, France	Encadrant de thèse
Frédéric DUFAUX	CNRS-Télécom Paris Tech, France	Examinateur
Pier Luigi DRAGOTTI	Imperial College London, UK	Examinateur
Jérôme VIERON	ATEME, France	Examinateur

A Albert.
A Nicolas.

Remerciements

Je souhaite en premier lieu exprimer toute ma reconnaissance à Joël Jung, ingénieur de recherche chez Orange Labs, pour son encadrement au cours de ces trois années de thèse. Les nombreuses heures à partager son savoir et son expérience tant professionnelles que personnelles ont grandement participé au succès de cette thèse.

Je tiens également à exprimer toute ma reconnaissance à Marc Antonini qui a honoré la fonction de directeur de thèse depuis le laboratoire I3S à Sophia-Antipolis. Ses conseils avisés et sa capacité à guider mes recherches ont fortement participé à la réussite de ces travaux.

Je tiens à remercier chaleureusement James Fowler, Philippe Salembier et Pascal Frossard d'avoir assumé la charge de rapporteur. Je veux de plus leur manifester ma très sincère reconnaissance quant au sérieux et à l'esprit critique dont ils ont fait preuve dans cet exercice. J'exprime également toute ma gratitude envers Pier-Luigi Dragotti, Frédéric Dufaux et Jérôme Viéron pour avoir accepté d'évaluer mon travail.

Je remercie maintenant l'ensemble des personnes cotoyées au cours de cette thèse. En particulier, les trois managers d'Orange Labs qui m'ont successivement accueilli dans leur équipe, Jean-Claude Schmitt, Ludovic Noblet et Didier Gaubil. Par ordre chronologique, j'adresse mes pensées aux membres de l'équipe Service des Images Animées à Issy-les-Moulineaux ainsi qu'à ceux de l'équipe Codage Vidéo Avancée à Rennes, les trois Patrick, Erwan, Olivier, Stéphane, Francis, Guillaume, Thomas et Chantal pour sa gentillesse puis Stéphane, Gordon, Félix, Patrick, Gilles, Maryline, Christophe et Joël. Une pensée particulière à mes camarades devenu docteurs ou futur docteurs, les deux Mathieu, Thomas, ShaSha, Younes et bien-sûr Elie. Je remercie aussi les stagiaires avec qui j'ai eu l'occasion de travailler, Thomas, Laurent et plus particulièrement Julien dont le travail a servi de bases à la seconde étude de ce manuscrit. Mêmes si nos échanges ont été davantage limités du fait de la distance, je tiens à saluer mes camarades de l'I3S, Khaled, Aymen, Leonardo, et Anis. J'adresse un remerciement particulier à Régine pour les nombreux échanges par mail concernant le volet administratif.

Je salue finalement mes nouveaux collègues au sein d'ATEME pour leur accueil et leur soutien dans mes derniers mois de rédaction et de préparation de la soutenance. En particulier, Jérôme Viéron, membre du jury aux questions pointilleuses, mon collègue de

Remerciements

bureau Mathieu ainsi qu'Etienne qui a fait le déplacement à Sophia-Antipolis pour ma soutenance.

Un tel travail ne serait pas possible sans le soutien des proches, famille et amis. Je remercie ainsi les amis Toulousains sédentaires, et ceux expatriés à Paris ou ailleurs, les membres de la Colocachan, de la mayliste et ceux de la GGGold pour les succès du lundi soir. Mes pensées vont vers ma mère, mon père que je remercie pour les commentaires avisés sur le manuscrit, et mon frère, tous trois présents à mes côtés depuis le début. Enfin, un grand merci à Mathilde pour son appui au quotidien pendant ces trois années, ses relectures et sa présence.

Table des matières

Remerciements	iii
Table des matières	v
Liste des figures	xi
Liste des tableaux	xvii
Glossaire	xxi
Introduction	1
1 Le codage vidéo, les normes H.264/AVC et HEVC	7
1.1 Codage vidéo au travers de la norme H.264/AVC	7
1.1.1 Contexte et synthèse du schéma d'encodage	7
1.1.2 Calcul du résiduel	9
1.1.3 Représentation du résiduel	12
1.1.4 Codage du résiduel	13
1.2 Schéma général de la future norme HEVC	14
1.2.1 Contexte	14
1.2.2 Description détaillée	15
1.2.2.1 Structure de codage	15
1.2.2.2 Prédiction Intra	17
1.2.2.3 Prédiction Inter	19
1.2.2.4 Autres outils	20
1.2.3 Evolution des standards de MPEG-2 à HEVC	21
1.3 Schémas alternatifs	22
1.3.1 Codecs divers	23
1.3.2 Autres voies d'amélioration majeures	24
1.4 Bilan	25

2	Analyse comportementale des schémas de codage vidéo modernes	27
2.1	Analyse de l'encodeur H.264/AVC	27
2.1.1	Analyse globale	27
2.1.2	Les informations de signalisation	32
2.1.3	L'information de mouvement	34
2.2	De MPEG-2 à HEVC	37
2.2.1	MPEG-2, H.264/AVC, HEVC : évaluation des performances . . .	37
2.2.2	Vers une limite des approches conventionnelles ?	43
2.3	Bilan	48
I	Approches conventionnelles pour les schémas de codage vidéo émergents	51
3	Signalisation, codage du mouvement et tatouage : un état de l'art	53
3.1	Etat de l'art sur la signalisation des choix de codage	53
3.2	Etat de l'art sur le codage du mouvement	55
3.2.1	Prédiction des vecteurs de mouvement	55
3.2.2	Codage du résiduel de mouvement	57
3.2.3	Autres approches pour le codage du mouvement	59
3.3	Un bref état de l'art sur le tatouage en vidéo	60
3.3.1	Tatouage	60
3.3.2	Tatouage et compression	62
3.4	Bilan	63
4	Tatouage pour le codage vidéo	65
4.1	Choix préliminaires	65
4.1.1	Technique de tatouage	66
4.1.2	Informations à masquer	67
4.2	Description de la méthode proposée	68
4.2.1	Schéma global	68
4.2.2	Modification des coefficients transformés et quantifiés	70
4.2.2.1	Optimisation débit-distorsion classique	70
4.2.2.2	Optimisation de Pareto	72
4.2.3	Amélioration de la méthode	74
4.2.4	Complexité de la méthode	74
4.3	Résultats expérimentaux	75
4.3.1	Masquage d'informations de mouvement	75
4.3.1.1	Résultats objectifs	75
4.3.1.2	Résultats subjectifs	86
4.3.2	Masquage d'informations Intra	89
4.3.2.1	Résultats objectifs	89

4.3.2.2	Résultats subjectifs	92
4.4	Bilan	94
5	Codage des vecteurs de mouvement par redistribution des résiduels basée sur un pronostic du mouvement	97
5.1	Description de la méthode proposée	98
5.1.1	Schéma global	98
5.1.2	Etape 1 : Création de l'information de mouvement	99
5.1.3	Etape 2 : Affinage de l'information de mouvement	103
5.1.3.1	Théorie de Parzen	103
5.1.3.2	Application à notre méthode	105
5.1.4	Etape 3 : Redistribution des résiduels de vecteurs de mouvement .	106
5.2	Discussion sur l'intégration dans un encodeur HEVC	110
5.2.1	Interaction avec la méthode de compétition de prédicteurs de mouvement (MV-Comp)	110
5.2.1.1	Comparaison des deux méthodes	110
5.2.1.2	Proposition de combinaison des méthodes	111
5.2.2	Interaction avec le mode Merge	112
5.2.3	Interaction avec le codeur entropique CABAC	113
5.3	Résultats expérimentaux	114
5.3.1	Evaluation de notre méthode	114
5.3.1.1	Configuration de tests	114
5.3.1.2	Analyse préliminaire	115
5.3.1.3	Performances de la méthode	118
5.3.1.4	Analyse	119
5.3.2	Evaluation de notre méthode et l'outil MV-Comp	123
5.4	Bilan	125
II	Approches en rupture basées sur l'exploitation d'un décodeur intelligent	127
6	Transfert d'intelligence au décodeur et échantillonnage compressé : un état de l'art	129
6.1	Etat de l'art sur l'exploitation d'un décodeur intelligent	130
6.1.1	Codage vidéo distribué	130
6.1.2	Applications au schéma de codage classique	132
6.1.2.1	Approches par « Template Matching »	132
6.1.2.2	Traitement des informations de compétition	134
6.1.2.3	Applications en codage multivues	135
6.2	Echantillonnage compressé : un bref état de l'art	136
6.2.1	Théorie	136

TABLE DES MATIÈRES

6.2.2	Echantillonnage compressé et vidéo	140
6.2.2.1	Echantillonnage compressé appliqué à un signal vidéo . .	141
6.2.2.2	Echantillonnage compressé intégré dans un codeur vidéo	143
6.3	Bilan	144
7	Approche décodeur intelligent avec application au codage Intra	147
7.1	Description de l'approche proposée	147
7.1.1	Motivation	147
7.1.2	Schéma global	149
7.1.3	Sélection de la partition causale	153
7.1.4	Sélection des paramètres optimaux	156
7.1.5	Traitement de la partition courante	158
7.2	Application pour l'héritage de modes Intra	159
7.2.1	Description de la méthode	159
7.2.2	Résultats expérimentaux	160
7.2.2.1	Configuration de tests	160
7.2.2.2	Evaluation sur des séquences 2D	161
7.2.2.3	Evaluation sur des séquences de cartes de profondeur . .	168
7.3	Perspectives	171
7.3.1	Perspectives générales	171
7.3.2	Perspectives dédiées à l'application aux modes Intra	172
7.4	Bilan	173
8	Approche décodeur intelligent appliquée à l'échantillonnage compressé pour le codage vidéo	175
8.1	Description de la méthode proposée	175
8.1.1	Schéma global	175
8.1.2	Constitution de la liste des paramètres en compétition	179
8.1.3	Détails de la méthode	182
8.1.3.1	Sélection de la partition causale	182
8.1.3.2	Sélection des paramètres optimaux	184
8.1.3.3	Codage et reconstruction de la partition courante	186
8.2	Résultats expérimentaux	188
8.2.1	Expérimentation théorique	188
8.2.1.1	Détails de l'implémentation	188
8.2.1.2	Résultats	189
8.2.2	Evaluation dans un codeur vidéo moderne (HEVC)	199
8.2.2.1	Détails de l'implémentation	199
8.2.2.2	Résultats	201
8.3	Bilan	204

Conclusion et perspectives	207
Annexe	213
A Hybrid-1D : un nouveau mode de prédiction	215
A.1 Etat de l'art	215
A.2 Méthode proposée : Hybrid 1D	216
A.2.1 Intra 1D + Skip 1D	216
A.2.1.1 Partitionnement 1D	216
A.2.1.2 Ordres de parcours	217
A.2.1.3 Prédiction	217
A.2.1.4 Transformée 1D	218
A.2.2 Inter 1D	218
A.2.3 Signalisation du mode proposé	219
A.2.4 Critère de compétition	220
A.3 Résultats expérimentaux	221
A.3.1 Configuration de test	221
A.3.2 Gains sur le débit	221
A.3.3 Analyse des résultats	222
A.4 Bilan	224
Bibliographie	225
Publications de l'auteur	239

Liste des figures

1.1	Vue d'ensemble du schéma de codage du standard H.264/AVC.	8
1.2	Prédiction Intra 16×16.	10
1.3	Structure de Gop IBBBP hiérarchique.	11
1.4	Parcours zig-zag d'un bloc de coefficients.	13
1.5	Planning des activités de standardisation HEVC.	15
1.6	Exemple de partitionnement en PU(bleu) et TU(rouge) d'une LCU 64×64.	17
1.7	Liste des prédicteurs Intra de HEVC.	17
1.8	Représentation du prédicteur Planar.	18
1.9	Interpolation utilisée pour la prédiction de la luminance en Inter dans HEVC.	19
1.10	Représentation des PU d'origines des vecteurs de mouvement prédicteurs candidats parmi le voisinage spatial de la PU courante.	20
2.1	Représentations visuelles d'un encodage d'une image extraite de la séquence Crowdrun 720p : image reconstruite (a), coût de chaque bloc (b) et répartition des modes de codage (Rouge : Intra, Vert : Inter, Bleu : Skip) (c).	29
2.2	Pourcentages moyens du débit des principales données de codage par rapport au flux complet au sein des images Intra à différents débits.	31
2.3	Pourcentages moyens du débit des principales données de codage par rapport au flux complet au sein des images Inter P à différents débits.	31
2.4	Pourcentages moyens du débit des principales données de codage par rapport au flux complet au sein des images Inter B à différents débits.	32
2.5	Pourcentages moyens du débit des différentes informations de signalisation par rapport au flux complet au sein des images Intra à différents débits.	33
2.6	Classification de l'origine des vecteurs de mouvement selon le type de leur bloc d'origine et la distance en quart de pixel à un vecteur déjà codé dans le voisinage (en pourcentages).	35
2.7	Représentation des champs de mouvements de deux images successives de la séquence Runners 720p codée avec H.264/AVC à QP=22.	37

2.8	Courbes débit-distorsion des trois standards (MPEG-2, H.264/AVC, HEVC) pour la séquence Crowdrun 720p (a), images tirées de la séquences encodée à 6.7Mbits/s avec H.264/AVC (b) et HEVC (c).	40
2.9	Courbes débit-distorsion des trois standards (MPEG-2, H.264/AVC, HEVC) pour la séquence Soccer 720p (a), images tirées de la séquences encodée à 4Mbits/s avec H.264/AVC (b) et HEVC (c).	41
2.10	Courbes débit-distorsion des trois standards (MPEG-2, H.264/AVC, HEVC) pour les séquences Victoria (a), Runners (b) et Stockholm (c).	42
2.11	Résiduels d'une image B de la séquences Crowdrun 720p codée avec les encodeurs MPEG-2 (a) et H.264/AVC (b) pour un même niveau de qualité (30dB).	44
2.12	Pourcentages moyens du débit des principales données de codage de l'encodeur HEVC par rapport au flux complet au sein des images Intra à différents débits.	45
2.13	Pourcentages moyens du débit des principales données de codage de l'encodeur HEVC par rapport au flux complet au sein des images Inter à différents débits.	45
2.14	Pourcentage de gains sur le débit à différents niveaux de débit pour plusieurs nombre de prédicteurs Intra en compétition (9, 18 et 35) par rapport à un prédicteur unique (testés dans HEVC).	47
2.15	Pourcentage de gains sur le débit à différents niveaux de débit pour plusieurs nombre d'images de référence en compétition (2 et 4) par rapport à une référence unique (testés dans HEVC).	48
3.1	Problématique du tatouage.	61
3.2	Schéma de compression basé sur le tatouage (extrait de [Campisi <i>et al.</i> 02]). 63	63
4.1	Valeurs des multiplicateurs de Lagrange déterminées pour chaque QP, pour les modes Inter 16×16 à 8×8 dans le cas d'une modification de la chrominance (a) et de la luminance (b).	71
4.2	Frontière de Pareto pour un ensemble donné de couples débit-distorsion. 73	73
4.3	Pourcentage d'indices MV-Comp transmis ou non-transmis dû au test d'égalité et au masquage d'indice pour les deux schémas : masquage dans la chroma (a) et masquage à la fois dans la chroma et la luma (b), pour trois valeurs de QP.	77
4.4	Pourcentage moyen de réduction de la proportion de l'indice dans le débit de l'information de mouvement comparé à la référence pour l'ensemble des séquences proposés. Masquage dans la chroma (rouge). Masquage à la fois dans la chroma et la luma (bleu).	80
4.5	Echelle d'évaluation de la qualité.	86
4.6	Evaluation subjective des séquences à haut (a) et moyen (b) débit. Comparaison de la référence avec notre méthode.	87

4.7	Séquence <i>Bbc_news CIF</i> : (a) compressée avec le codeur de référence; (b) avec le masquage dans la chroma; (c) différence entre (a) et (b); (d) zoom de (c). Test réalisé à haut débit.	88
4.8	Pourcentage d'indices MPM transmis ou masqués dans les coefficients AC ou DC, pour trois valeurs de QP.	90
4.9	Schéma de masquage en YUV 4 :2 :0 (à droite) et en YUV 4 :4 :4 (à gauche).	91
4.10	Pourcentage d'indices masqués en YUV 4 :2 :0 et YUV 4 :4 :4 selon le QP.	92
4.11	Evaluation subjective des séquences à haut (a) et moyen (b) débit. Comparaison de la référence avec notre méthode.	93
4.12	Séquence <i>RaceHorses WVGA</i> : (a) compressée avec le codeur de référence; (b) avec le masquage dans la chroma; (c) différence entre (a) et (b); (d) zoom de (c). Test réalisé à haut débit.	94
5.1	Vue globale de la méthode proposée.	99
5.2	Exemple de distribution des résiduels observés pour un bloc donné.	101
5.3	Représentation du vecteur translaté.	102
5.4	Représentation du vecteur colocalisé déplacé.	102
5.5	Exemple de densités de probabilité estimées par noyau pour différentes valeurs du paramètre de lissage (h_1) et différents nombres de mesures (n).	104
5.6	Exemple type de mise en œuvre des deux phases du processus d'association.	109
5.7	Sélection du mode Merge pour la séquence <i>ParkScene</i> tirée de [Marpe <i>et al.</i> 10]. Sur la figure de droite, les bordures blanches délimitent l'ensemble des partitions codées avec la même information de mouvement.	112
5.8	Classification de l'origine des vecteurs résiduels codés (en pourcentage) selon le niveau de quantification. Evaluation moyenne pour l'ensemble des séquences du CfP (a); pour la séquence <i>BasketBallDrill WVGA</i> (b); et la séquence <i>BQSquare QWVGA</i> (c).	121
5.9	Evolution du gain de codage des vecteurs résiduels (en bits) en fonction du déplacement moyen au quart de pixel réalisé lors de l'étape 3 (évalué par la norme 2). Résultats obtenus avec les séquences <i>PartyScene</i> (bleu) et <i>BasketBallDrill</i> (rose) WVGA pour le QP=32.	123
6.1	Schéma de codage vidéo distribué.	131
6.2	Exemple de représentation de l'estimation de mouvement au niveau du décodeur.	134
6.3	Schéma global de l'échantillonnage compressé.	138
6.4	Exemple de contenu adapté pour la méthode [Zhang <i>et al.</i> 08].	143
7.1	Schéma de codage d'une partition courante de façon classique (à gauche) ou selon notre proposition (à droite).	150
7.2	Représentation de l'approche <i>décodeur intelligent</i>	151

7.3	Partitions causales voisines spatialement (en hachuré) de la partition courante.	155
7.4	Partition causale pointée par le vecteur médian.	155
7.5	Modification de la signalisation HEVC intégrant le mode Intra DI proposé.	160
7.6	Pourcentage de sélection des différents modes de codage selon le QP calculés en moyenne pour l'ensemble des images Inter des séquences testées. Statistiques obtenues avec le codeur de référence (a), et le codeur modifié avec notre nouveau mode en moyenne sur l'ensemble des séquences (b); pour la séquence <i>BasketBallDrive</i> 1080p (c); et la séquence <i>BasketBallPass</i> qwvga (d).	165
7.7	Pourcentages de sélection des différents modes de codage selon le type d'image (Intra ou Inter) et de partition ($N \times N$ ou $2N \times 2N$). Moyenne calculée à partir des résultats des quatre QP.	167
7.8	Pourcentages de sélection des différents modes de codage en fonction du nombre de prédicteurs en compétition dans la méthode proposée. Moyenne calculée à partir des résultats des quatre QP.	169
7.9	Image compressée avec la méthode proposée au QP=32 tirée de la séquence de carte de profondeur Kendo 1024×768	170
7.10	Exemple d'estimation de mouvement dans le domaine des modes Intra.	172
8.1	Schéma simplifié de la méthode d'échantillonnage compressé combinée avec l'approche décodeur intelligent.	178
8.2	Exemples, pour un bloc donné (a) transformée par une ondelette avec 3 niveaux de décomposition (b), du premier et second masques (c et d respectivement) obtenus pour l'opérateur ondelette avec un échantillonnage de 50% (la zone blanche correspond aux valeurs conservées).	180
8.3	Exemple d'estimation dans les domaines transformés par DCT et ondelettes.	183
8.4	Evolution du gain débit-distorsion (%) de la méthode comparée à HEVC selon le nombre d'itérations pour un nombre d'itérations fixé (courbe bleue) et trois solutions hybrides. Tests avec l'implémentation HEVC pour la séquence <i>RaceHorses</i> qwvga.	185
8.5	Schéma détaillé de la méthode d'échantillonnage compressé combinée avec l'approche décodeur intelligent.	186
8.6	Variation du paramètre de sous-échantillonnage.	190
8.7	Résultat complet de la méthode comparée aux méthodes de l'état de l'art en échantillonnage compressé. Evolution du PSNR en fonction du pas d'échantillonnage pour trois séquences Cif testées : <i>Susie</i> (a), <i>Foreman</i> (b) et <i>Soccer</i> (c).	193

8.8	Seconde image de la séquence <i>Susie cif</i> originale (a) et encodée avec la méthode (b), répartition des opérateurs Φ_i optimaux DCT (en noir) et ondelette (en blanc) (c) et des échantillonnages optimaux S_i des plus forts échantillonnages (en noir) aux plus faibles (en blanc) (d).	195
8.9	Résultat de la méthode dans un schéma incluant une quantification des mesures. Comparaison à une méthode de l'état de l'art en échantillonnage compressé (BCS-PL) et à une méthode de codage vidéo classique (MC-BPC). Evolution du PSNR en fonction du débit pour trois séquences Cif testées : <i>Susie</i> (a), <i>Foreman</i> (b) et <i>Soccer</i> (c).	197
8.10	Répartition des opérateurs Φ_i optimaux avec la DCT en noir et l'ondelette en blanc (a) et des échantillonnages optimaux S_i avec les plus forts échantillonnages en noir et les plus faibles échantillonnages en blanc (b) pour la seconde image de la séquence <i>Susie cif</i>	198
8.11	Mode de codage basé sur l'échantillonnage compressé.	200
A.1	Exemple de partitionnements obtenus à partir d'une partition initiale p_1	217
A.2	Représentation de la mise en correspondance de partitions 1D.	219
A.3	Signalisation Hybrid-1D.	220
A.4	Exemples de distributions des sous-modes proposés.	224

Liste des tableaux

1.1	Principaux outils des standards MPEG-2, H.264/AVC et HEVC.	22
2.1	Proportion de l'indice de compétition de prédicteurs de mouvement dans l'information de mouvement et dans le flux total (en pourcentage).	34
2.2	Pourcentages de gain moyen sur le débit calculés avec la métrique de Bjontegaard pour les débits moyens entre les standards H.264/AVC et MPEG-2 ainsi qu'entre H.264/AVC et HEVC.	39
4.1	Pourcentage de réduction du débit moyen pour la méthode de masquage dans la composante de chroma. Modifications sélectionnées à l'aide de l'optimisation débit-distorsion classique.	78
4.2	Pourcentage de réduction du débit moyen pour la méthode de masquage à la fois dans les composantes de chroma et de luma. Modifications sélectionnées à l'aide de l'optimisation débit-distorsion classique.	79
4.3	Pourcentage de réduction du débit moyen dans une configuration IBBP pour les deux masquages, l'optimisation débit-distorsion classique et les modes Inter 16×16 à 8×8	81
4.4	Pourcentage de réduction du débit moyen pour les deux méthodes d'optimisation pour l'ensemble de QP du Test 1 et les modes Inter 16×16 à 8×8	81
4.5	Temps de calcul moyen (en millisecondes) sur un bloc pour les deux méthodes d'optimisation à différentes valeurs du pas de quantification (QP).	82
4.6	Position de la marque (en pourcentage) pour le masquage dans la chroma uniquement, et à la fois dans la chroma et la luma, pour 3 valeurs de QP.	83
4.7	Pourcentage de sélection de l'amplitude de la modification m_j relativement à différentes valeurs de coefficients résiduels originaux a_n pour le masquage dans la chroma uniquement, et à la fois dans la chroma et la luma.	84

4.8	Δ -PSNR des composantes de chroma U et V (en dB) ainsi que le pourcentage de gain sur le débit moyen pour le schéma proposé comparé à la méthode de référence. Masquage dans la chroma uniquement pour les modes Inter 16×16 à 8×8	85
4.9	Pourcentage de gains sur le débit pour les deux ensembles de QPs.	91
5.1	Résumé de la comparaison MV-Comp/MV-FMap.	111
5.2	Comparaison de différentes configurations de notre méthode pour les différentes résolutions des séquences du CfP HEVC. Pourcentage de gain moyen sur le débit.	116
5.3	Pourcentage de gains sur le débit de notre méthode MV-FMap dans la configuration optimale (Test 6) pour trois catégories de débits cibles. Application à l'ensemble des séquences du CfP HEVC et à 4 séquences additionnelles.	119
5.4	Pourcentage de gains sur le débit des méthodes MV-Comp et MV-FMap ainsi que de deux combinaisons des méthodes. Tests réalisés à moyens débits.	124
7.1	Pourcentages de réduction de débit maximums de notre méthode Intra DI pour quatre configurations non-décodables à moyen débit.	162
7.2	Pourcentages de réduction de débit moyen de notre méthode d'héritage des modes Intra basée décodeur intelligent pour deux configurations différentes à moyen et bas débit.	163
7.3	Pourcentages de réduction de débit moyen de notre méthode décodeur intelligent appliquée au mode Intra pour différents nombres de prédicteurs en compétition (4, 17 et 34). Test réalisé pour un ensemble de QP à moyen débit.	167
7.4	Pourcentages de réduction du débit moyen de notre méthode appliquée aux séquences de cartes de profondeur et textures associées.	170
8.1	Evaluation de la mise en compétition des opérateurs d'acquisition DCT et Ondelette pour nos deux propositions de masque ainsi que l'estimation de mouvement réalisée dans le domaine transformé (EM*). Tests réalisés avec $\nu = 5$	191
8.2	Pourcentages de réduction de débit maximums de notre méthode basée sur l'échantillonnage compressé et l'approche décodeur intelligent pour trois configurations non-décodables.	202
8.3	Pourcentages de réduction du débit de notre méthode (DI-CS) et des méthodes de l'état de l'art intégrées dans HEVC.	203
A.1	Pourcentages de gains sur le débit pour un encodage avec un GOP de 12 sur les QP 22 à 37 des trois méthodes <i>Intra 1D</i> , <i>Intra 1D+Skip 1D</i> et <i>Hybrid 1D</i> par rapport à la référence H.264/AVC.	222

A.2	Pourcentages de sélection moyens des modes du standard H.264/AVC et du mode proposé pour quatre QP sur l'ensemble des séquences testées.	222
A.3	Pourcentages de sélection de chaque sous-mode proposé dans les schémas <i>Intra 1D+Skip 1D</i> and <i>Hybrid 1D</i> pour quatre QP en moyenne sur les séquences testées.	223

Glossaire

AVC	<i>Advanced Video Coding</i>
BMA	<i>Block Matching Algorithm</i>
CABAC	<i>Context Adaptive Binary Arithmetic Coding</i>
CAVLC	<i>Context Adaptive Variable Length Coding</i>
CfP	<i>Call for proposal</i>
CIF	<i>Common Intermediate Format</i> - format de résolution 352×288 pixels
CS	<i>Compressed Sensing</i> - Echantillonnage Compressé
CU	<i>Coding Unit</i> - Unité de codage de la future norme HEVC
dB	<i>Décibel</i>
DCT	<i>Discrete Cosine Transform</i>
EM	<i>Estimation de mouvement</i>
GOP	<i>Group of Pictures</i>
HD	<i>High Definition</i>
HEVC	<i>High Efficiency Video Coding</i>
HM	<i>High Model</i> - Logiciel de référence associé à la future norme HEVC
ITU	<i>International Telecommunications Union</i>
JCT-VC	<i>Joint Collaborative Team on Video Coding</i>
KTA	<i>Key Technical Area</i>
MERGE	Outil utilisé pour le codage du mouvement dans la future norme HEVC
MPM	<i>Most Probable Mode</i>
MV-Comp	<i>Motion Vector predictor Competition</i>
PSNR	<i>Peak Signal to Noise Ratio</i>
PU	<i>Prediction Unit</i> - Unité de prédiction de la future norme HEVC
QP	<i>Quantization Parameter</i> - Pas de quantification
QWVGA	<i>Quarter WVGA</i> - format de résolution 416×240 pixels
SAD	<i>Sum of Absolute Differences</i>

SSD	<i>Sum of Squared Differences</i>
TU	<i>Transform Unit</i> - Unité de transformée de la future norme HEVC
WVGA	<i>Wide VGA</i> - format de résolution 832×480 pixels

Introduction

Contexte

Le développement de la vidéo numérique est croissant depuis deux décennies et plus particulièrement depuis l'année 1995 qui a vu l'arrivée des premières chaînes de télévision numérique, par câble et satellite, ainsi que l'avènement du DVD. De par ses atouts, le numérique a ainsi pris le pas sur l'analogique sur l'ensemble de la chaîne de traitement du signal vidéo : capture, stockage, transmission et visionnage. Il offre en effet davantage de fonctionnalités, d'applications et de services malgré une arrivée accompagnée d'une dégradation de la qualité de l'image. Bien que certaines spécificités héritées de l'analogique demeurent présentes, le numérique est actuellement implanté massivement dans notre quotidien au travers, notamment des écrans plats et du support de stockage DVD, ainsi que son successeur le Blue-Ray. Multiplication des offres de contenus et perfectionnement de la qualité visuelle sont ainsi les deux axes de progrès ouverts par la vidéo numérique. Nombre de défis théoriques et technologiques ont été relevés au cours des précédentes années mais l'évolution incessante de la demande entretient la nécessité de progression.

Ainsi, concernant l'avenir de la vidéo, plusieurs analyses récentes, dont une étude menée par le constructeur Cisco en 2011, font état d'une croissance encore plus élevée du trafic vidéo pour les années à venir. Au niveau mondial, le trafic Internet total serait ainsi multiplié par quatre d'ici 2015 et la vidéo serait le principal moteur de cet accroissement significatif. En effet, elle représentait 40% du trafic total en 2010 et atteindrait 90% en 2015 selon ces prévisions : 60 millions de secondes de vidéo transiteraient ainsi sur le réseau chaque seconde. Cette forte croissance s'explique notamment par la progression des plateformes de partages de vidéos telles que YouTube qui recense un milliard de vidéos regardées par jour et d'autre part, par le développement des services de vidéos à la demande (VOD) pour lesquels des pics à plus d'un million de visionnages par jour sont par exemple recensés sur le réseau IPTV Orange en France. Un défi très important est donc à relever du côté de l'impact des vidéos sur le réseau Internet mondial.

En termes de type de contenu et de modes de consommation, la diversité croît elle-aussi de façon significative, notamment avec le développement de la vidéo en relief d'une part et la généralisation de la HD d'autre part. Celle-ci est relayée par l'apparition de formats d'images plus grands tels que la 2K (2048×1080 pixels) et la 4K (4096×2160 pixels) qui

équipent d'ores et déjà de nombreuses salles de cinéma. Cette course à la résolution va même plus loin avec le format Ultra-HD (UHDTV) ou 8K (7680×4320 pixels) qui correspond à une résolution seize fois supérieure à la HD actuelle et dont les premières expérimentations sont réalisées par la NHK et la BBC. Les modes de consommation tendent pour leur part vers l'exploitation de la diversification des terminaux : l'utilisateur souhaite accéder à ses contenus sur sa télévision, son ordinateur, sa tablette ou son smartphone. Ainsi, parallèlement aux problèmes liés au trafic sur le réseau, la vidéo numérique fait face à de grands défis concernant la qualité des multiples expériences vidéo, sujettes à une demande toujours plus grande.

Face à ces enjeux importants, la nécessité de la compression des médias numériques et de la vidéo en particulier est évidente. Elle représente ainsi un levier majeur pour atteindre ces nombreux objectifs, et ce, malgré l'accroissement des capacités de stockage et l'introduction de débits beaucoup plus élevés grâce à la fibre optique ou au futur réseau sans fil LTE/4G¹. Alors que la compression permettait de diviser par trente la taille d'une vidéo tout en restant visionnable au début des années 1990, le facteur de compression n'a cessé d'augmenter au prix d'évolutions théorique et technologique significatives. L'amélioration de ce ratio demeure cependant un sujet de recherche crucial qui anime une forte communauté de laboratoires académiques et de compagnies internationales. En résultat de ces activités de recherche, un nouveau standard de compression vidéo nommé HEVC (*High Efficiency Video Coding*, appelé H.265 par l'ITU-T et MPEG-H part 2 par MPEG) est en train d'être spécifié et permettra de répondre à une partie des défis à venir.

C'est dans ce contexte qu'un opérateur de télécommunication tel qu'Orange porte un vif intérêt au perfectionnement des techniques d'encodage dans le but de réduire le débit alloué aux vidéos. En particulier, il est nécessaire d'avoir un standard compatible avec ses besoins et respectant les contraintes propres à l'opérateur, à la fois côté encodeur et côté décodeur. De plus, la maîtrise de ces nouveaux outils et dans une plus faible mesure le développement de la propriété intellectuelle du groupe sont autant de raisons motivant des activités de recherche soutenues dans le domaine. Ces travaux de thèse ont ainsi été effectués au sein du Service des Images Animées puis de l'équipe Codage Vidéo Avancé du département Orange Labs du groupe Orange, en collaboration avec le laboratoire I3S, UMR 7271 de l'Université de Nice Sophia-Antipolis et du CNRS.

Contributions

Les travaux réalisés au cours de cette thèse s'articulent autour de deux phases de recherche ayant pour but de proposer de nouveaux outils améliorant les performances en compression des codeurs vidéo actuels avec l'ambition de les voir intégrés dans les futures générations d'encodeurs, à commencer par le standard HEVC. Nos études ont pour point

1. LTE : Long Term Evolution

commun la volonté de limiter l'utilisation d'informations de signalisations additionnelles. Une attention particulière a de plus été accordée à l'exploitation de méthodes issues de domaines connexes à la compression vidéo et s'avérant pertinents dans le cadre de nos études.

La première phase de nos travaux est ainsi davantage orientée vers la normalisation avec la proposition de méthodes visant à améliorer l'encodeur standard H.264/AVC de façon incrémentale. Dans cette phase, nous identifions principalement trois contributions qui apportent des gains significatifs par rapport à la référence :

1. Un module dédié à la transmission d'informations de codage. Bâti de façon originale sur une technique issue du tatouage numérique, nous appliquons notre outil au codage Inter ainsi qu'au codage Intra. Notre méthode repose en outre sur une solution d'optimisation multi-critères originale dans le cadre de la compression vidéo.
2. Une méthode de codage des vecteurs de mouvements. Elle exploite les redondances spatiales et temporelles du champ de mouvement pour modifier de façon judicieuse la distribution des vecteurs résiduels à coder. Cette modification tire de plus profit de l'absence de transmission d'informations additionnelles pour coder efficacement le mouvement.
3. Une variante de notre méthode précédente, obtenue par combinaison avec une méthode de codage du mouvement par compétition de prédicteurs.

La seconde phase de nos travaux est davantage consacrée à une étude amont. Afin de contourner une limite des schémas de codage conventionnels liée à l'augmentation systématique des choix en compétition, nous proposons une approche en rupture pour réduire le coût de l'information de signalisation associé. Basée sur une exploitation accrue du décodeur, cette phase exploratoire a conduit à trois contributions :

1. Une approche dédiée à la détermination des paramètres de codage. Elle est fondée sur un schéma reproductible au niveau du décodeur permettant d'économiser la transmission d'informations très coûteuses tout en s'affranchissant de la contrainte liée au nombre de choix en compétition.
2. Une application de notre approche à la transmission des modes de prédiction Intra.
3. Une mise en œuvre de l'approche décodeur intelligent dans le cadre de l'échantillonnage compressé. Notre travail se concrétise au travers, d'une part, d'un nouveau schéma de codage uniquement construit sur la méthode proposée et d'autre part, d'un mode de codage intégré au sein d'un encodeur de nouvelle génération.

L'ensemble de ces méthodes ont été intégrées au sein d'encodeurs déjà très performants et évaluées de façon rigoureuse afin d'envisager leur utilisation dans de futurs standards. Bien que la complexité ne soit pas évaluée directement au sein de ce manuscrit, nous avons de plus accordé une certaine attention à cet aspect tout en considérant la forte progression des appareils numériques qui offrent davantage de ressources matérielles.

Organisation du document

Ce manuscrit s'organise en deux parties faisant suite à une introduction à la compression vidéo et à une analyse pratique des encodeurs vidéo. Les deux parties proposées présentent respectivement nos travaux orientés vers la standardisation à court terme, et nos travaux plus amont liés à l'approche décodeur intelligent. Chacune de ces parties débute par ailleurs par un état de l'art et une revue du vocabulaire de circonstance. En détails, le manuscrit est organisé comme suit :

- Le chapitre 1 introduit de façon théorique la compression vidéo au travers du standard H.264/AVC. Une description détaillée du futur standard HEVC (horizon 2013), comparé à ses deux prédécesseurs, est ensuite le point d'orgue de cette revue des encodeurs vidéo normalisés ces vingt dernières années. Enfin, mention est faite, non seulement d'autres encodeurs existants, mais également d'autres voies d'améliorations majeures.
- Le chapitre 2 est dédié à l'analyse comportementale des encodeurs. La place des différentes données composant le flux vidéo ainsi que l'organisation du mouvement sont ainsi étudiées. Afin d'illustrer l'analyse théorique du chapitre précédent, une évaluation des performances des trois dernières générations de standards est ensuite réalisée et permet notamment, de situer les travaux présentés dans la suite du manuscrit. Enfin, une étude de l'évolution des schémas de codage est proposée pour mesurer les limites potentielles de ces approches.

Première partie :

- Le chapitre 3 introduit les notions nécessaires à cette première partie. Il commence par un état de l'art sur la signalisation des informations de codage ainsi que sur le codage du mouvement afin de positionner nos travaux. S'ensuit une brève introduction au domaine du tatouage numérique et ses quelques applications en compression afin de servir de base à notre chapitre 4.
- Le chapitre 4 décrit notre méthode de tatouage appliquée au codage vidéo. Après une explication des choix préliminaires, la méthode qui consiste à masquer des informations de codage dans les résiduels de pixels est détaillée dans un cadre général. Deux méthodes de sélection du tatouage par optimisation multi-critères sont alors proposées avant de présenter les résultats obtenus pour deux applications réalisées respectivement dans un cadre Inter et Intra.
- Le chapitre 5 expose notre solution pour le codage des vecteurs de mouvement à partir d'une fine analyse statistique du mouvement. Les trois étapes qui constituent le schéma proposé sont longuement détaillées. Nous discutons ensuite plus précisément du positionnement de la méthode qui se distingue des approches actuelles par l'absence de signalisation additionnelle. Les résultats très prometteurs de notre méthode et ceux d'une variante obtenue par combinaison avec une autre solution sont finalement rapportés.

Seconde partie :

- Le chapitre 6, à la manière du chapitre 3, introduit les notions nécessaires à cette seconde partie. Le transfert d’intelligence au niveau du décodeur est tout d’abord abordé : il peut être réalisé de façon totale dans le cadre du codage vidéo distribué ou bien de façon plus contrainte dans un cadre plus proche des encodeurs conventionnels. La théorie émergente de l’échantillonnage compressé est ensuite présentée à titre de pré-requis pour notre étude du chapitre 8.
- Le chapitre 7 est dédié à notre approche décodeur intelligent. Les motivations sont tout d’abord définies avant d’introduire l’environnement proposé qui est réparti en trois points. S’ensuit, une application directe de notre approche aux modes de prédiction Intra dans HEVC. Des perspectives prometteuses clôturent finalement ce chapitre.
- Le chapitre 8 applique notre approche décodeur intelligent dans le cadre de l’échantillonnage compressé. Exploitée dans un schéma de compression vidéo, cette technique s’inscrit parfaitement dans l’approche proposée du fait des nombreux paramètres à déterminer pour en tirer profit. Une première mise en œuvre théorique démontre l’apport de notre modèle puis, son évaluation, au sein de l’encodeur HEVC, clôture ce chapitre.

Nous terminons ce manuscrit avec les conclusions de nos travaux et les perspectives associées. Finalement, nous présentons en annexe un mode de codage étendant un précédent travail réalisé au sein de notre équipe avant de lister les publications associées à nos travaux.

Chapitre 1

Le codage vidéo, les normes H.264/AVC et HEVC

Une part importante des travaux réalisés dans cette thèse se positionnent dans un cadre normatif entre le standard H.264/AVC et son futur remplaçant HEVC. Nous débutons par conséquent ce manuscrit par une introduction au codage vidéo au travers de la norme H.264/AVC qui est actuellement le schéma de compression le plus utilisé. Basé sur une approche hybride exploitant les redondances spatiales et temporelles du signal vidéo à l'aide de multiples choix de codage en compétition, H.264/AVC a permis d'obtenir des gains en compression significatifs par rapport à ses prédécesseurs. Fort de cette réussite, le nouveau standard HEVC tire encore plus efficacement profit des mêmes principes avec pour objectif une réduction du débit de 50% à qualité subjective équivalente. Ainsi, dans la mesure où le sujet HEVC est très récent, nous consacrons une seconde section à la description détaillée des principaux outils et des innovations qui le composent. Il sera ensuite mis en regard de ses prédécesseurs de manière à appréhender l'évolution des principales générations de standards.

Enfin, afin de disposer d'un état de l'art complet du codage vidéo et dans la mesure où nous aborderons dans nos travaux des sujets au-delà de la normalisation, nous présentons dans la dernière section de ce chapitre un point sur les autres codecs existants ainsi que sur des approches davantage en rupture. Cette étude apporte des éléments de réponses à notre réflexion sur le futur de la compression vidéo.

1.1 Codage vidéo au travers de la norme H.264/AVC

1.1.1 Contexte et synthèse du schéma d'encodage

Le standard H.264/AVC (AVC pour *Advanced Video Coding*), aussi connu sous le nom MPEG-4 Part 10, a été publié dans sa première version en 2003 par un rapprochement entre les groupes VCEG (*Video Coding Experts Group*) ITU-T SG16-Q6 et MPEG (*Mo-*

ving Picture Expert Group) ISO/IEC JTC 1/SC 29/WG regroupés au sein du groupe JVT (*Joint Video Team*). En 2005, l'extension H.264/AVC F_{RE}Ext (*Fidelity Range Extensions*) apportant de nouveaux outils de codage fût finalisée et constitue le profil High de cette norme qui est maintenant l'un des profils le plus utilisé.

Le schéma d'encodage de ce standard est résumé dans la figure 1.1. Il fonctionne en boucle fermée, c'est à dire que les données déjà codées sont disponibles pour traiter la partie courante : on parle de données causales qui sont obtenues par une boucle de décodage intégrée dans l'encodeur. La vidéo en entrée est traitée image par image de façon séquentielle, chaque image étant divisée en *slïces* (représentant une partie de l'image ou l'image entière) eux-mêmes divisés en macroblocs de taille 16×16 pixels pouvant être partitionnés en rectangles de tailles plus petite. Les macroblocs sont traités selon un parcours dit *raster*, débutant en haut à gauche de l'image pour terminer en bas à droite.

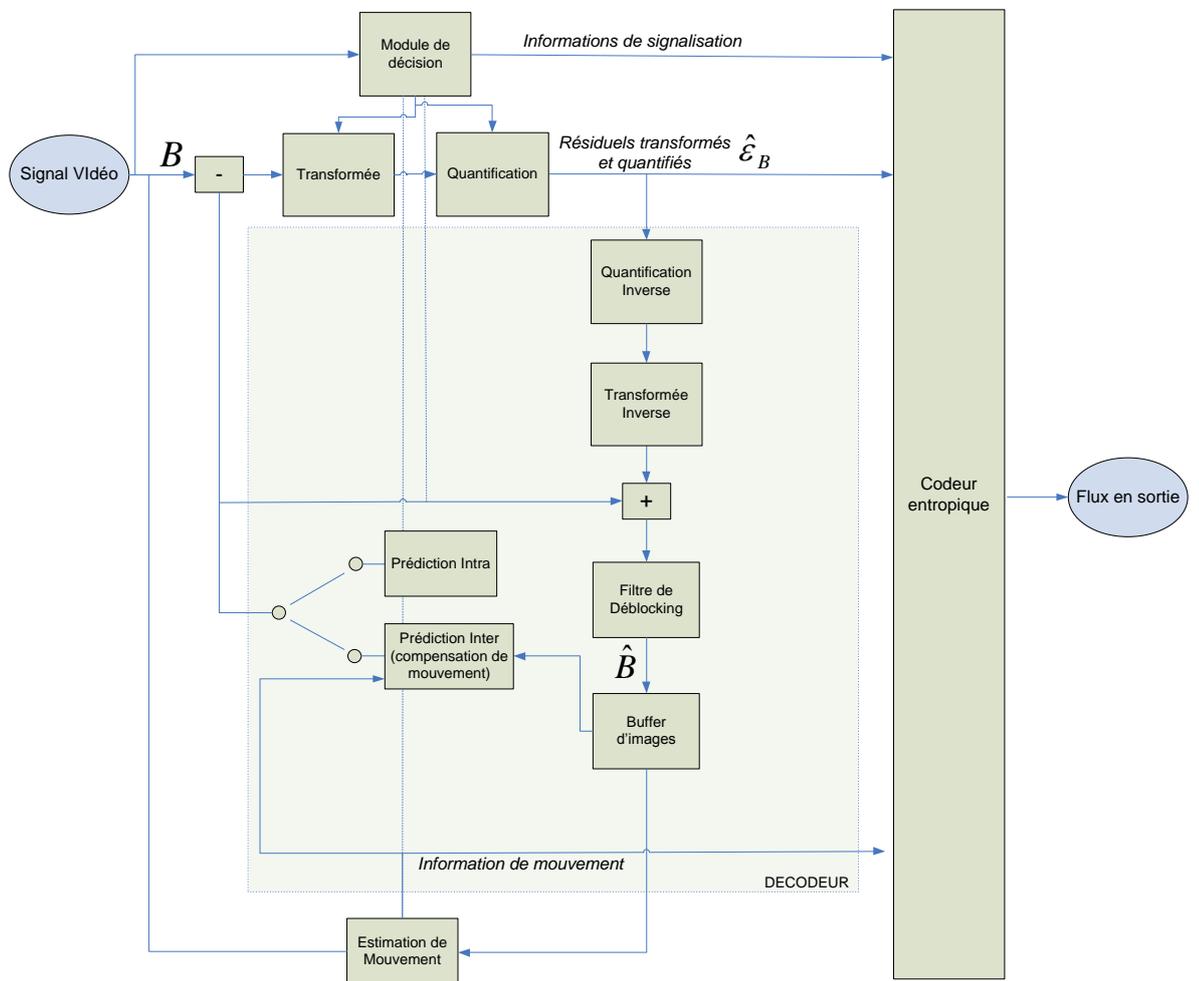


FIGURE 1.1: Vue d'ensemble du schéma de codage du standard H.264/AVC.

Chacun des macroblocs subit alors les étapes de prédiction générant un résiduel de pixels à partir des données causales, transformation et quantification du résiduel puis finalement encodage à l'aide d'un codeur entropique des résiduels de pixels et des informations de signalisation nécessaires au décodage du flux. Le flux binaire obtenu en sortie de l'encodeur peut alors être lu par un décodeur qui restitue une vidéo en répétant ces étapes de façon inverse et en incluant une étape de filtre anti-bloc (en anglais, *deblocking filter*) permettant de diminuer les dégradations du signal. Insistons sur le fait que le décodeur est normatif tandis que l'encodeur ne l'est pas et possède par conséquent une liberté dans ses choix de codage. En effet, les trois étapes clés nécessitent à différents niveaux des choix de codage réalisés par l'encodeur parmi un ensemble de modes et de paramètres en compétition. C'est notamment ce procédé de décision qui n'est pas normatif, il vise dans la majorité des applications à optimiser le compromis entre qualité de reconstruction et coût de codage selon des contraintes d'utilisations données.

Nous décrivons plus en détails les trois étapes principales du schéma de compression dans les sous-sections suivantes.

1.1.2 Calcul du résiduel

Les schémas de codage vidéo tels qu'H.264/AVC exploitent les redondances spatiales et temporelles du signal vidéo à l'aide d'un procédé de prédiction. La prédiction consiste à approximer un ensemble de pixels courant à partir d'informations causales appartenant à une partie de l'image courante déjà codée ou bien à une image précédemment codée. Cette étape de prédiction génère un résiduel, noté ϵ_n , obtenu par une différence entre l'ensemble de pixels original p_n et la prédiction \hat{p}_n :

$$\epsilon_n(x, y) = p_n(x, y) - \hat{p}_n(x, y),$$

où (x, y) indique la position du pixel considéré.

Nous définissons l'étape inverse de reconstruction de l'ensemble de pixels obtenu à partir du résiduel et de la prédiction par :

$$p_n(x, y) = \epsilon_n(x, y) + \hat{p}_n(x, y). \quad (1.1)$$

Formellement, la prédiction \hat{p}_n constitue une estimation statistique de la variable aléatoire p_n à partir d'un ensemble de N variables aléatoires précédemment observées $\Omega_n = \{p_{n-1}, p_{n-2}, \dots, p_{n-N}\}$. On définit alors le prédicteur P_n qui est une fonction déterministe générant la prédiction :

$$\hat{p}_n = P_n(\Omega_n).$$

Dans la suite de ce manuscrit, nous confondrons prédiction et prédicteur. L'enjeu de cette étape de prédiction est majeur dans les schémas de compression tels qu'H.264/AVC dans la mesure où une prédiction efficace réduira considérablement la quantité d'informa-

tions à transmettre, améliorant d'autant le niveau de compression. Ainsi, deux types de prédiction pixelliques sont définies, Intra et Inter, selon que l'on exploite respectivement les redondances spatiales ou bien temporelles :

Prédiction Intra Elle s'apparente au codage d'image fixe et exploite les données causales au sein de l'image courante et plus particulièrement voisines de l'ensemble de pixels courant. La ligne du dessus et la colonne de gauche constituent ainsi l'ensemble des pixels de référence. Dans le standard H.264/AVC, trois tailles de blocs sont utilisées pour ce type de prédiction : 16×16 , 8×8 et 4×4 . Ces partitionnements permettent de s'adapter à la difficulté de prédiction du contenu du macrobloc, une zone texturée utilisera par exemple une prédiction 16×16 alors qu'une zone plus délicate avec des bords sera plutôt codée en Intra 4×4 .

Différents prédicteurs sont définis pour chaque taille de bloc : quatre en Intra 16×16 et neuf en Intra 8×8 et 4×4 . La figure 1.2 présente les prédicteurs utilisés pour le partitionnement 16×16 où le prédicteur vertical est une copie des pixels de la ligne du dessus pour chacune des lignes du prédicteur, le prédicteur horizontal est une copie de la colonne de gauche, le prédicteur DC est une moyenne des pixels de référence et finalement le prédicteur Plane est une moyenne pixels à pixels selon les diagonales entre la colonne de gauche et la ligne du dessus.

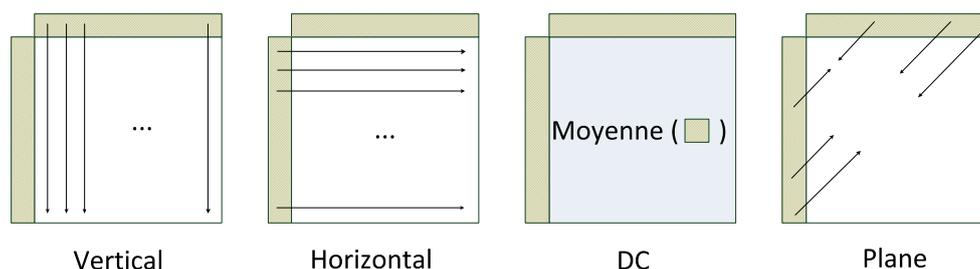


FIGURE 1.2: Prédiction Intra 16×16 .

L'ensemble des prédicteurs des deux autres partitionnements sont semblables avec l'addition d'autres directions de prédictions de manière à capter davantage de redondances dans le signal.

Prédiction Inter Elle réalise une compensation de mouvement à partir d'une image précédemment encodée sur laquelle a été réalisée une estimation de mouvement. Le procédé de compensation de mouvement consiste à utiliser un bloc de l'image de référence en tant que prédicteur. Cette technique s'appuie sur l'observation que deux images successives sont très corrélées dès lors que l'on prend en compte le mouvement. Ce dernier est ainsi déterminé par l'étape d'estimation de mouvement qui consiste à rechercher dans l'image de référence l'ensemble de pixels le plus proche de l'ensemble à prédire selon un critère d'appariement. Comme nous l'étudierons davantage en détail, ce mouvement est alors représenté par un vecteur qui est

transmis au décodeur qui peut réaliser la compensation de mouvement. Par conséquent, l'estimation de mouvement a lieu uniquement à l'encodeur et l'algorithme utilisé tout comme le critère d'appariement ne sont pas normatifs. À la manière de l'Intra, plusieurs partitionnements sont définis dans H.264/AVC en Inter avec l'addition de partitions rectangulaires qui permettent de segmenter efficacement le mouvement. Les tailles de blocs disponibles sont : 16×16 , 16×8 , 8×16 et 8×8 pouvant lui-même être partitionné en 8×4 , 4×8 et 4×4 . Le standard utilise de plus une représentation sous-pixellique du mouvement qui peut ainsi être réalisée jusqu'au $\frac{1}{4}$ de pixel et permet de palier la représentation discrète du signal utilisée dans le numérique alors que la captation a lieu dans le domaine continu. Enfin, un mode particulier est défini en Inter, il s'agit du mode Skip qui s'applique sur les blocs de taille 16×16 et a la particularité de ne nécessiter aucune information additionnelle (ni vecteur de mouvement, ni résiduels de texture).

Enfin, H.264/AVC dispose de trois types d'images selon le moyen de prédiction utilisé :

- Les images I utilisent exclusivement une prédiction Intra. Ces images permettent d'initier le procédé de compression et dans la mesure où elles sont indépendantes des autres images encodées, elles permettent par exemple un accès aléatoire à la séquence.
- Les images P utilisent la prédiction Intra ainsi qu'une prédiction Inter avec une image de référence unique pour chaque bloc. Cette image est nécessairement antérieure temporellement à l'image courante et les deux techniques de prédiction sont en compétition pour chaque macrobloc.
- Les images B utilisent la prédiction Intra ainsi qu'une prédiction Inter avec deux images de référence pour chaque bloc pouvant être antérieur et/ou postérieur temporellement. La compression de ces images est particulièrement efficace grâce à l'exploitation de redondances multiples.

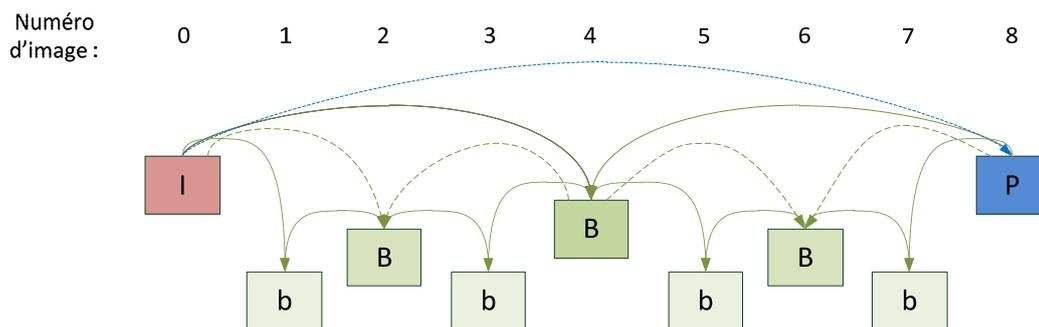


FIGURE 1.3: Structure de Gop IBBBP hiérarchique.

Différentes structures d'images peuvent alors être définies à partir de ces trois types d'images, on parle de GOP (*group of pictures*). La structure IPP utilise une image I puis des images P, la structure IPB utilise une image I puis une image P éloignée de l'image I

avec des images B insérées entre les deux et finalement la structure hiérarchique (figure 1.3) permet d'insérer des images B ayant différents niveaux de directions permettant de réduire significativement le coût de codage des dernières images de la hiérarchie.

1.1.3 Représentation du résiduel

Une fois le macrobloc courant prédit, le résiduel obtenu est représenté dans le domaine transformé puis quantifié afin d'optimiser son codage en réduisant à l'essentiel la quantité d'information à transmettre.

Ainsi, l'opération de transformation permet tout d'abord de décorrélérer le signal dont la représentation dans le domaine spatial ne permet pas d'isoler l'information importante. Le standard H.264/AVC utilise la transformée en cosinus discrète (DCT pour l'anglais *Discrete Cosine Transform*) comme de nombreuses normes de compression à l'exception du standard de compression d'images fixes JPEG2000 qui utilise une transformée en ondelettes. La DCT [Rao *et al.* 07] a en effet pour particularités de compacter efficacement l'énergie du signal observé et de permettre une localisation fréquentielle exploitable pour la suite du codage. La formule de la DCT pour un bloc de pixels p de taille $N \times N$ est donnée dans (1.2) où $F(u, v)$ est le coefficient transformé à la position (u, v) .

$$F(u, v) = \frac{2}{N} c(u) c(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} p(x, y) \cos \left[\frac{\pi u}{N} \left(x + \frac{1}{2} \right) \right] \cos \left[\frac{\pi v}{N} \left(y + \frac{1}{2} \right) \right] \quad (1.2)$$

$$\text{où } c(w) = \begin{cases} \frac{1}{\sqrt{2}} & \text{si } w = 0 \\ 1 & \text{sinon} \end{cases}$$

Précisons que la DCT utilisée dans H.264/AVC [Malvar *et al.* 03] est différente de celle des autres standards puisqu'il s'agit d'une DCT entière qui permet, en plus des propriétés avantageuses de cette transformée, de bénéficier d'une implémentation aisée nécessitant uniquement des décalages et des additions et délivrant des coefficients entiers. Enfin, deux tailles de transformée, 4×4 et 8×8 , sont disponibles dans ce standard et permettent de capter efficacement les redondances entre pixels voisins sans engendrer une complexité calculatoire trop élevée.

Suite à la transformation, le résiduel est alors quantifié avant d'être transmis au codeur entropique. En effet, la transformation permet de représenter le résiduel dans une base appropriée mais ne réalise pas de compression en tant que telle, c'est la quantification qui permet de conserver l'information la plus importante pour l'oeil humain. Ainsi, elle consiste à représenter un signal prenant des valeurs parmi un dictionnaire de grande taille N_d dans un nouveau dictionnaire de taille N'_d telle que $N'_d < N_d$. Le signal nécessite alors moins de bits pour être codé. La quantification introduit par conséquent une erreur de quantification et on parle donc de compression avec *pertes* dans la mesure où le procédé n'est pas inversible.

H.264/AVC utilise une quantification scalaire utilisant un pas de quantification (QP) noté q_k qui définit le niveau de dégradation du signal.

$$Q : \mathbb{R} \rightarrow D = \{x'_1, x'_2, \dots, x'_{N_D}\} \quad (1.3)$$

L'opérateur de quantification Q associe alors un sous-ensemble de l'espace d'entrée à une unique valeur x'_i :

$$Q(x) = x'_i \text{ si } x \in \left[x'_i - \frac{q_k}{2}, x'_i + \frac{q_k}{2} \right[. \quad (1.4)$$

Notons que suite à ces deux étapes, la formule de reconstruction du bloc courant \tilde{p}_n (1.1) est modifiée dans la mesure où le résiduel reconstruit $\tilde{\epsilon}_n$ est différent du résiduel initial ϵ_n , soit :

$$\tilde{p}_n(x, y) = \tilde{\epsilon}_n(x, y) + \hat{p}_n(x, y). \quad (1.5)$$

1.1.4 Codage du résiduel

Finalement, l'ensemble des codeurs de source sont constitués d'une étape de codage sans perte consistant à représenter de façon efficace les symboles à coder au sein du train binaire. Préalablement à cette étape, le bloc de résiduels de pixels à coder est parcouru selon un ordre en zigzag (figure 1.4) permettant de collecter les coefficients transformés et quantifiés selon les fréquences croissantes. En effet, d'après les propriétés de la DCT, l'information essentielle se trouve dans la zone en haut à gauche du bloc tandis que les autres coefficients sont en majorité nuls ou de faibles valeurs. Les coefficients ainsi réordonnés forment alors un vecteur qui est efficacement exploité par l'algorithme de codage RLC (en anglais, *Run Length Coding*) qui remplacent les valeurs nulles par la distance entre deux coefficients non-nuls avant d'être envoyé dans le flux binaire.

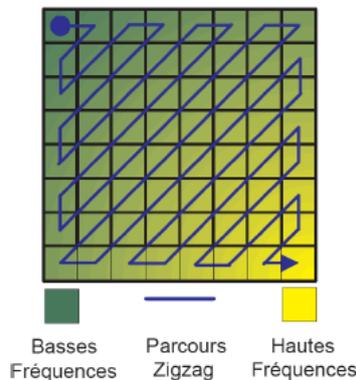


FIGURE 1.4: Parcours zig-zag d'un bloc de coefficients.

Ainsi, les résiduels de pixels tout comme les différentes informations de signalisation sont transmis à l'aide d'un codeur entropique qui exploite les statistiques du signal à coder (la source) pour construire un code composé de mots de code de différentes tailles. On parle ainsi de codes à longueurs variables obtenus avec un codeur tel que le CAVLC (*Context Adaptive Variable Length Coding*) qui est utilisé dans H.264/AVC et affecte les mots de code les plus courts possibles aux symboles les plus fréquents, réduisant de fait le coût de codage de la source. Un second codeur entropique est utilisé dans ce standard et apporte des gains très significatifs de 10% à 15% par rapport au CAVLC au prix d'une augmentation de la complexité. Il s'agit du codeur CABAC (*Context Adaptive Binary Arithmetic Coding*) qui est basé sur l'exploitation de différents modèles en fonction de l'information à coder et de probabilités mises à jour au cours du codage. Nous détaillerons notamment son fonctionnement pour le codage des vecteurs de mouvement dans le chapitre 5.

Dans la sous-section suivante, nous présentons le futur standard HEVC et nous envisageons son positionnement en regard des précédents standards. Ces analyses nous permettront d'isoler une problématique dont nous discuterons dans la seconde partie de ce manuscrit.

1.2 Schéma général de la future norme HEVC

1.2.1 Contexte

A la fin de la période de normalisation du standard H.264/AVC et tout en se concentrant sur les annexes scalables et multivues, le groupe VCEG a créé en avril 2005 un encodeur de travail afin de préparer la future génération d'encodeur normalisé. L'encodeur KTA (Key Technical Area) [KTA] regroupe ainsi les outils les plus efficaces acceptés par le groupe depuis la fin de la normalisation d'H.264/AVC, citons notamment le $\frac{1}{8}$ de pixel [Wedi 00] qui étend la représentation sous-pixellique du mouvement, le filtre d'interpolation adaptatif (en anglais, *Adaptive Interpolation Filter AIF*) [Vatis & Ostermann 06] dont nous parlerons ci-dessous et l'outil de compétition des prédicteurs de mouvement MV-Comp (en anglais, *Motion Vector predictor Competition*) [Jung & Laroche 06].

Sur la base des progrès significatifs démontrés par les résultats du KTA qui apporte alors un gain de plus de 20% en efficacité de codage, le nouveau groupe JCT-VC (Joint Collaborative Team on Video Coding) est créé en juillet 2009 et décide de réaliser un appel à propositions lançant ainsi une nouvelle phase de standardisation. Au même moment, des objectifs précis sont définis afin de guider cette activité :

1. Réduction du débit de 50% à qualité subjective équivalente comparé à H.264/AVC.
2. Multiplication de la complexité par deux ou trois au maximum.

Vingt-sept solutions sont ainsi reçues et évaluées en avril 2010, il s'agit de propositions de qualité variable provenant d'entreprises ou de regroupement d'entreprises ainsi que

d'académiques. Cinq de ces propositions se détachent significativement avec des gains situés entre 30% et 35%. Elles sont retenues et leurs meilleurs outils respectifs sont regroupés au sein du TMuC (Test Model under Consideration) pour former une première version de ce qui constituera le futur standard HEVC (High Efficiency Video Coding). Une période stimulante d'activités de standardisation a ainsi débuté depuis 2010 avec le planning représenté sur la figure 1.5 qui résume globalement le processus depuis la fin des activités sur H.264/AVC. Alors qu'une version finale de la norme HEVC est attendue pour Janvier 2013, les travaux sont d'ores et déjà très avancés et font état de gains significatifs proches des 50% attendus. Preuve de l'enjeu majeur que représente ce standard et plus généralement la compression vidéo, les réunions du groupe JCT-VC qui ont lieu de façon trimestriel font maintenant état de plus de mille contributions par meeting et d'une participation accrue de nombreux acteurs industriels dont des fabricants de chipsets très actifs tels que Texas Instrument et Broadcom.

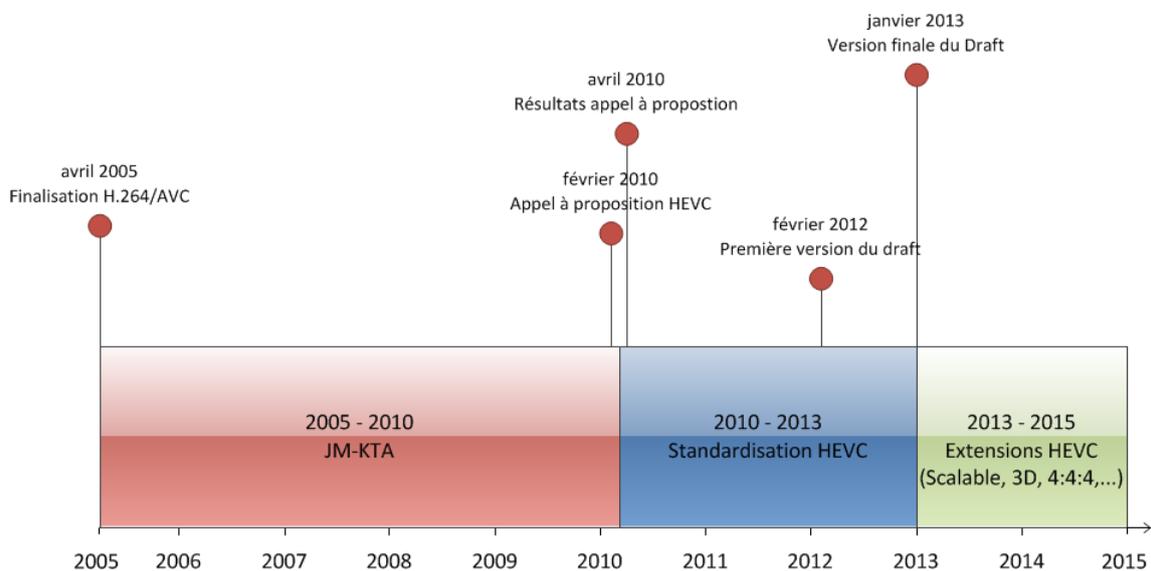


FIGURE 1.5: Planning des activités de standardisation HEVC.

Les travaux actuels étant suffisamment avancés, la structure générale et les principaux outils de ce futur standard sont par conséquent déjà connus, nous les détaillons dans la sous-section suivante à partir des spécifications du HM 5.0 datant du mois de novembre 2011 [Bross *et al.* 11].

1.2.2 Description détaillée

1.2.2.1 Structure de codage

Intéressons-nous tout d'abord à la structure de codage qui repose sur une représentation en arbre quaternaire (en anglais, *quadtree*) permettant de disposer d'un partition-

nement en multiples tailles de blocs s'adaptant aisément au contenu des images tout en ayant un coût de signalisation limité. Trois unités de bases sont définies pour le traitement de chaque image :

Unité de Codage (en anglais, *Coding Unit* notée CU), elle définit la taille de base des blocs pour l'ensemble de la séquence traitée et permet des tailles allant de 8×8 à 64×64 pixels à adapter selon l'application. En particulier, ces tailles élevées d'unité de codage apportent une efficacité de codage significative pour les séquences de résolution supérieure pour lesquelles la corrélation entre pixels augmente significativement. Notons de plus que de telles tailles de blocs sont rendues possibles par le développement des capacités de calcul des équipements numériques.

Unité de Prédiction (en anglais, *Prediction Unit* notée PU), elle définit la taille des partitionnements possibles pour chaque niveau de décomposition de l'arbre et varie en fonction du type de prédiction de la feuille. Ainsi une unité de prédiction peut par exemple être décomposée en quatre partitions de tailles égales ($N \times N$) ou bien peut ne pas être partitionnée ($2N \times 2N$). L'ensemble des tailles possibles est résumé ci-dessous en fonction du type de prédiction :

- Skip : $2N \times 2N$.
- Intra : $2N \times 2N$, $N \times N$.
- Inter : $2N \times 2N$, $2N \times N$, $N \times 2N$, $2N \times U$, $2N \times D$, $nL \times 2N$ et $nR \times 2N$ avec la taille 4×4 retirée.

Unité de Transformée (en anglais, *Transform Unit* notée TU), elle définit la taille pour la transformée et la quantification appliquées à une unité de prédiction. Trois niveaux de décompositions sont possibles au maximum pour cette TU qui prend des tailles allant de 4×4 à 32×32 . Les TU sont exclusivement carrées en Intra et notons que des transformées rectangulaires ont été ajoutées en Inter (4×16 , 16×4 , 8×32 et 32×8) afin d'éviter des cas où la transformée coupe une frontière de bloc et produit des coefficients de hautes fréquences. Enfin, précisons qu'elles peuvent être plus larges que la taille d'une PU et englober ainsi plusieurs PU.

Pour l'encodage d'une séquence, il est nécessaire de définir la taille de la plus grande unité de codage (en anglais, *largest coding unit*, LCU) et la profondeur de partitionnement des CU et des TU. Les tailles de partitionnements des trois unités CU, PU et TU sont alors déterminées de manière récursive au cours du codage. Un exemple de partitionnement d'une LCU 64×64 est donné dans la figure 1.6 avec sa représentation en arbre à droite de la figure. Cette représentation est très avantageuse au vue de la taille élevée des unités de codage principales et permet de plus une flexibilité tant au niveau codage qu'en termes d'implémentation. Le type de prédiction et la décision étant réalisée au niveau de chaque PU (les *feuilles* de l'arbre) il en résulte une diversité des moyens de prédiction qui s'adaptent habilement au contenu à coder.

Calqué sur le principe exploité dans les précédents standards, HEVC s'appuie en effet sur la mise en compétition de différents modes de codage pour traiter chaque PU. Des

moyens de prédiction Intra et Inter sont ainsi utilisés pour capter les redondances spatiales et temporelles du signal vidéo. Nous les détaillons dans les sous-sections suivantes.

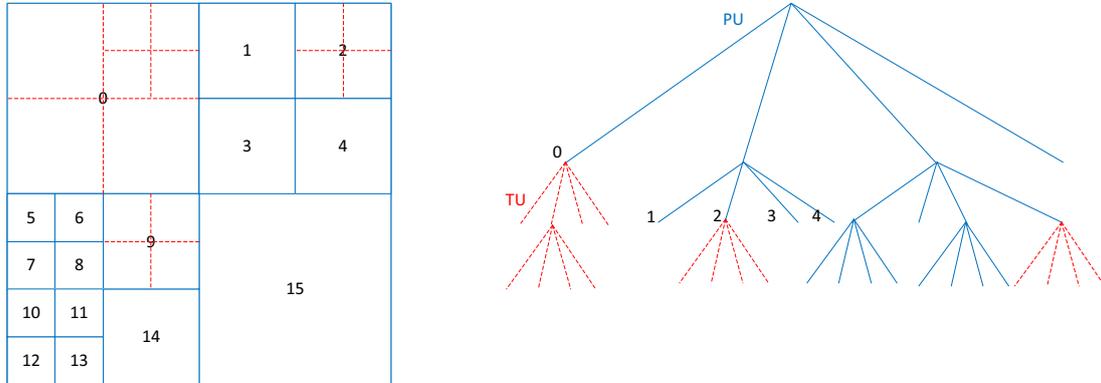


FIGURE 1.6: Exemple de partitionnement en PU(bleu) et TU(rouge) d'une LCU 64×64 .

1.2.2.2 Prédiction Intra

La prédiction Intra est basée sur une nette augmentation de la compétition avec jusqu'à 34 prédicteurs Intra (figure 1.7) pour la Luma composés de 32 directions spatiales, du prédicteur DC (moyenne des pixels de référence) et d'un nouveau prédicteur nommé *Planar*. L'ensemble de ces modes ne sont pas systématiquement activés pour les différentes tailles de PU qui comportent 3 prédicteurs pour les PU 64×64 , 17 pour les PU 4×4 et 34 pour les autres tailles de PU. Ces nombreux modes permettent de capter efficacement les redondances pixelliques dans le voisinage au prix d'une augmentation de la signalisation.

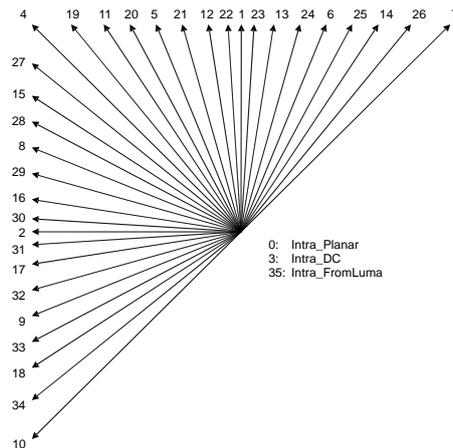


FIGURE 1.7: Liste des prédicteurs Intra de HEVC.

Le prédicteur Planar dont le fonctionnement est représenté sur la figure 1.8 permet d'avoir une prédiction multi-directionnelle. Chaque pixel est ainsi interpolé à partir de la colonne de droite et de la ligne du dessous au sein de la PU courante qui sont substituées respectivement par les pixels causaux au-dessus (TR) et à gauche (LR) de la PU. Ce mode de prédiction qui est particulièrement adapté pour les zones texturées apporte une solution au problème de la prédiction Intra classique qui impose une distance non-négligeable entre le signal de référence et les pixels situés le plus en bas et à droite des blocs.

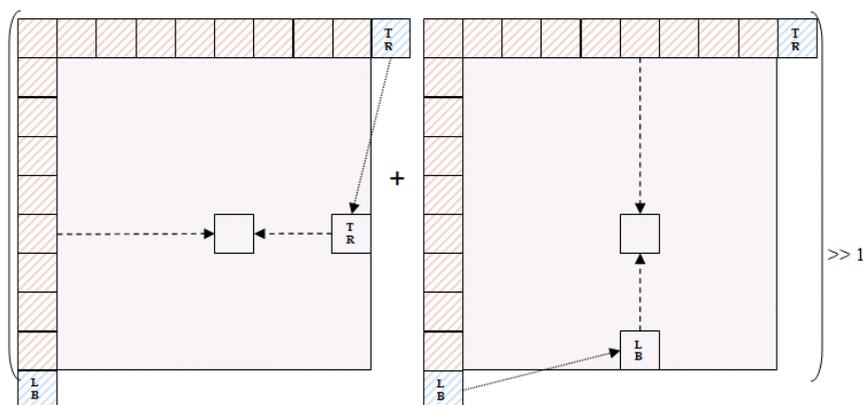


FIGURE 1.8: Représentation du prédicteur Planar.

Pour la composante de chrominance, on retrouve un nombre plus restreint de prédicteurs en compétition avec six modes. L'un de ces prédicteurs réalise une prédiction à partir du bloc reconstruit de luminance selon la formule :

$$Pred_C[x, y] = \alpha \cdot Rec'_L[x, y] + \beta,$$

où $Pred_C$ représente le prédicteur de la PU de chrominance considérée, Rec'_L est la PU de luminance reconstruite après sous-échantillonnage et le couple (α, β) est un couple de coefficients déterminés à partir du voisinage causal.

Ce mode s'avère particulièrement pertinent pour prédire la chrominance et a aussi un impact significatif sur l'efficacité de codage globale grâce à une réduction du coût des résiduels associés.

Pour terminer cette revue des outils Intra, citons la mise en compétition des transformées DCT et DST pour le codage des résiduels de blocs de petites tailles (4×4), la définition de trois ordres de parcours des coefficients transformés (Zig-Zag, Horizontal et Vertical) adaptés en fonction de la direction de prédiction Intra pour améliorer leur traitement par le codeur entropique ainsi que l'outil MDIS (*Mode Dependent Intra Smoothing*) qui applique un filtre sur les pixels de référence afin de réduire les hautes fréquences utilisées pour la prédiction. Finalement, insistons sur le fait qu'un effort particulier a été mené sur la signalisation afin de compenser la nette augmentation de choix

en compétition et on retrouve notamment deux indices most probable mode pour le codage des prédicteurs de la composante de luminance.

1.2.2.3 Prédiction Inter

La prédiction Inter s'articule autour de deux aspects, le calcul de la prédiction et le codage du vecteur de mouvement. Ces deux points ont été améliorés avec un effort particulier sur le codage du vecteur mouvement.

Ainsi, par rapport au calcul de la prédiction pour la compensation de mouvement, un filtre d'interpolation perfectionné est utilisé pour la luminance à la précision $\frac{1}{4}$ pixel avec un filtre 12-tap basé sur la DCT et représenté sur la figure 1.9. Un filtre moins complexe est aussi défini pour des applications plus légères à l'aide d'un filtre 6-tap. La chrominance bénéficie pour sa part d'un filtre bi-linéaire au $\frac{1}{8}$ pixel qui améliore considérablement sa prédiction.

$A_{-1,-1}$				$A_{0,-1}$	$a_{0,-1}$	$b_{0,-1}$	$c_{0,-1}$	$A_{1,-1}$				$A_{2,-1}$
$A_{-1,0}$				$A_{0,0}$	$a_{0,0}$	$b_{0,0}$	$c_{0,0}$	$A_{1,0}$				$A_{2,0}$
$d_{-1,0}$				$d_{0,0}$	$e_{0,0}$	$f_{0,0}$	$g_{0,0}$	$d_{1,0}$				$d_{2,0}$
$h_{-1,0}$				$h_{0,0}$	$i_{0,0}$	$j_{0,0}$	$k_{0,0}$	$h_{1,0}$				$h_{2,0}$
$n_{-1,0}$				$n_{0,0}$	$p_{0,0}$	$q_{0,0}$	$r_{0,0}$	$n_{1,0}$				$n_{2,0}$
$A_{-1,1}$				$A_{0,1}$	$a_{0,1}$	$b_{0,1}$	$c_{0,1}$	$A_{1,1}$				$A_{2,1}$
$A_{-1,2}$				$A_{0,2}$	$a_{0,2}$	$b_{0,2}$	$c_{0,2}$	$A_{1,2}$				$A_{2,2}$

FIGURE 1.9: Interpolation utilisée pour la prédiction de la luminance en Inter dans HEVC.

Le codage du vecteur de mouvement a été un sujet particulièrement traité et qui s'articule autour de deux outils, la prédiction de vecteur de mouvement avancée et le mode Merge. Ces deux outils seront exposés en détails dans le chapitre 3 afin de servir de base à notre étude du chapitre 5. Précisons néanmoins que le premier outil est une extension du module MV-Comp et qu'il est basé sur l'utilisation de prédicteurs de mouvement en compétition (jusqu'à cinq) issus des PU voisines spatialement mais aussi temporellement. La figure 1.10 représente l'origine des blocs voisins spatialement parmi lesquelles

des prédicteurs sont recherchés et auxquelles il faut ajouter un candidat temporel venu de la position centrale du bloc colocalisé. Le vecteur ainsi prédit, il en résulte un résiduel de mouvement réduit qui est transmis au décodeur en compagnie de l'indice du prédicteur optimal.

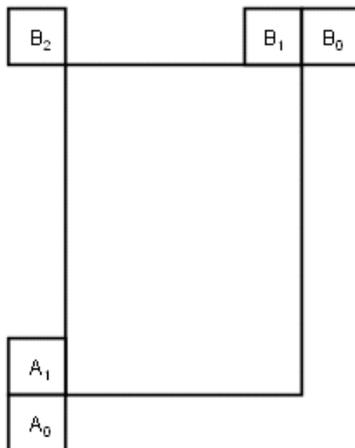


FIGURE 1.10: Représentation des PU d'origines des vecteurs de mouvement prédicteurs candidats parmi le voisinage spatial de la PU courante.

Le mode Merge est quant à lui une extension du mode Direct d'H.264/AVC. Il permet de signaler pour une PU que son vecteur de mouvement doit être hérité d'une PU voisine parmi un ensemble de PU candidate à la manière de l'outil précédent. Ce mode permet ainsi de corriger un défaut lié à la représentation en arbre quaternaire qui a tendance à sur-partitionner l'image. Une PU codée avec le mode Merge économise alors le coût du vecteur de mouvement ainsi que l'indice de l'image de référence, elle ne nécessite alors que la transmission des résiduels de textures.

1.2.2.4 Autres outils

En addition de la structure de codage et de ces outils concernant la prédiction, un effort particulier a été mené sur les filtres et le parallélisme du standard.

Concernant les filtres intégrés au codage, un filtre anti-bloc est toujours utilisé et permet de réduire les artefacts liés au traitement par bloc bien que les tailles plus grandes utilisées dans HEVC réduisent partiellement ce défaut. Deux filtres additionnels sont aussi définis, le Sample-Adaptive Offset Filter et l'Adaptive Loop Filter. Le premier est appliqué après le filtre anti-bloc et consiste à classifier les pixels de la LCU en fonction de leurs intensités puis à appliquer un offset pour chaque catégorie qui est dérivé par rapport à la LCU originale et transmis au décodeur. Le second filtre est quant à lui appliqué avant la copie de l'image dans le buffer et permet d'améliorer la qualité à la

fois objective et subjective de l'image. L'image est partitionnée en zones de différents niveaux d'activité auxquelles est appliqué un filtrage selon deux formes sélectionnées en fonction de leur efficacité.

Par rapport au parallélisme, deux outils sont notamment définis pour permettre un codage efficace pour des applications exploitant le parallélisme : le codage en Wavefront et les Tiles. Le premier outil se base sur une technique bien connue des concepteurs d'encodeurs consistant à traiter les CU en parallèle ligne par ligne avec un décalage du nombre de CU nécessaires pour conserver les dépendances. En particulier, cette méthode permet d'avoir disponible les CU voisines pour la prédiction Intra ainsi que la prédiction du vecteur de mouvement. Sur le même principe, l'outil ajoute en plus une gestion astucieuse des probabilités nécessaires au codeur entropique CABAC. L'outil des Tiles voit quant à lui le problème différemment en divisant les images en plusieurs rectangles à encoder en parallèle. L'intégration de tels outils dans la norme et donc leur exploitation lors du décodage est un réel avantage pour le développement de futurs encodeurs compétitifs.

1.2.3 Evolution des standards de MPEG-2 à HEVC

Afin de positionner ce nouveau standard et le mettre en perspective par rapport à l'évolution des normes de compression depuis une vingtaine d'années, nous discutons ici des différences entre ces schémas de codage. Le standard MPEG-2 [mpe 94] ayant été déposé en 1994, il s'est écoulé dix ans pour définir le standard H.264/AVC et la même période est en train d'être observée pour la définition d'HEVC. Ainsi, la norme MPEG-2 avait été principalement créée pour les applications liées à la télévision numérique et est encore exploitée actuellement. H.264/AVC a alors permis d'améliorer significativement l'efficacité de compression, en divisant par deux le débit nécessaire à qualité équivalente au prix d'une augmentation de la complexité du décodeur d'environ quatre fois celle de son prédécesseur. Cette progression s'est opérée de façon incrémentale en conservant la structure du schéma de codage du MPEG-2 dont quasiment chaque partie a été optimisée et en ajoutant de nombreux nouveaux outils. Les principales différences entre les trois standards sont résumées dans le tableau 1.1. Concernant les deux premiers citons en particulier la modification de la DCT flottante de MPEG-2 en une DCT entière et l'introduction du codeur entropique CABAC. Plus globalement, H.264/AVC se distingue par la nette augmentation des choix de codage en compétition comme les nombreux partitionnements des macroblocs, l'introduction de modes Intra et de deux tailles de DCT. Le passage à H.264/AVC a d'ailleurs nécessité en parallèle un effort particulier mené sur le module de décisions non-normatives permettant de sélectionner les choix optimaux selon un critère donné.

Similairement, la présentation d'HEVC faite précédemment indique que ce futur standard est conceptuellement proche de son prédécesseur qui a été amélioré de façon incrémentale sans addition d'outils en rupture significative. Un schéma hybride exploitant les redondances spatiales et temporelles à l'aide de multiples choix en compétition est

ainsi de nouveau utilisé. Néanmoins, la définition d'une structure de codage plus appropriée et l'augmentation significative des choix de codage à tous les niveaux font d'HEVC l'encodeur le plus abouti. Ainsi, les performances de compression que nous étudierons dans le chapitre suivant ainsi que celles rapportées dans [de Simone *et al.* 10, de Simone *et al.* 11, Li *et al.* 12] confirment ces promesses. De la même manière que pour MPEG-2, les dernières colonnes du tableau 1.1 résument les principaux outils des standards H.264/AVC et HEVC et permettent de mettre en avant l'augmentation de la compétition entre ces deux générations d'encodeurs.

	MPEG-2	H.264/AVC	HEVC
Unité de codage	Macrobloc 16×16	Macrobloc 16×16	LCU 8×8 à 64×64
Partitionnement	Inter 16×8 Intra 8×8	Sous-blocs jusqu'à 4×4	Arbre quaternaire
Transformée	DCT flottante	DCT entière 8×8 et 4×4	TU de 32×32 à 4×4 carré et rectangulaire
Prédiction Intra	Prédicteur DC	9 prédicteurs maximum	34 prédicteurs maximum
Prédiction du mouvement	Vecteur d'un bloc voisin	Médian spatial (trois blocs voisins)	Compétition de 5 prédicteurs
Précision du mouvement	$\frac{1}{2}$ pixel bi-linéaire	$\frac{1}{2}$ pixel 6-tap, $\frac{1}{4}$ pixel bi-linéaire	$\frac{1}{4}$ pixel 12-tap DCT $\frac{1}{8}$ pixel 6-tap chroma
Autres	/	Mode Direct	Mode Merge
Références	Une en P, deux en B	Multiples	Multiples
Codeur entropique	Simple VLC	CAVLC, CABAC	CABAC
Filtres	/	Filtre anti-bloc	Filtre anti-bloc <i>Adaptive Loop Filter</i> <i>Sample Adaptive Offset</i>

TABLE 1.1: Principaux outils des standards MPEG-2, H.264/AVC et HEVC.

De manière à réaliser un tour d'horizon complet des techniques actuelles, nous poursuivons ce chapitre par une présentation des divers schémas alternatifs existants et des principales voies d'améliorations étudiées par la communauté.

1.3 Schémas alternatifs

De façon générale, le développement des diverses générations de standards de façon incrémentale à partir d'un schéma initial unique limite l'apparition d'autres encodeurs suffisamment performants, en particulier ceux basés sur une approche en rupture. L'efficacité de compression des derniers standards a en effet été atteinte suite à de multiples

améliorations minutieuses de l'ensemble du schéma réalisées par une communauté très active stimulée par de forts enjeux économiques. Ainsi, les autres codecs existants sont principalement basés sur un schéma semblable tandis que les approches davantage en rupture fournissent des performances souvent en décalage avec les derniers standards malgré des promesses intéressantes.

Dans cette section, nous proposons un bref état des lieux de ces autres solutions.

1.3.1 Codecs divers

Historiquement, l'un des principaux concurrents du standard H.264/AVC est VC-1 porté par l'entreprise Microsoft qui a été normalisé en 2006 par l'organisme SMPTE (*Society of Motion Picture and Television Engineers*). Il a été notamment normalisé pour être exploité dans les futures générations de DVD (HD DVD et Blu-ray) et servait à l'origine de base au codec Windows Media Video (WMV) 9. Les principales innovations par rapport à H.264/AVC concernent notamment la transformée qui peut être de taille rectangulaire (8×4 et 4×8), la transmission d'un indice indiquant la précision sous-pixellique des vecteurs de mouvement et l'utilisation, addition du filtre anti-bloc, d'une transformée (*overlap transform*) de taille supérieure à celle des macroblocs pour les blocs à forts résiduels. Le schéma de codage de VC-1 demeure cependant très proche d'H.264/AVC et ses performances voisines ont donné lieu à de nombreux débats dans la communauté.

Un autre encodeur standard a vu le jour en 2005 sous l'impulsion du gouvernement chinois avec pour but de rendre la Chine plus autonome dans le secteur multimédia et d'économiser les redevances liées aux autres formats. Désigné AVS (*Audio Video coding Standard*), cet encodeur a une structure et des outils quasiment identiques à H.264/AVC et son utilisation demeure limitée à la Chine.

On retrouve ensuite plusieurs acteurs du côté de la communauté du logiciel libre. Désireux de créer une alternative aux standards soumis à de fortes contraintes d'utilisation liées aux nombreux brevets, différents projets ouverts ont ainsi été développés. Citons notamment les encodeurs Theora et Dirac. Le premier, sous licence BSD est basé sur l'encodeur VP3 de On2 Technologies tandis que le second a été principalement conçu par la BBC et a la particularité d'être basé sur une transformation de type ondelettes. Les performances de ces encodeurs sont cependant limitées et les positionnent au niveau de MPEG-2 ce qui rend leur utilisation marginale.

Finalement, plus récemment et toujours dans cette catégorie des encodeurs libres, Google a poussé une nouvelle solution : le format multimédia ouvert WebM dont la cible principale concerne un usage sur le réseau internet. Pour la partie vidéo, ce format est basé sur l'encodeur VP8 dont Google est propriétaire depuis le rachat de On2 Technologies et qu'il a rendu libre en mai 2010. Les spécifications de ce dernier ont été publiées sous licence Creative Commons et son exploitation relève d'une licence délivrée gratuitement selon certaines conditions interdisant notamment toutes revendi-

cations de propriété intellectuelle. C'est cette gratuité qui en fait un concurrent crédible au standard H.264/AVC qui fonctionne pour sa part sur un principe de redevances. Cependant, WebM est pour l'instant limité par ses performances qui le place légèrement en-dessous d'H.264/AVC et par le relativement faible support industriel par rapport aux standards. D'un point de vue technique, cet encodeur a une structure proche de celle d'H.264/AVC dans son profil baseline et ne dispose notamment pas d'outils majeurs tels que la prédiction bi-directionnelle, la transformée 8×8 , la quantification adaptative au niveau macrobloc (du moins de façon aisée) et l'adaptivité lors de la phase de filtrage.

En conclusion, d'autres encodeurs existent et sont principalement basés sur le même schéma que les standards H.261 à l'exception de l'encodeur Dirac. Leurs performances sont cependant limitées par rapport à H.264/AVC et devraient les positionner loin derrière HEVC mais ces solutions disposent d'autres atouts liés à leur exploitation qui maintiennent une relative concurrence.

1.3.2 Autres voies d'amélioration majeures

A côté des travaux dédiés à la normalisation, plusieurs approches en rupture avec les schémas conventionnels sont étudiées. L'étape de transformation est ainsi un sujet particulièrement traité pour laquelle l'utilisation de la transformée en ondelettes [Cagnazzo *et al.* 04] plutôt que d'une DCT par blocs est une voie souvent envisagée. En particulier, les propriétés des ondelettes liées à la scalabilité ont longtemps semblé prometteuses mais cette transformée n'a pour l'instant pas été retenue en codage vidéo du fait notamment des difficultés liées à la définition d'une compensation de mouvement aussi performante [Cagnazzo *et al.* 07, Agostini *et al.* 06]. Ainsi, les résiduels de textures des standards actuels sont tellement faibles que l'impact de la transformée est trop réduit. Des travaux basés sur le schéma de lifting généralisé [Sole & Salembier 07] qui permet d'implémenter une telle transformée sont néanmoins étudiés pour être appliqué en compression vidéo. Un autre sujet souvent considéré est la définition d'un schéma de codage orienté objet permettant de représenter la séquence par la définition d'une collection d'objets traités séparément qui sont ensuite superposés pour reconstruire la séquence. Une telle approche est définie dans la norme MPEG-4 mais se trouve confrontée au problème très délicat de la segmentation des objets. Le codage par analyse/synthèse de contenu s'appuie sur une idée commune au codage orienté objet et consiste à supprimer des zones d'images lors de l'encodage et de les reconstruire lors du décodage à partir de techniques de synthèses qui exploitent les données déjà décodées [Dumitras & Haskell 04]. Enfin, beaucoup de travaux de recherche dans le domaine du traitement des images, et notamment dans le domaine de la compression vidéo, s'inspirent aujourd'hui du fonctionnement du système visuel humain et des propriétés du code neural. L'étude des mécanismes neuronaux utilisés pour représenter le flux continu d'images nous parvenant apportera certainement de nouvelles solutions algorithmiques de traitement d'information bio-inspiré en potentielle rupture avec l'existant et

qui permettront de représenter plus efficacement un signal vidéo, et donc le compresser [Van Rullen & Thorpe 01, Masmoudi *et al.* 11, Masmoudi *et al.* 12].

D'autres solutions en rupture ont été développées suite à des besoins de nouvelles fonctionnalités. Citons notamment, le codage par description multiple [Goyal 01] introduit avec le développement d'Internet pour les applications de type streaming qui nécessitent une forte résistance aux erreurs pour limiter l'impact de la perte de paquets sur la qualité. Le flux est ainsi divisé en plusieurs descriptions indépendantes contribuant à différentes caractéristiques du signal et de niveau d'importances variés. Le décodage est alors d'autant plus efficace que le nombre de descriptions reçues est élevé [Agostini & Antonini 08], on parle de *dégradation élégante*. De même, le codage vidéo distribué répond pour sa part à la problématique liée à l'utilisation de systèmes légers tels que les téléphone portable et propose de basculer l'ensemble de la complexité au niveau du décodeur. Enfin, le développement de la vidéo 3D entraîne nécessairement des travaux de recherche [Smolic *et al.* 07] pour optimiser la compression de tels contenus. Même si une annexe (H.264/MVC) a été définie, ce sujet reste particulièrement ouvert tant sur le sujet de la représentation des scènes 3D que sur celui de l'exploitation des nouvelles corrélations engendrées. L'utilisation de cartes de profondeur est l'une des solutions envisagées dans la mesure où elle réduit significativement la taille du flux à transmettre comparée à la transmission de deux vues. Dans le cadre de notre travail présenté dans le chapitre 7, nous avons d'ailleurs évalué l'une de nos solutions sur ce type de contenu.

Nous terminons cette sous-section avec une voie d'amélioration récente mais très pertinente qui consiste à exploiter davantage le décodeur de manière à économiser la transmission de certaines données à l'encodeur. Ce type de solution a l'avantage de pouvoir faire évoluer les performances des standards actuels en s'intégrant habilement dans les schémas de codage existants. Cette approche constituera la base de nos travaux présentés dans la seconde partie de ce manuscrit et un état de l'art détaillé lui sera ainsi consacré dans la première section du chapitre 6 qui abordera également le sujet du codage vidéo distribué.

Dans la seconde section de ce chapitre 6, un état de l'art de l'échantillonnage compressé sera aussi proposé. Même si son but initial n'est pas la compression comme nous l'avons présenté ici, cette technique en plein essor offre des possibilités prometteuses dont nous discuterons en détail.

1.4 Bilan

Dans ce chapitre nous avons introduit les notions de bases de la compression vidéo au travers notamment du standard H.264/AVC. Ce dernier est basé sur une approche hybride exploitant les redondances spatiales et temporelles du signal à l'aide de multiples compétitions. Une attention particulière a ensuite été accordée à son futur remplaçant HEVC qui à l'aide de nombreuses améliorations incrémentales constitue l'encodeur existant le plus performant, multipliant par deux les performances de compression de son

prédécesseur. HEVC bénéficie notamment d'une structure de codage plus structurée et de nombreuses modifications qui augmentent considérablement les choix de codage en compétition. Cette constatation est le point de départ de la dernière partie de ce manuscrit et nous en réalisons une analyse plus détaillée dans le chapitre suivant où nous étudions le fonctionnement de ces deux codeurs.

Enfin, pour clore cet état de l'art nous avons évoqué différentes approches se positionnant en alternative aux encodeurs standardisés qui demeurent néanmoins éloignées en performance. Ainsi, les nombreuses activités autour de la compression vidéo prouvent qu'il s'agit d'un domaine très dynamique et certaines des approches en rupture évoquées inspireront probablement les prochaines générations de standards, jusqu'à présent bâtis sur un schéma qui pourrait atteindre ses limites.

Chapitre 2

Analyse comportementale des schémas de codage vidéo modernes

L'amélioration des techniques de compression vidéo nécessite une fine connaissance du comportement des codeurs vidéo existants. Afin d'orienter nos travaux, nous présentons dans ce chapitre les résultats d'une analyse du standard H.264/AVC et de son futur remplaçant HEVC. Après avoir réalisé une courte étude globale du codeur de référence H.264/AVC, nous examinons la répartition des différentes informations de signalisation au sein du flux généré par ce codeur avant d'analyser en détails les caractéristiques du mouvement. La suite du chapitre se concentre sur l'évolution des schémas de codage conventionnels depuis une vingtaine d'années en débutant par une évaluation des performances des trois standards MPEG-2, H.264/AVC et HEVC. Différents points sont ensuite abordés pour notamment examiner l'impact de la nette augmentation de la compétition au sein de ces encodeurs.

2.1 Analyse de l'encodeur H.264/AVC

2.1.1 Analyse globale

A titre d'analyse pratique à mettre en regard de la présentation théorique du chapitre précédent, nous étudions tout d'abord l'encodeur H.264/AVC de façon globale. Rappelons que cet encodeur dispose de trois types d'images (plus rigoureusement, de *slices*) I, P et B respectivement pour les images prédites spatialement, temporellement avec images de référence dans le passé et temporellement avec références dans le passé et le futur (ces deux derniers types d'images exploitant aussi les redondances spatiales). Une fois le type des images de la séquence défini, différents choix de codage sont mis en compétition pour traiter chaque macrobloc. Un encodage est alors caractérisé par la *qualité de reconstruction* d'une image et le *débit* correspondant. Ce débit est distribué parmi l'ensemble des macroblocs de l'image et dépend directement de la répartition des multiples *choix de*

codage en compétition sélectionnés classiquement selon un critère débit-distorsion. La figure 2.1 résume ces trois caractéristiques pour une image P extraite de la séquence Crowdrun 720p encodée avec un profil IBBBP et une quantification $QP = 32$:

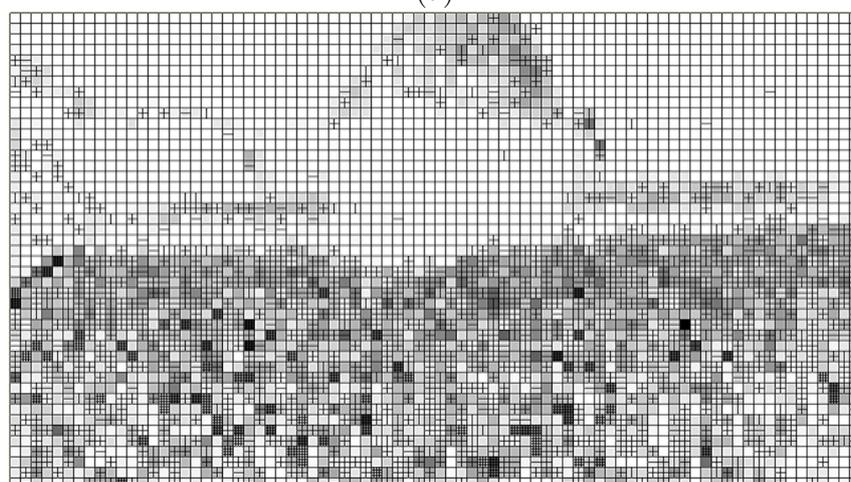
- La figure (a) correspond à l'image reconstruite obtenue et retranscrit ainsi la qualité.
- La figure (b) représente, en niveaux de gris, les coûts de codage de chaque macrobloc. Les coûts comprennent l'ensemble des données transmises (informations de signalisation, résiduels de texture et de mouvement éventuels). Les zones blanches correspondent aux zones les moins coûteuses et les noires sont les plus coûteuses.
- La figure (c) représente la répartition des modes de codage et des partitionnements sélectionnés. Les macroblocs de couleur rouge, verte et bleue correspondent respectivement aux macroblocs codés avec les modes Intra, Inter et Skip.

Ces figures sont instructives pour appréhender le fonctionnement de l'encodeur. En débutant par la figure (c) en regard de la figure (a), nous observons que la prédiction temporelle (Inter et Skip) est la plus utilisée au sein de l'image. Plus précisément, la prédiction Inter (zones vertes) est utilisée pour l'ensemble des zones avec un mouvement relativement complexe, en particulier les coureurs et la zone supérieure de l'arbre qui est agitée. Le mode Skip (zones bleues) est peu coûteux dans la mesure où il consiste à recopier le macrobloc de l'image de référence pointé par le vecteur de mouvement prédicteur : il est ici principalement utilisé pour les parties statiques de l'image, notamment le ciel et les zones d'arbres figées. Notons d'ailleurs que des macroblocs Inter sont utilisés pour séparer des ensembles distincts tels que le ciel et les arbres dans cet exemple. Les vecteurs de mouvements hérités de ces macroblocs permettent ensuite d'obtenir un Skip suffisamment efficace pour compléter les zones correspondant aux arbres. Enfin, quelques macroblocs sont codés avec la prédiction Intra (en rouge), il s'agit de blocs particulièrement détaillés parmi les coureurs où l'estimation de mouvement n'a pas permis de déterminer le mouvement correct. Quelques macroblocs dans le ciel sont aussi codés en Intra avec le mode Intra 16×16 qui est peu coûteux et a par conséquent été privilégié au Skip pour la qualité de reconstruction. Concernant la sélection des différentes tailles de macroblocs, nous observons que des partitionnements plus précis sont utilisés pour les zones les plus détaillées, notamment le bandeau horizontal situé au centre de l'image. La zone supérieure dispose pour sa part de partitionnements plus grands, notamment car le mode Skip n'est disponible que pour une taille de 16×16 pixels.

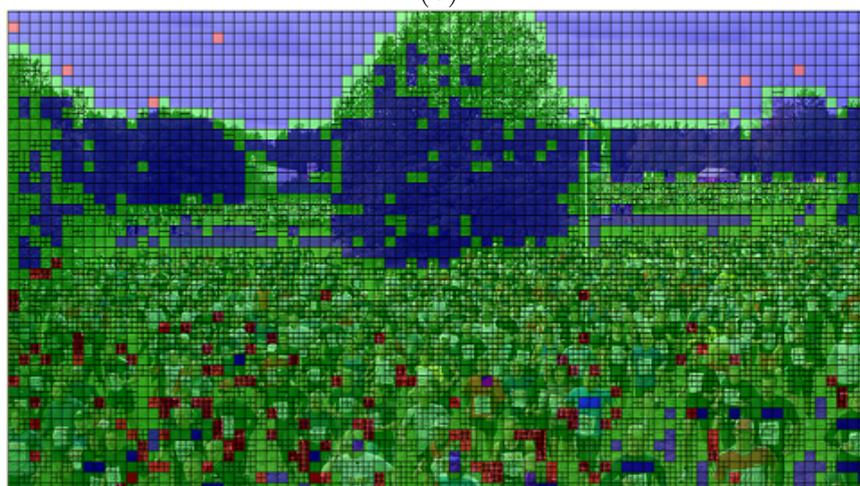
En mettant en correspondance la figure (c) avec la figure (b), nous pouvons alors interpréter la répartition des coûts. De façon générale, la partie supérieure, qui correspond en majorité à des zones codées en Skip, est effectivement la moins coûteuse à l'exception des quelques macroblocs codés en Inter. La partie inférieure de l'image qui a principalement été codée en Inter est plus coûteuse du fait de la transmission de résiduels de textures et de mouvement. Notons enfin que les quelques macroblocs qui ont utilisé des modes Intra sont particulièrement coûteux du fait du coût des résiduels mais aussi des prédicteurs à transmettre. Aussi, les macroblocs les plus partitionnés contiennent généralement davantage d'informations à transmettre notamment ceux en Intra 4×4 qui correspondent aux macroblocs les plus coûteux de l'image (en noirs).



(a)



(b)



(c)

FIGURE 2.1: Représentations visuelles d'un encodage d'une image extraite de la séquence Crowdrun 720p : image reconstruite (a), coût de chaque bloc (b) et répartition des modes de codage (Rouge : Intra, Vert : Inter, Bleu : Skip) (c)₂₉

Cette analyse démontre l'efficacité du standard H.264/AVC qui repose sur une habile sélection de multiples modes et de partitionnements.

Le lecteur pourra notamment se reporter à [Laroche 09] pour obtenir plus de détails sur la sélection des différents modes de codage et des partitionnements. Dans le cadre de notre travail, nous avons souhaité nous attarder sur la répartition des données transmises par l'encodeur au sein du flux en nous appuyant sur l'analyse ci-dessus. Nous avons ainsi distingué les informations selon les catégories suivantes :

- Résiduels de pixels, comprenant les résiduels de luminance et de chrominance.
- Informations de mouvement, comprenant les vecteurs résiduels et les index des images de référence.
- Modes de codage, comprenant la signalisation de l'ensemble des modes Intra, Inter et Skip. Précisons que nous avons inclus dans cette catégorie l'Intra 16×16 qui dispose d'une signalisation particulière dans la mesure où le mode et le prédicteur sont signalés conjointement.
- Prédicteurs Intra, correspondant à la signalisation des prédicteurs optimaux de l'Intra 8×8 et de l'Intra 4×4 . Précision que nous avons distingué cette donnée des modes de codage uniquement pour les images Intra (figure 2.2).
- Autres informations, comprenant notamment l'indice de transformée 8×8 ou 4×4 , le CBP (*Coded Block Pattern*) et le delta QP.

Les proportions de ces données au sein du débit total pour différents niveaux de débits sont représentées sur les figures 2.2, 2.3 et 2.4. Celles-ci correspondent respectivement aux proportions au sein d'images Intra, Inter P et Inter B mesurées à l'aide d'un encodeur H.264/AVC sur un ensemble de dix séquences 720p de contenus variés avec la configuration IBBBP et un GOP de trente-trois images.

Pour la figure 2.2, correspondant aux images Intra, nous notons la proportion majeure des résiduels de textures qui représentent de 55% à 95% du débit total. Les informations de choix de codage, à la fois les modes et les prédicteurs, représentent ensuite une part non-négligeable du flux avec jusqu'à 40% à bas débits où les résiduels sont nettement quantifiés. En particulier, les prédicteurs occupent systématiquement une place significative et atteignent 20% à moyen débit : quatre ou seize prédicteurs sont en effet nécessaires pour chaque macrobloc utilisant les modes Intra 8×8 ou Intra 4×4 . Enfin, nous remarquons que la proportion des modes dépasse celle des prédicteurs à très bas débits, ceci s'explique par la plus forte sélection pour ce type de débit de l'Intra 16×16 qui est inclus dans les données de mode.

La figure 2.3 correspond au flux d'une image Inter P. Elle apporte tout d'abord le même type de conclusion avec la proportion importante des résiduels qui diminue pour les bas débits où les diverses informations de codage prennent une part plus importante. Cette part dépasse ici celle des résiduels de texture à très bas débit puisque l'ensemble des informations représentent environ 60% du flux. En particulier, les informations de mouvement occupent une part significative atteignant un peu plus de 20%. L'enjeu que représente le codage du mouvement est ainsi mis en avant en considérant de plus que son efficacité influe directement sur les résiduels de textures. Un codage efficace

du mouvement réel des macroblochs permet en effet de réaliser une compensation de mouvement judicieuse. Enfin, nous notons que les informations de modes, qui incluent ici les prédicteurs Intra, occupent toujours une proportion significative du flux total.

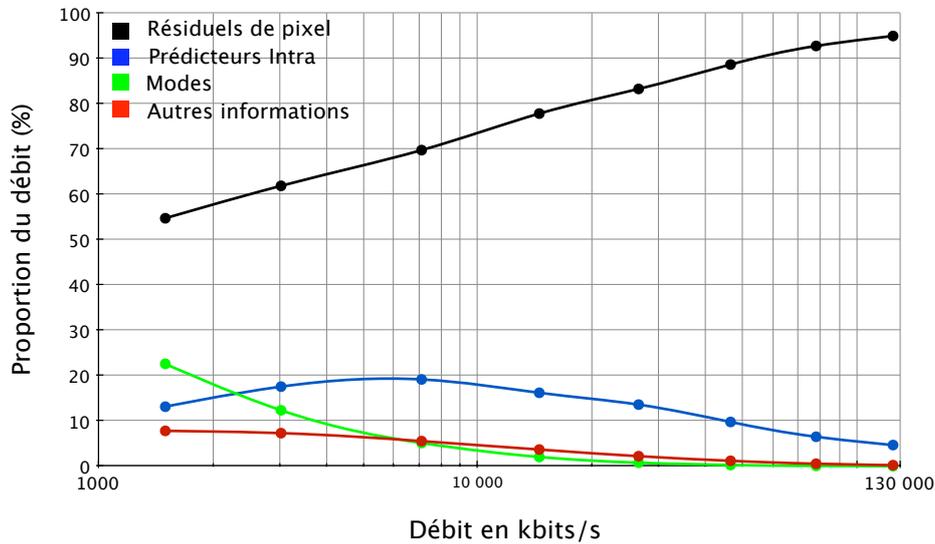


FIGURE 2.2: Pourcentages moyens du débit des principales données de codage par rapport au flux complet au sein des images Intra à différents débits.

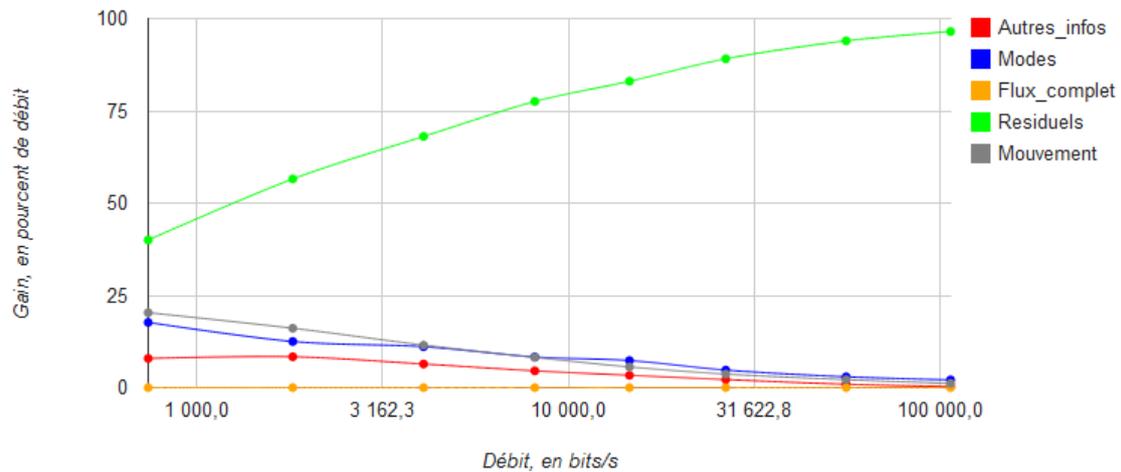


FIGURE 2.3: Pourcentages moyens du débit des principales données de codage par rapport au flux complet au sein des images Inter P à différents débits.

Enfin, la dernière figure (2.4) a été obtenue sur les images Inter de type bi-prédicatif où le mouvement tient une place encore plus importante puisque jusqu'à deux vecteurs peuvent être codés par partition. Nous observons en effet que la proportion du mouvement y est plus élevée que dans les images P ce qui confirme l'intérêt de s'intéresser à l'amélioration de son codage. En particulier, les informations de mouvement sont plus importantes que les informations de modes à la différence des images P où les deux types de données avaient une proportion voisine.

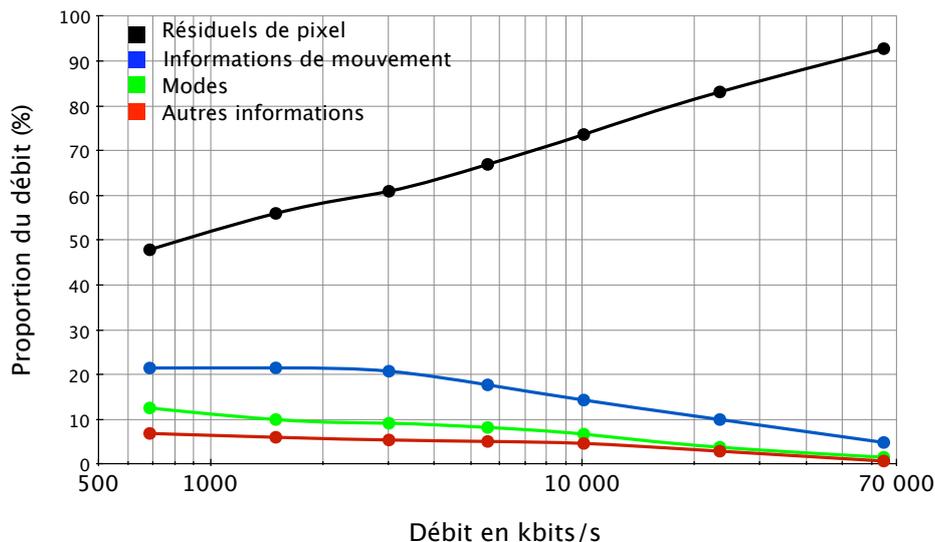


FIGURE 2.4: Pourcentages moyens du débit des principales données de codage par rapport au flux complet au sein des images Inter B à différents débits.

En résumé, les résiduels de textures représentent évidemment les données majeures du flux vidéo. Cependant, leur proportion diminue nettement pour les bas débits au profit des informations de codage qui ont alors un rôle essentiel. En particulier, la transmission des choix en compétition, notamment en Intra, et des données de mouvement en Inter représentent un véritable enjeu pour l'amélioration des performances de compression.

2.1.2 Les informations de signalisation

L'objet de cette sous-section est d'étudier le coût de signalisation des différentes informations de signalisation composant un flux typique généré par l'encodeur H.264/AVC. Cette étude servira de base pour notre travail présenté dans le chapitre 4 qui traite le problème de la réduction de ces coûts. La précédente sous-section a en effet montré que ceux-ci représentent une proportion très significative du flux total, en particulier à bas débits où ils atteignent de 30% à 45% selon le type d'image.

Nous avons mené un test afin d'isoler les données de signalisation les plus importantes au sein du flux et définir celles qui sont les plus pertinentes à traiter. Une collecte a ainsi

été faites pour les images Intra avec la même configuration que celle de la précédente sous-section et ces résultats sont présentés dans la figure 2.5. Nous observons que les informations de compétition telles que les indices des prédicteurs Intra ont une proportion particulièrement significative. Celle-ci évolue avec le débit selon le type de mode, Intra 8×8 ou 4×4 , ce dernier étant moins utilisé à bas débit voit en effet sa proportion diminuée au profit des prédicteurs de l'Intra 8×8 qui représentent plus de 10% du flux total à bas débit. Enfin, une donnée occupe de façon constante une part élevée du flux, il s'agit de l'indice MPM (*most probable mode*) qui est utilisé pour la transmission des prédicteurs Intra et que nous définissons dans le chapitre suivant. Cet indice binaire intervient dans chaque partition à l'exception des macroblocs 16×16 où le prédicteur est transmis avec la signalisation du mode et on observe d'ailleurs, qu'excepté à très bas débit, cette donnée (la courbe violette) occupe une proportion relativement faible du flux. Ainsi, nous considérons l'indice MPM pour l'application en codage Intra de notre méthode de transmission d'informations de signalisation présentée dans le chapitre 4.

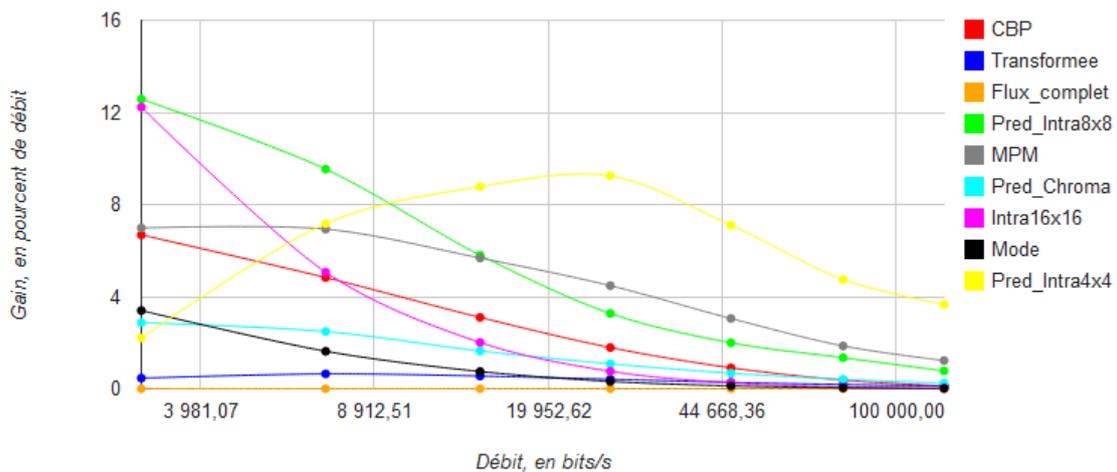


FIGURE 2.5: Pourcentages moyens du débit des différentes informations de signalisation par rapport au flux complet au sein des images Intra à différents débits.

Une étude similaire a été menée pour les images de type Inter où les prédicteurs Intra ont toujours une part significative. Les autres données majeures sont les indices des images de référence et l'indice de la transformée 8×8 ou 4×4 . Ce dernier indice n'est cependant pas compatible avec la solution que nous proposons par la suite car celle-ci nécessite la lecture des résiduels de texture lors du décodage qui dépend elle-même de l'indice de transformée. Comme nous le présentons dans le chapitre 4, en Inter la méthode est appliquée à la réduction du coût de signalisation d'un outil développé au sein de notre équipe : l'outil MV-Comp [Laroche *et al.* 08]. Il s'agit d'un outil dédié au codage du mouvement utilisant une compétition entre plusieurs prédicteurs. Associé à chaque

vecteur de mouvement, un indice binaire est ainsi transmis pour signaler le prédicteur optimal. Nous étudierons plus en détails cette méthode dans le chapitre suivant mais nous avons souhaité faire apparaître ici la proportion de l'indice généré au sein de l'information de mouvement et du débit total afin de positionner nos résultats du chapitre 4.

Les statistiques ont été générées pour neuf séquences des formats CIF à 1080p encodées avec le codec JM KTA 2.1 qui est une version d'H.264/AVC incluant notamment l'outil MV-Comp. Les résultats sont résumés dans le tableau 2.1 pour trois valeurs de pas de quantification et seuls les modes Inter 16×16 à 8×8 ont été activés. Nous observons que la proportion de cet indice est non négligeable puisqu'elle représente jusqu'à 22.8% en moyenne de l'information de mouvement à très haut débit. Ce pourcentage diminue à moyen débit où il représente 11.7%. Evidemment, cette information a une proportion réduite dans le flux total avec 0.3% à très haut débit et 1.5% à moyen débit. L'augmentation de la proportion avec le pas de quantification s'explique par la proportion de l'information de mouvement dans le flux total qui augmente elle aussi avec le QP.

	Pourcentage de l'information de mouvement			Pourcentage du débit total		
	QP 12	QP 22	QP 32	QP 12	QP 22	QP 32
Foreman CIF	15.0	8.0	7.9	0.4	0.8	1.9
Bbc_news CIF	20.7	17.0	7.6	0.3	0.7	1.0
Schumacher CIF	13.1	11.8	7.4	0.3	1.1	1.3
Drama CIF	29.2	15.5	10.6	0.3	0.7	1.4
News_car CIF	20.8	20.2	3.3	0.2	0.6	0.3
Rugby CIF	11.1	10.3	10.4	0.2	0.5	1.4
Keiba3 WVGA	69.4	64.0	31.4	0.9	1.9	3.3
Crowdrun 720p	15.5	14.3	14.0	0.2	0.5	1.5
Crowdrun 1080p	10.1	10.7	12.3	0.1	0.4	1.5
Moyenne	22.8	19.1	11.7	0.3	0.8	1.5

TABLE 2.1: Proportion de l'indice de compétition de prédicteurs de mouvement dans l'information de mouvement et dans le flux total (en pourcentage).

2.1.3 L'information de mouvement

Dans la sous-section 2.1.1, nous avons mis en évidence le rôle majeur de l'information de mouvement dont la proportion dans le flux est significative, en particulier pour les moyens et bas débits. L'enjeu que représente l'information de mouvement étant établi, nous nous intéressons maintenant à ses caractéristiques.

La figure 2.6 présente une classification des vecteurs de mouvement selon leur corrélation avec les vecteurs mouvements du voisinage (en pourcentage). Les vecteurs codés

sont tout d'abord classifiés en fonction de leur distance (en quart de pixel) au vecteur le plus proche au sein d'un voisinage constitué des blocs voisins spatialement et temporellement définis ci-dessous, les vecteurs trop éloignés sont dans la catégorie « Autres ». Nous classifions ensuite chaque vecteur effectivement codé selon qu'il appartient :

- à un bloc voisin spatialement : bloc à gauche du bloc courant, au dessus, au-dessus à gauche ou au-dessus à droite.
- à un bloc voisin temporellement : bloc colocalisé du bloc courant dans l'image précédente ou l'un des huit blocs autour de ce bloc colocalisé.
- à la fois à un bloc voisin spatialement et un bloc voisin temporellement.

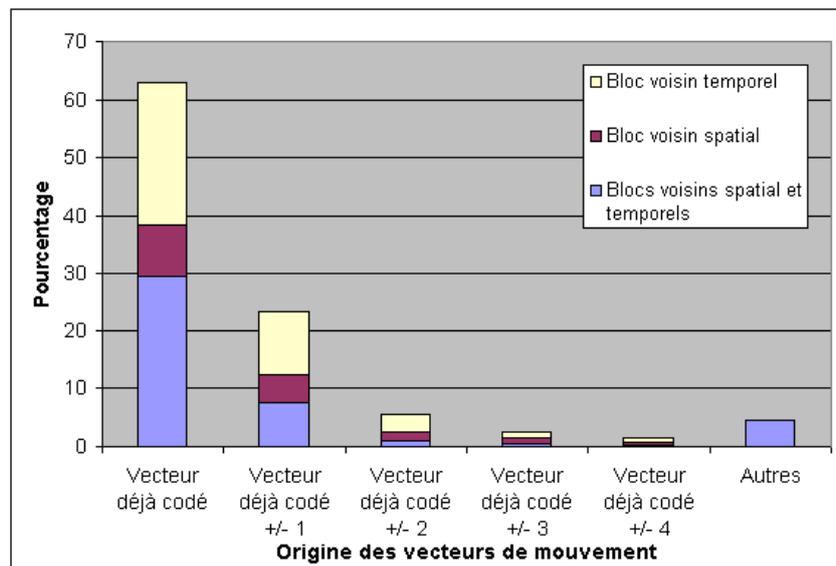


FIGURE 2.6: Classification de l'origine des vecteurs de mouvement selon le type de leur bloc d'origine et la distance en quart de pixel à un vecteur déjà codé dans le voisinage (en pourcentages).

L'analyse a été menée au sein du codec H.264/AVC JM pour un ensemble de quatre séquences QWVGA de contenus variés. Nous observons tout d'abord qu'une proportion très élevée des vecteurs est fortement corrélée aux vecteurs déjà codés. Ainsi, seulement 5% des vecteurs correspondent à un mouvement réellement nouveau (de norme supérieure ou égale à 5 positions au quart de pixel), mieux, plus de 60% des vecteurs ont déjà été codés parmi le voisinage local que nous avons étudié. Ce résultat est très important et sera à l'origine de notre travail du chapitre 5. Il indique que le principe de la prédiction par un vecteur déjà codé comme le médian spatial est pertinent mais cependant elle ne permet pas de couvrir l'ensemble des redondances dans le voisinage. D'autre part, notons qu'une telle prédiction a ses limites dans la mesure où elle a tendance à biaiser l'apparition de mouvements différents puisqu'ils sont alors codés à coût très élevé.

Quant à la répartition des vecteurs selon leur origine, le point principal est qu'une

majorité (plus de 40%) des vecteurs sont très corrélés aux vecteurs issus de blocs voisins à la fois spatialement et temporellement. Ces situations correspondent à des faibles mouvements entre images successives et suggèrent que le vecteur courant peut être codé très efficacement en considérant ces redondances. Nous notons ensuite la corrélation très importante avec les vecteurs issus de blocs voisins temporellement. Ce résultat doit être analysé en tenant compte du fait que les vecteurs issus de neuf blocs de l'image précédente ont été considérés contre seulement quatre blocs voisins spatialement du fait du traitement de l'image en ordre de parcours. Enfin, les vecteurs issus des blocs voisins spatialement sont évidemment eux-aussi corrélés au vecteur du bloc courant avec 20% en moyenne auxquels s'ajoutent les 40% en communs avec les vecteurs temporels.

A titre d'illustration de cette analyse, nous avons représenté sur la figure 2.7 les champs de vecteurs de deux images successives de la séquence Runners 720p qui a été codée avec un encodeur H.264/AVC à débit suffisamment élevé (QP =22) pour avoir un nombre de vecteurs significatifs sans avoir trop de macroblocs Intra. La figure (a) correspond à une image antérieure à l'image de la figure (b). Les images sont quadrillées en macroblocs qui sont eux-mêmes éventuellement partitionnés selon le choix de codage réalisé et un vecteur par partition Inter est représenté en rouge avec pour origine la pointe fine (les zones sans vecteurs correspondent à des zones codées en Intra). La séquence considérée est intéressante pour étudier le mouvement, elle est constituée de coureurs allant de la droite de l'image vers la gauche et situés à différentes distances de la caméra avec deux types de mouvement, l'un global qui est dirigé horizontalement dans le sens droite vers gauche et d'autres mouvements plus locaux correspondant à certaines parties du corps des coureurs (mains et pieds notamment). Nous observons tout d'abord la très forte corrélation entre les vecteurs voisins au sein des deux images respectives, en particulier pour les nombreuses zones correspondant au mouvement global. Environ deux tiers des vecteurs de l'image ont ainsi la même direction et le même sens, seul la norme du vecteur varie légèrement selon la vitesse associée aux différentes zones. Les vecteurs correspondant aux mouvements plus locaux sont pour leur part significativement différents des vecteurs voisins spatialement mais ils se retrouvent dans l'image précédente pour des blocs relativement proches du bloc colocalisé. C'est en particulier le cas pour les pieds du coureur au premier plan et les jambes ainsi que le bras du second coureur. Ces champs de mouvement sont très corrélés entre les deux images et peuvent être pronostiqué en prolongeant les vecteurs localement.

L'intuition confirmée par les résultats présentés sur la figure 2.6 est donc qu'une très forte majorité des vecteurs de mouvement se retrouvent dans les blocs voisins à la fois spatialement et temporellement. La prédiction par médian spatial utilisée dans H.264/AVC est par conséquent pertinente mais seulement de façon partielle dans la mesure où elle permet en effet de couvrir une majorité des redondances spatiales mais échoue cependant dès lors qu'une rupture dans le mouvement intervient. Les techniques par compétition de prédicteurs, telles que celles utilisées dans HEVC et que nous détaillons dans le chapitre suivant, permettent de corriger une partie de ce problème en augmentant le nombre de

prédicteurs et en intégrant des prédicteurs temporels. Ce type de solutions apporte des gains très significatifs, cependant elles n'apparaissent pas encore optimales puisqu'il faudrait faire intervenir au moins une quinzaine de prédicteurs en compétition pour capter quasiment la totalité des redondances mais cela serait alors moins performant du fait du coût de signalisation engendré. Dans le chapitre 5, nous proposons une solution originale permettant de franchir ces limites pour coder très efficacement le mouvement.



(a)



(b)

FIGURE 2.7: Représentation des champs de mouvements de deux images successives de la séquence Runners 720p codée avec H.264/AVC à QP=22.

2.2 De MPEG-2 à HEVC

2.2.1 MPEG-2, H.264/AVC, HEVC : évaluation des performances

Afin de débiter cette section, nous réalisons une évaluation des performances des trois dernières générations d'encodeurs standardisés : MPEG-2, H.264/AVC et HEVC. Cette

comparaison pratique est à mettre en regard de la comparaison théorique proposée dans la sous-section 1.2.3 du précédent chapitre.

Cette évaluation a été réalisée dans le cadre de cette thèse lors d'une étude préliminaire du codeur HEVC effectuée en collaboration avec l'entreprise ATEME¹ que nous remercions. Les encodeurs H.264/AVC et MPEG-2 utilisés correspondent à des versions de recherche développées par ATEME aux performances très poussées proches de solutions utilisées de façon professionnelle. Le but de l'étude est de démontrer la pertinence de l'arrivée sur le marché du standard HEVC, elle permet dans le cadre de cette thèse de positionner nos travaux et discuter de l'évolution des schémas de compression. Le protocole expérimental utilisé pour ces tests est précisé ci-dessous, un large intervalle de débits ont été visés pour cinq séquences 720p réputées pour leur difficulté de codage et chaque encodeur a été configuré dans le meilleur profil disponible tout en veillant à utiliser des outils comparables. En particulier, aucun outil de contrôle de débit n'est utilisé pour les trois encodeurs pour lesquelles nous avons seulement défini de façon identique un *delta* entre les pas de quantification de chaque type d'image. En outre, l'outil RDO-Q qui fait varier le pas de quantification au niveau macrobloc est désactivé pour les trois encodeurs de manière à ne pas biaiser l'étude.

MPEG-2 Meilleur profil (prédiction bi-directionnelle avec 3 images B).

H.264/AVC Profil High (prédiction bi-directionnelle hiérarchique avec 15 images B).

HEVC HM 4.0, Random Access High Efficiency (prédiction bi-directionnelle hiérarchique).

Débits Large intervalle de débits pour les codeurs MPEG-2 et H.264/AVC, QP= {22, 27, 32, 37, 42} pour HEVC.

Séquences 5 séquences 720p à 50Hz : Crowdrun, Soccer, Victoria, Runners, Stockholm.

Evaluation Gains débit-distorsion calculés avec la métrique de Bjontegaard et courbes débit-distorsion (PSNR-Y).

Les résultats obtenus sont tout d'abord résumés dans le tableau 2.2 qui présente les pourcentages de gains entre deux générations successives d'encodeurs. En moyenne, des performances du même ordre de grandeur sont obtenues avec 42.0% de gains pour H.264/AVC face à MPEG-2 et 43.9% pour HEVC en comparaison avec H.264/AVC. Ces gains sont proches des 50% annoncées et ont été obtenus sur des séquences difficiles à coder, en particulier la séquence Crowdrun et la séquence Victoria pour lesquelles le gain entre les deux premiers standards évalués était lui-aussi plus limité. La première est en effet une séquence de course à pied avec de nombreux détails et beaucoup de mouvement tandis que la seconde séquence inclut une partie avec des confettis complexe en termes de reconstruction. L'apport des nouveaux filtres sur ce type de séquence est prépondérant

1. ATEME est une société française leader mondial des solutions d'encodage vidéo professionnelles H.264/AVC et MPEG-2 pour le marché "Broadcast & Broadband", supportant tous les formats, du mobile à la ultra HD et la 3D.

tandis que les séquences avec un mouvement difficile, telles que les séquences de sport, bénéficient de l'amélioration du codage du mouvement.

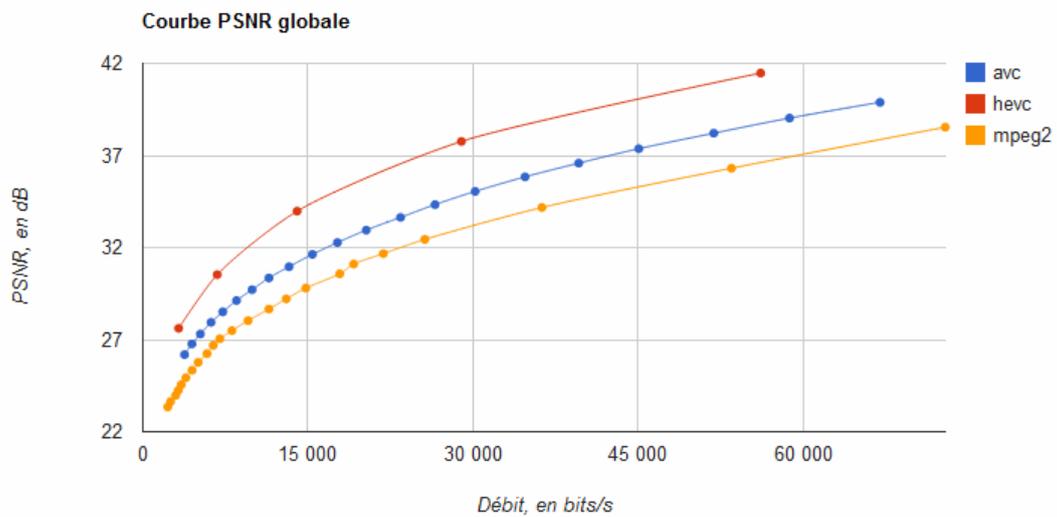
Cette expérimentation tend donc à confirmer les résultats annoncés par le groupe JCT-VC et prouvent la pertinence du nouveau standard HEVC dans le mesure où l'on observe le même type de gains obtenus après une période d'activité similaire d'environ dix ans. Notons de plus que la version d'HEVC que nous avons testée ne correspond pas à la version finale du standard qui sera défini dans un an et dont les performances seront certainement supérieures. Enfin, HEVC est encore mieux conçu pour des séquences de plus grands formats tel que le 1080p où les grandes tailles de partitions sont d'autant plus judicieuses qu'il y a un nombre plus élevé de pixels.

	H.264/AVC vs MPEG-2	HEVC vs H.264/AVC
Crowdrun	31.5	43.6
Soccer	57.8	45.2
Victoria	18.7	31.9
Runners	39.0	44.8
Stockholm	63.1	53.9
Moyenne	42.0	43.9

TABLE 2.2: Pourcentages de gain moyen sur le débit calculés avec la métrique de Bjontegaard pour les débits moyens entre les standards H.264/AVC et MPEG-2 ainsi qu'entre H.264/AVC et HEVC.

Afin de visualiser ces performances, nous avons représenté les résultats obtenus sur les figures suivantes. Les figures 2.8 et 2.9 présentent les courbes débit-distorsion des séquences Crowdrun et Soccer accompagnées d'images extraites des flux vidéo reconstruits tandis que la figure 2.10 regroupe les courbes des trois autres séquences testées, Victoria, Runners et Stockholm. Ces courbes confirment les performances annoncées, en particulier pour les bas débits qui permettent d'atteindre des niveaux de qualités intéressants de 30dB à 35dB pour les séquences de sports (2.8, 2.9 et 2.10 b) à environ 7Mbits/s et jusqu'à 40db pour ce même type de débit sur une séquence moins complexe telle que Stockholm (2.10 c).

Les images extraites de la séquence Crowdrun (2.8) encodée à 6.7Mbits/s avec H.264/AVC (b) et HEVC (c) montrent notamment l'efficacité d'HEVC à restituer des détails difficiles comme l'arbre au centre de l'image pour lequel les effets de blocs dans H.264/AVC sont très visibles. L'apport des grandes tailles de blocs combinées aux nouveaux filtres est quant à lui visible sur l'exemple de la séquence Soccer (2.9) pour lequel la pelouse est particulièrement bien reconstruite avec HEVC.



(a)

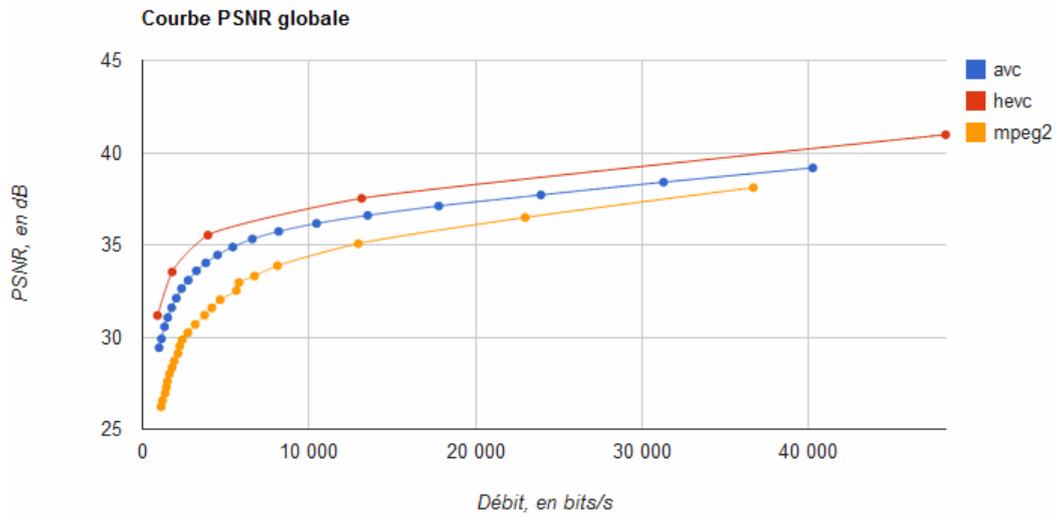


(b)



(c)

FIGURE 2.8: Courbes débit-distorsion des trois standards (MPEG-2, H.264/AVC, HEVC) pour la séquence Crowdrun 720p (a), images tirées de la séquences encodée à 6.7Mbits/s avec H.264/AVC (b) et HEVC (c).



(a)

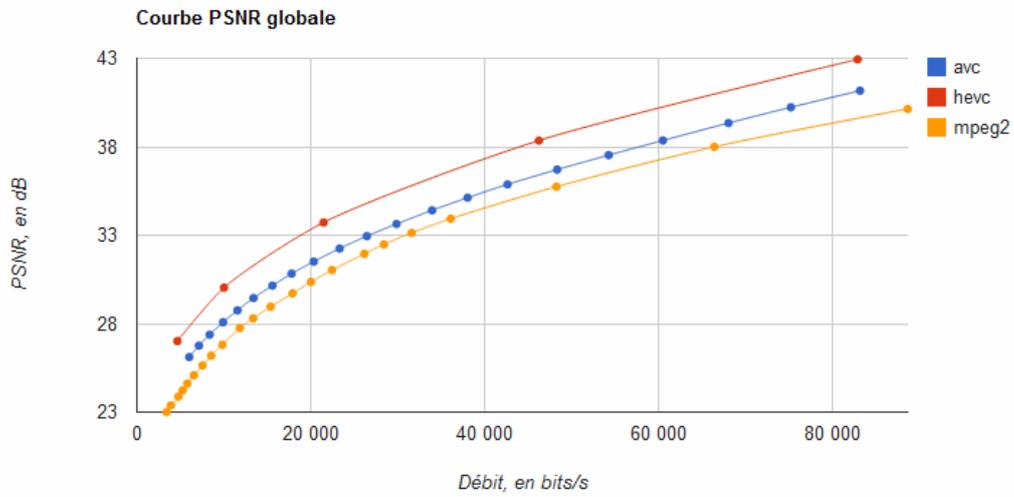


(b)

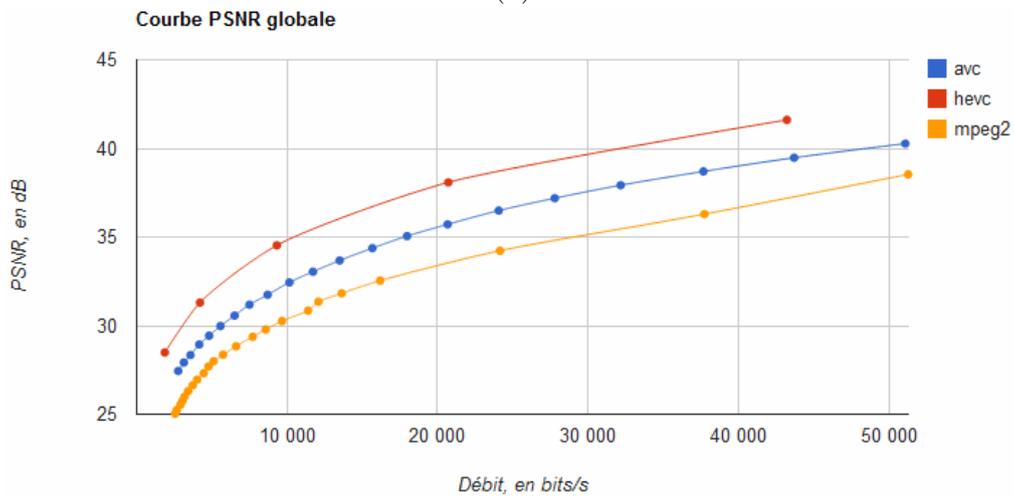


(c)

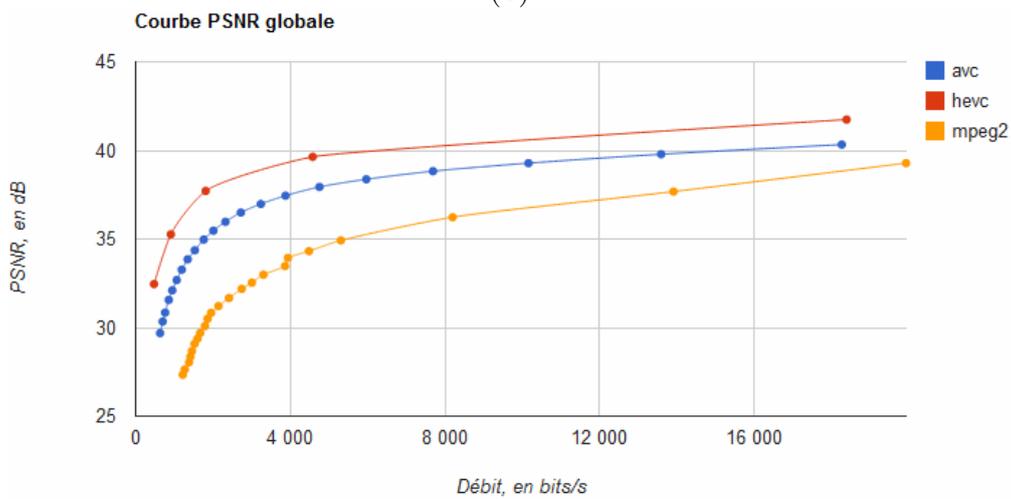
FIGURE 2.9: Courbes débit-distorsion des trois standards (MPEG-2, H.264/AVC, HEVC) pour la séquence Soccer 720p (a), images tirées de la séquences encodée à 4Mbits/s avec H.264/AVC (b) et HEVC (c).



(a)



(b)



(b)

FIGURE 2.10: Courbes débit-distorsion des trois standards (MPEG-2, H.264/AVC, HEVC) pour les séquences Victoria (a), Runners (b) et Stockholm (c).

2.2.2 Vers une limite des approches conventionnelles ?

L'évaluation des performances des trois générations d'encodeurs a montré que des sauts de performances proches de 50% entre chaque génération sont confirmés. Afin d'envisager la prochaine évolution, nous avons souhaité analyser ces encodeurs à travers leur efficacité de prédiction respective et l'impact de l'augmentation de la compétition. Cette étude sert de base à nos travaux de la seconde partie du manuscrit.

Evolution de la qualité de la prédiction

L'amélioration des techniques de prédiction entre chacune des générations d'encodeurs a donné lieu à une réduction des coûts de codage de la texture combinée à une meilleure qualité de reconstruction.

Tout d'abord, pour apprécier cette évolution entre les standards MPEG-2 et H.264/AVC, nous présentons sur la figure 2.11 deux images de résiduels obtenus avec chacun de ces encodeurs. Nous avons considéré une image extraite de la séquence Crowdrun 720p codée en prédiction Inter bi-directionnelle hiérarchique (image B) et appartenant au dernier niveau de la hiérarchie. Afin que le test soit pertinent, la même structure de GOP a été utilisée pour les encodeurs MPEG-2 et H.264/AVC avec trois images B suivies d'une image P. Nous avons réalisé l'encodage pour un débit classique de l'encodeur H.264/AVC, autour de 6.6Mbits/s, délivrant un PSNR de 30dB ; le même PSNR a ensuite été obtenu à l'aide de l'encodeur MPEG-2 pour un débit de 13Mbits/s. La figure (a) correspond à l'encodeur MPEG-2 et la figure (b) à H.264/AVC. Dans les deux images, les zones grises sont des zones sans résiduels, le plus souvent obtenues avec le mode Skip. On peut observer que la densité de résiduels a nettement diminué avec H.264/AVC, en particulier dans la partie supérieure de l'image où il n'y a alors quasiment plus de résiduels. Ceci s'explique par la qualité supérieure des vecteurs de mouvement du fait de leur codage plus efficace dont le mode Skip bénéficie. De la même manière, les multiples partitionnements disponibles dans H.264/AVC permettent une prédiction plus efficace par rapport au MPEG-2 pour lequel il faut de plus noter la présence de blocs Intra particulièrement coûteux. Globalement, nous retiendrons que la qualité de la prédiction a effectivement progressé entre ces deux générations de standards mais il demeure des résiduels dans certaines zones complexes. L'exemple proposé ici correspond à un type d'image où les résiduels de texture sont les plus faibles, la comparaison entre d'autres types d'images (I et P) et par conséquent moins nette. Nous devinons néanmoins que les standards émergents auront davantage de difficultés pour améliorer la qualité de la prédiction qui est déjà particulièrement efficace avec les outils actuels, en témoigne l'effort particulier qui a été mené sur les filtres post-reconstructions dans le projet HEVC. Comme nous l'avons rapporté dans le chapitre précédent (1.2.2.4), trois filtres sont ainsi disponibles alors qu'il n'y en avait qu'un seul dans H.264/AVC et aucun dans MPEG-2. Ces filtres sont responsables d'une part très significative du gain total de HEVC car l'unique amélioration de la prédiction ne suffit plus pour faire progresser les performances de compression.

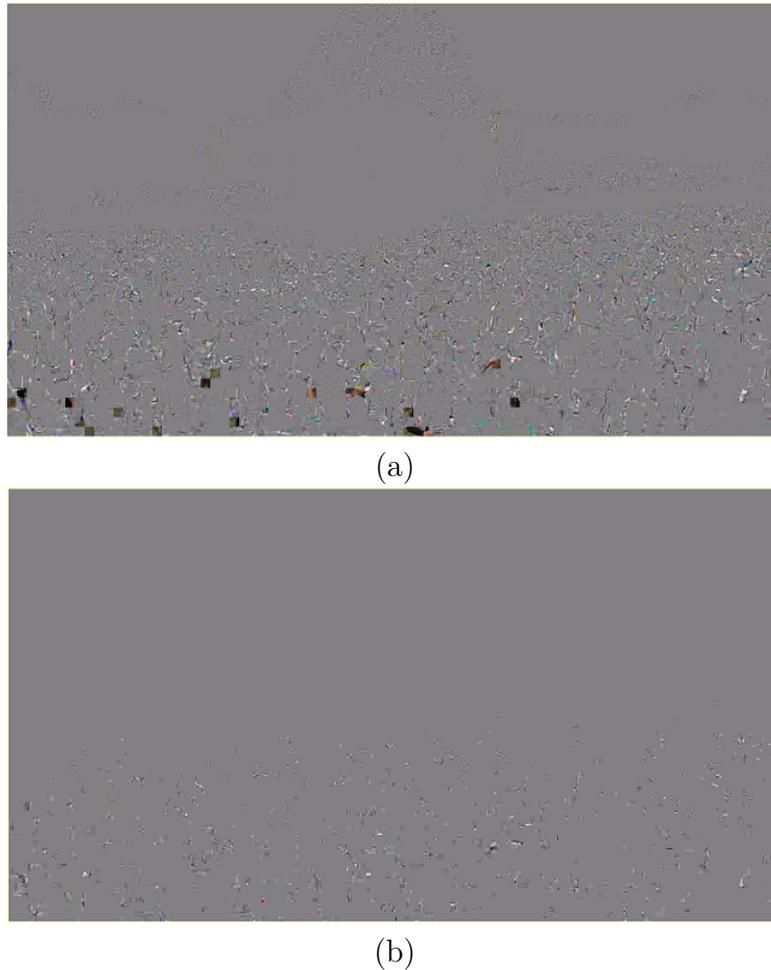


FIGURE 2.11: Résiduels d'une image B de la séquences Crowdrun 720p codée avec les encodeurs MPEG-2 (a) et H.264/AVC (b) pour un même niveau de qualité (30dB).

Pour juger de cette évolution avec le futur standard HEVC, nous avons représenté sur les figures 2.12 et 2.13 la répartition des principales données d'un flux HEVC pour les images de type Intra et Inter respectivement. Ces figures sont à mettre en regard de celles de la sous-section 2.1.1 obtenues avec H.264/AVC pour les mêmes séquences. Tout d'abord, nous observons sur la figure (a) que les résiduels de texture pour les images Intra conservent une proportion significative qui décroît moins rapidement avec le débit en comparaison des résiduels dans H.264/AVC (figure 2.2). Ainsi, la prédiction Intra constitue le type de prédiction où il y a le plus à gagner au vue de la taille particulièrement élevée des images Intra mais on observe néanmoins que les gains en efficacité de codage y sont plus faibles. Les auteurs de [Li *et al.* 12] rapporte en effet des gains de 23% pour HEVC comparé à H.264/AVC en profil Intra alors que 33% et 41% sont rapportés pour le même test avec des configuration Inter plus ou moins

complexes. Enfin, la figure 2.13 correspond au cas Inter. De façon un peu moins nette, nous observons que la proportion des résiduels de mouvement décroît avec le débit de façon moins rapide que pour H.264/AVC (figure 2.3 et 2.4).

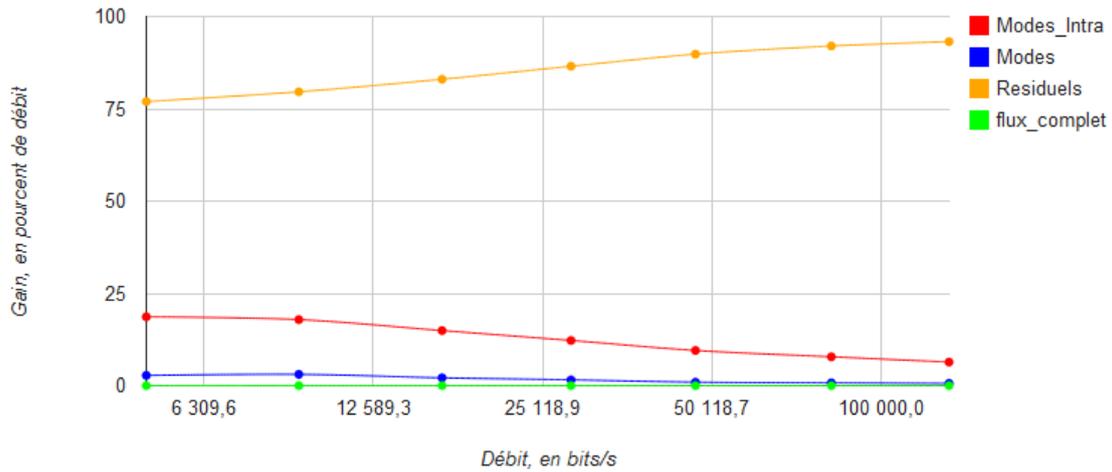


FIGURE 2.12: Pourcentages moyens du débit des principales données de codage de l'encodeur HEVC par rapport au flux complet au sein des images Intra à différents débits.

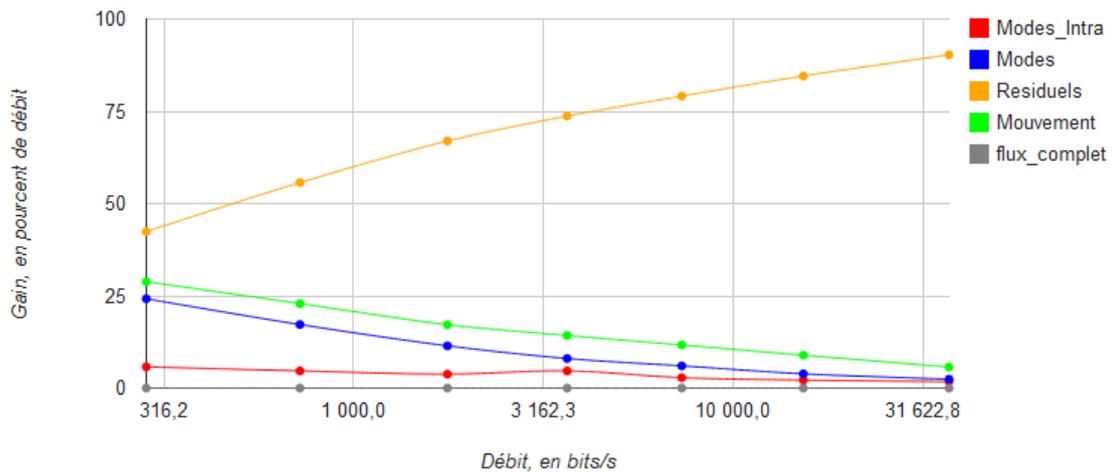


FIGURE 2.13: Pourcentages moyens du débit des principales données de codage de l'encodeur HEVC par rapport au flux complet au sein des images Inter à différents débits.

La qualité de la prédiction a significativement progressé au cours des deux dernières décennies pour permettre d'augmenter le ratio de compression. Cette efficacité élevée des outils existants rend cependant l'émergence de nouvelles techniques plus délicate, en particulier en Intra où de nombreuses redondances sont déjà exploitées ou dans les images Inter B où les résiduels sont d'ores et déjà très faibles. Comme nous l'étudions ci-dessous, cette limite est par ailleurs liée au coût de signalisation de techniques émergentes qui nécessitent une augmentation de la compétition.

Evolution des choix de codage en compétition

Finalement, nous souhaitons évaluer l'impact de l'augmentation des choix en compétition sur l'efficacité de codage. Notre description du chapitre précédent a en effet mis en lumière la forte évolution de la compétition depuis le standard MPEG-2 jusqu'au futur standard HEVC. Cette compétition combinée à une amélioration minutieuse des divers modules des encodeurs a conduit aux gains présentés précédemment mais cette approche pourra-t-elle être de nouveau reproduite pour les prochaines générations d'encodeurs ? Afin de mesurer cette évolution, nous avons fait varier le nombre de paramètres en compétition pour différents outils de codage de l'encodeur HEVC.

Le premier test concerne les modes de prédiction Intra qui sont au nombre de 35 dans HEVC. Nous avons comparé plusieurs ensembles de prédicteurs en compétition en prenant soin de conserver les prédicteurs les plus significatifs et d'adapter la signalisation. Quatre configurations ont ainsi été testées avec respectivement un seul prédicteur (DC), neuf, dix-huit et trente-cinq prédicteurs. Notons que les deux premières configurations et la dernière simulent respectivement les modes des standards MPEG-2, H.264/AVC et HEVC. Le test a été mené en codage Intra uniquement et pour une séquence 720p composée de plusieurs images uniques issues de séquences classiques, la séquence a ensuite été encodée pour les pas de quantification $QP = \{17, 22, 27, 32, 37, 42\}$ et nous avons mesuré à l'aide de la métrique de Bjontegaard le gain sur le débit à différents niveaux de débit par rapport au test avec un prédicteur unique. Les résultats sont présentés sur la figure 2.14. Nous observons tout d'abord que l'augmentation du nombre de prédicteurs en compétition est efficace puisqu'elle améliore significativement les performances par rapport au test avec un unique mode. Les gains vont ainsi de 5% en moyenne pour le test avec neuf prédicteurs à 8% en moyenne avec les trente-cinq prédicteurs. Ces gains impliquent une amélioration de la prédiction réellement significative en considérant que le premier test ne nécessite pas de signalisation alors que dans la configuration correspondant à HEVC, jusqu'à six bins sont nécessaires pour signaler le choix optimal parmi trente-cinq. Comparées au coût de signalisation, la qualité de la distorsion et la diminution de la quantité de résiduels résultants sont ainsi suffisamment fortes au sens du compromis débit-distorsion. Cependant, il est important de relever la nette atténuation des gains avec l'augmentation de la compétition. En effet, alors que les neuf prédicteurs apportaient un gain de 6%, la configuration avec dix-huit prédicteurs est seulement 1.5% au-dessus de la précédente tandis que la dernière configuration

n'apporte qu'environ 0.5%. Notons d'ailleurs que dans un souci de limitation de la complexité, il est probable que certaines implémentations d'encodeurs HEVC n'exploiteront pas l'ensemble des directions de prédiction et seront par conséquent sous-optimales dans la mesure où la signalisation sera toujours réalisée pour l'ensemble des choix. Bien que le nombre de modes soit multiplié par deux entre chacune des trois dernières configurations, ces résultats montrent donc qu'il sera délicat d'obtenir un nouveau gain significatif en multipliant de nouveau par deux le nombre de prédicteurs *directionnels* sans modifications structurelles supplémentaires. Bien que l'ajout de prédicteurs différents des prédicteurs directionnels pourrait aussi permettre un nouveau progrès, ce type de prédicteurs nécessitent généralement une signalisation additionnelle qu'il sera aussi délicat de compenser. Un tel constat a été fait pour de nombreux prédicteurs Intra alternatifs proposés récemment et nous observons que le codeur HEVC n'inclut qu'un seul nouveau prédicteur « original » : le Planar (1.2.2.2). Cet outil, dans sa configuration initiale, nécessitait d'ailleurs la transmission d'informations supplémentaires qui ont depuis été supprimées pour rendre le mode efficace du point de vue du compromis débit-distorsion malgré une réduction de la qualité de reconstruction par rapport à la proposition initiale.

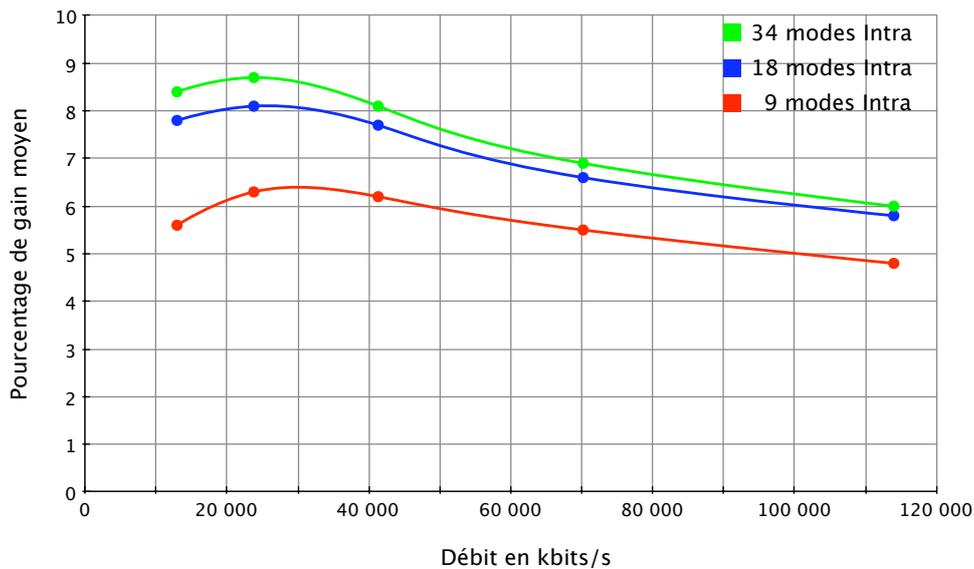


FIGURE 2.14: Pourcentage de gains sur le débit à différents niveaux de débit pour plusieurs nombre de prédicteurs Intra en compétition (9, 18 et 35) par rapport à un prédicteur unique (testés dans HEVC).

De la même manière, nous présentons un second test réalisé en Inter où nous avons fait varier le nombre d'images de référence en compétition : une seule référence, deux et quatre références ont été testées. Nous observons que le passage d'une seule référence à deux apporte un gain de 2% en moyenne alors que la multiplication du nombre de références n'apporte que 0.3% de plus environ. Bien que la distance temporelle par

rapport à l'image courante qui augmente avec le nombre de références doit être pris en compte pour interpréter ce résultat, l'impact du coût de signalisation dans la limitation du gain pour le passage à quatre références est réel.

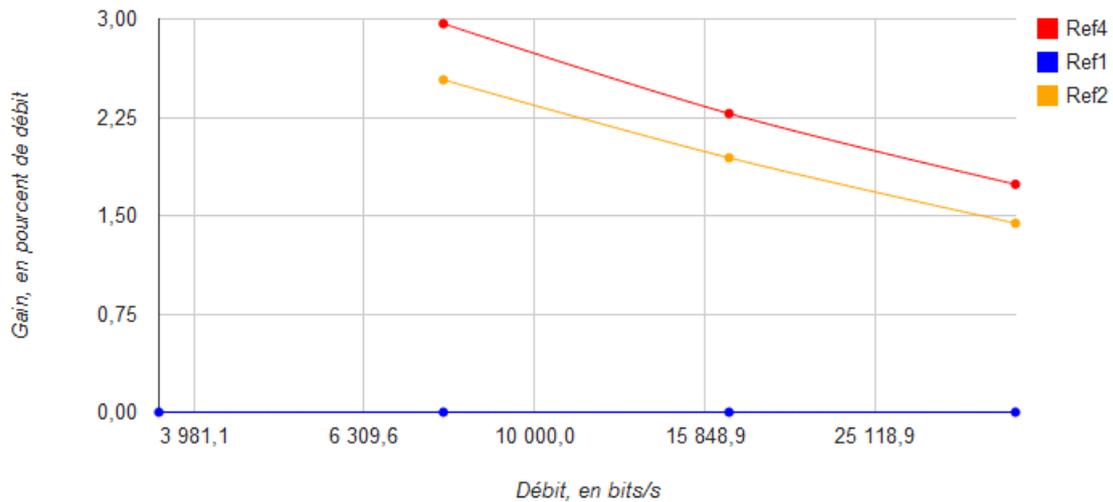


FIGURE 2.15: Pourcentage de gains sur le débit à différents niveaux de débit pour plusieurs nombre d'images de référence en compétition (2 et 4) par rapport à une référence unique (testés dans HEVC).

2.3 Bilan

Dans ce chapitre nous avons tout d'abord étudié le schéma de codage de la norme H.264/AVC. Deux points d'améliorations ont été isolés avec d'une part le sujet de la signalisation et d'autre part celui du codage du mouvement. L'ensemble des informations à transmettre en dehors des résiduels de pixels et de mouvement représente en effet une proportion significative du flux, notamment à bas débit. En particulier, les informations issues de la mise en compétition de choix de codage constitue un enjeu particulier pour l'amélioration de l'efficacité de compression. Nous abordons ce point dans le chapitre 4 où une approche originale pour réduire les coûts de signalisation est proposée. La transmission des vecteurs de mouvement est elle-aussi importante au vue de leur proportion et de leur importance pour la qualité de la prédiction. Le perfectionnement du codage du mouvement impacte sur la qualité de la prédiction de la texture dont il faut par conséquent tenir compte. Une piste de recherche que nous explorons dans le chapitre 5 concerne l'exploitation accrue des redondances à la fois spatiales et temporelles au sein du mouvement.

La dernière section du chapitre, nous a ensuite permis de donner une vue d'ensemble

sur l'évolution des trois générations d'encodeurs standards, MPEG-2, H.264/AVC et HEVC. Dans une première évaluation, nous avons ainsi mesuré les performances des trois encodeurs dont les gains respectifs en efficacité de compression sont effectivement proches des 50% annoncées. L'étude croisée de la qualité de la prédiction et de l'augmentation de la compétition au sein de ces encodeurs nous a finalement montré une limite potentielle de ces schémas. A terme, l'augmentation systématique des choix en compétition semble en effet pouvoir être un frein aux prochaines évolutions. Afin de proposer une solution à ce problème, nous étudions le potentiel d'une approche exploitant davantage le décodeur dans la seconde partie de ce manuscrit.

Première partie

Approches conventionnelles pour les schémas de codage vidéo émergents

Chapitre 3

Signalisation, codage du mouvement et tatouage : un état de l'art

Nos travaux que nous exposons dans cette première partie du manuscrit ont été réalisés dans le cadre du développement de codeurs vidéo émergents tels que le futur standard HEVC. Afin de positionner nos solutions par rapport aux méthodes existantes, nous proposons un état de l'art sur le sujet de la signalisation et du codage du mouvement. Nous avons montré dans le chapitre précédent que la signalisation des choix de codage représente une proportion importante du flux, notamment à bas débit, et que ce constat tend à progresser avec la nouvelle génération d'encodeurs standardisés. Différentes solutions pour améliorer la transmission de ces informations ont été envisagées et sont rapportées dans ce chapitre. Nous étudions ensuite les méthodes développées pour traiter le mouvement qui représente lui aussi un fort enjeu dans les schémas de codage prédictifs.

La dernière section de ce chapitre est quant à elle consacrée à l'introduction au domaine du tatouage numérique et plus particulièrement à ses applications en vidéo. Compression et tatouage sont deux domaines aux buts contradictoires et nous nous appuyons sur ce dernier pour proposer une méthode originale de codage dans le chapitre suivant.

3.1 Etat de l'art sur la signalisation des choix de codage

La signalisation des choix de codage est un véritable enjeu pour les codeurs vidéos émergents où la proportion de cette information augmente significativement à chaque génération d'encodeur en faisant diminuer le coût des résiduels de textures. La transmission est principalement gérée par l'utilisation de codeurs entropiques qui réduisent le coût des informations les plus probables. Le codeur entropique CABAC intégré dans H.264/AVC dispose ainsi de multiples contextes mis à jour au cours du codage dont l'efficacité est significative. Nous décrirons son fonctionnement pour le traitement des vecteurs de mouvement dans le chapitre 5 (5.2.3). Malgré cela, la proportion des informations de signalisations, notamment celles résultantes de la compétition entre choix de

codage, demeure très coûteuse.

Quelques techniques sont mises en œuvres dans H.264/AVC en dehors du codage entropique avec en particulier l'utilisation du MPM (*Most Probable Mode*) [Karczewicz & Lainema 02] pour la transmission des modes de prédiction en Intra 8×8 et 4×4 . Cette méthode utilise le voisinage du bloc courant pour déterminer le mode le plus probable qui peut alors être codé sur un faible nombre de bits. A partir des modes de prédiction des blocs voisins du dessus (A) et de gauche (B), le MPM est dérivé comme suit :

$$MPM(A, B) = \begin{cases} DC & \text{si A ou B indisponibles} \\ \min(pred_A, pred_B) & \text{sinon} \end{cases}, \quad (3.1)$$

où $pred_A$ et $pred_B$ sont les modes de prédiction respectifs des blocs A et B, l'ensemble des modes de prédiction Intra ayant été affecté à un indice fixe (1 : Horizontal, 2 : Vertical, 3 : DC, ...).

Un indice est alors dérivé pour signaler si le mode de prédiction du bloc courant correspond au MPM. Celui-ci est alors codé sur un seul symbole binaire le cas échéant tandis qu'il est nécessaire de signaler l'indice du prédicteur parmi huit modes (les neuf prédicteurs moins le MPM). Dans le cadre de travaux proposant de nouveaux modes de prédiction Intra, l'auteur de [Moinard 11] a par ailleurs introduit une modification de ce MPM permettant de signaler efficacement un prédicteur parmi dix : un second MPM est utilisé. Le futur standard HEVC pour lequel le nombre de prédicteurs à significativement augmenté utilise une méthode de ce type avec jusqu'à trois niveaux de MPM.

Au-delà de cette solution qui constitue en réalité une prédiction des moyens de prédiction, nous pouvons citer les quelques travaux consistant à exploiter davantage le décodeur pour réduire la quantité d'informations à transmettre à l'encodage. Un état de l'art exhaustif sur ce sujet sera proposé dans le chapitre 6. Peu d'autres techniques originales pour réduire le coût de signalisation ont été proposées. Citons les approches cherchant à regrouper des informations à transmettre telle [Shukla *et al.* 05] qui analyse notamment un défaut lié à la représentation en arbre quaternaire pour le codage d'image 2D : la décomposition systématique d'un bloc en quatre sous-blocs empêche deux sous-blocs ayant des paramètres identiques d'être représentés ensemble. Un algorithme de fusion de sous-blocs partageant les mêmes données est alors proposé et permet d'améliorer les performances de codage. C'est cette idée qui est reprise dans le mode Merge introduit dans HEVC. Plus généralement, la technique consistant à regrouper des paramètres pour un ensemble donné peut intervenir au niveau slice ou au niveau image. C'est notamment le cas pour certains paramètres liés aux filtres.

3.2 Etat de l'art sur le codage du mouvement

L'étude des schémas de codage vidéo dans le chapitre 2 nous a enseigné que la proportion de l'information de mouvement était importante puisqu'elle représente un tiers du flux environ. La réduction du coût de codage des vecteurs de mouvement et plus généralement de l'information de mouvement est par conséquent un sujet largement traité dans les travaux récents.

Nous présentons dans cette sous-section une sélection de travaux sur le codage du mouvement dont l'étude nous a aidé à développer notre méthode introduite dans le chapitre 5. Ces travaux sont répartis selon trois catégories traitant respectivement de la prédiction des vecteurs de mouvement, de leur codage et enfin de nouvelles approches notamment basées sur une représentation différente du mouvement.

3.2.1 Prédiction des vecteurs de mouvement

A la manière des résiduels de texture, un codage prédictif des vecteurs de mouvement a été introduit dans les codeurs vidéo afin de réduire l'entropie de l'information à coder. Un résiduel de mouvement ε_{vm} obtenu à l'aide du vecteur prédictif vm_{pred} est ainsi transmis à la place du vecteur vm à coder :

$$\varepsilon_{vm} = vm - vm_{pred}, \quad (3.2)$$

où vm_{pred} est généré par l'opérateur de prédiction $p : X \rightarrow vm_{pred}$, où X représente les informations nécessaires pour déterminer la valeur du prédictif. Typiquement, ces informations représentent les vecteurs de mouvement des blocs voisins pour le médian spatial de H.264/AVC comme nous l'avons défini dans le premier chapitre. Ce prédictif classique exploite les redondances spatiales du mouvement et permet effectivement de réduire le coût de codage des vecteurs dans de nombreux cas. Il n'est cependant pas optimal dans plusieurs autres cas, en particulier en présence d'une rupture du mouvement dans le champ de vecteurs, ce qui laisse par conséquent la place à une amélioration de la prédiction. Après l'étude d'autres prédictifs alternatifs [Ebrahimi 94, Krishnamurthy 97, Yeh *et al.* 95] qui s'avèrent très efficaces pour certains mouvements mais moins performants en moyenne face au médian spatial, la communauté se dirige maintenant vers un système de mise en compétition de plusieurs prédictifs comme en atteste notamment les travaux de standardisation du codeur HEVC.

Une telle idée a été entrevue dans [Chen & Jr 97, Kim & Ra 99, Natale & Granelli 01] avant d'être intégrée dans le codeur JM KTA [Jung & Laroche 06] dans le but d'améliorer la prédiction par médian spatial d'H.264/AVC et donc de servir de base à l'outil AMVP du standard HEVC en cours d'élaboration. Ainsi, l'outil de compétition des vecteurs de mouvement prédictifs (MV-Comp) tel que proposé dans [Laroche *et al.* 08], spécifie un cadre pour la compétition des prédictifs du mode SKIP (défini dans le chapitre 1) et des vecteurs de mouvement prédictifs. L'outil sélectionne de façon op-

timale le vecteur prédicteur à l'aide d'un critère débit-distorsion et tire profit des redondances à la fois spatiales et temporelles des champs de mouvement en utilisant plusieurs prédicteurs en compétition. Ces vecteurs de mouvement prédicteurs sont notés $p_i, \forall i \in \{1, \dots, n\}$, où p_i est l'opérateur de prédiction qui génère le vecteur vm_{pred_i} .

Ce schéma nécessite alors la transmission d'indices de signalisation du prédicteur sélectionné. Le débit R_{vm_i} généré par le codage de l'information de mouvement du vecteur de mouvement vm à l'aide du prédicteur $p_i, \forall i \in \{1, \dots, n\}$ est ainsi exprimé par :

$$R_{vm_i} = \varsigma(\varepsilon_{vm_i}) + \varsigma(i), \quad (3.3)$$

où $\varsigma(x)$ est le coût en débit de la donnée x dans le flux, i est l'indice du prédicteur p_i , et ε_{vm_i} le vecteur de mouvement résiduel à coder, défini par :

$$\varepsilon_{vm_i} = vm - vm_{pred_i}. \quad (3.4)$$

L'indice du prédicteur sélectionné i^* satisfait :

$$i^* = \operatorname{argmin} \{D_i + \lambda_{vm} R_{vm_i}\}_{i \in \{1, \dots, n\}}, \quad (3.5)$$

où D_i est la distorsion générée et λ_{vm} , le multiplicateur de Lagrange de l'estimation de mouvement.

Notons enfin qu'en cas d'égalité des prédicteurs p_i , l'indice n'est pas transmis : $\varsigma(i) = 0, \forall i \in \{1, \dots, n\}$. En pratique ce cas de figure arrive dans environ 20% des cas pour une compétition entre le médian spatial et le vecteur du bloc colocalisé temporellement. Dans cette configuration qui permet d'habilement exploiter les redondances spatiales et temporelles du mouvement, les auteurs font état d'un gain moyen de 4.3% comparé à H.264/AVC dans le profile High sous les conditions de test définies par le groupe VCEG [T. K. Tan & Wedi 07]. Il s'agit d'un gain significatif relativement à la faible complexité ajoutée mais la méthode est néanmoins limitée du fait de la signalisation additionnelle et d'un problème de lecture des données (*parsing*) qui anime les activités du groupe JCT-VC [Jung & Bross 11].

A partir de l'environnement défini par MV-Comp, de nombreuses contributions à la fois académiques et de standardisation ont traité le sujet de l'amélioration du codage des vecteurs de mouvement par une meilleure prédiction des vecteurs. Les principaux travaux sur le sujet sont introduits ci-dessous.

Dans [Dai *et al.* 09a], un ensemble de prédicteurs candidats est tout d'abord déterminé adaptativement à partir des caractéristiques des vecteurs de mouvement des blocs voisins, le prédicteur optimal est alors sélectionné dans cet ensemble de manière à minimiser le nombre de bits utilisés lors du codage du résiduel de mouvement. L'indice de signalisation du prédicteur sélectionné est finalement transmis de façon optimale selon les règles définies dans [Kim & Ra 99] et un critère basé sur une estimation de mouve-

ment aux frontières reproductible au décodeur permet de réduire le coût de signalisation du prédicteur. Un schéma similaire est proposé dans [Won *et al.* 09] où le concept de transfert de l'intelligence au décodeur est ici exploité au maximum : aucun indice n'est transmis au décodeur qui retrouve lui-même le prédicteur sélectionné. Cette technique améliore légèrement l'efficacité de MV-Comp mais est limitée par une moins bonne sélection des prédicteurs optimaux.

3.2.2 Codage du résiduel de mouvement

La prédiction des vecteurs demeure le sujet de recherche le plus abordé par rapport au mouvement mais certains travaux se sont aussi concentrés sur le codage du résiduel ou plus largement de l'information de mouvement. En effet, même avec un outil de prédiction des vecteurs très efficace, il subsiste des vecteurs coûteux en présence de mouvements irréguliers ou très complexes.

Parmi ces travaux, citons tout d'abord ceux abordant le sujet du codage du vecteur de mouvement avec pertes. A la manière des résiduels de texture, les vecteurs peuvent ainsi être quantifiés avant d'être transmis au décodeur. De tels travaux datant notamment du milieu des années 1990 ([Lee & Woods 95, Joshi *et al.* 95]) considèrent le vecteur de mouvement comme une information quantifiable et lui applique par conséquent une quantification vectorielle (QV). Cette quantification étant appliquée en boucle fermée, ces schémas sont limités par la qualité moindre des résiduels de texture calculés par compensation de mouvement à l'aide du vecteur quantifié. Plus récemment, les auteurs de [Agostini-Vautard *et al.* 11] ont introduit un nouveau mode de codage Inter utilisant une quantification des vecteurs de mouvement dans le standard H.264/AVC avec une différence majeure par rapport aux précédentes études : les résiduels de texture sont déterminés à partir des vecteurs non quantifiés. La quantification du mouvement vm y est définie par :

$$Q(vm, Q_v) = \left[\frac{vm}{Q_v} \right], \quad (3.6)$$

où $[.]$ est la partie entière.

Mis en compétition avec les modes classiques d'H.264/AVC, ce mode essaye alors d'optimiser la distribution des données à coder entre la texture et les résiduels de mouvement. Le schéma apporte des améliorations encourageantes mais il est cependant limité par la sélection du pas de quantification optimal Q_v^* et le problème de signalisation associé, en particulier lorsque les tailles de partitions plus petites (4×4) sont activées et augmentent par conséquent la quantité d'informations de mouvement.

Un second thème de recherche se rapprochant du précédent a attiré l'attention de la communauté, il s'agit de la notion de résolution adaptative. En s'appuyant sur des travaux précédents, les auteurs de [Guo *et al.* 10] ont ainsi mis en avant qu'une résolution différente pouvait être utilisée de façon adaptative pour coder chaque vecteur. En effet, il

peut par exemple être intéressant de rester à une résolution entière et économiser les bits correspondants pour certains mouvements qui ne requièrent pas une précision trop élevée. Plutôt que de signaler cette résolution à l'aide d'un indice supplémentaire qui impacte sur l'efficacité de la méthode comme dans les travaux précédents, la résolution est dérivée au niveau du décodeur à partir de la variance du bloc résiduel. Des gains intéressants sont obtenus comparé à H.264/AVC lorsque la méthode est appliquée uniquement sur des tailles de partitions 16×16 , en particulier pour des bas débits avec un gain moyen de 3.80%. Ce sujet, de résolution adaptative des vecteurs de mouvement, a eu un écho dans les activités actuelles de standardisation pour HEVC où un groupe de travail a étudié différentes contributions. Les résultats sont résumés dans le document [Chien 11] et font état de gains supérieurs à 1%.

Enfin, un peu à la marge des études précédentes, citons le travail [Takahashi & Yamaguchi 09] où les auteurs proposent deux améliorations du codage du mouvement. Ils traitent tout d'abord de façon relativement classique le problème de la prédiction en codant adaptativement l'indice de signalisation du prédicteur optimal selon la distribution des différentes composantes des vecteurs du voisinage. Ils proposent ensuite une amélioration intéressante du codage des résiduels de mouvement : après avoir catégorisé chaque bloc selon un pronostic de l'amplitude de ce résiduel, une table de codage des vecteurs résiduels est sélectionnée adaptativement parmi un ensemble de trois tables prédéfinies afin de coder les vecteurs à moindre coût. La catégorisation du bloc courant est réalisée en fonction du vecteur prédicteur p_i sélectionné. Typiquement, les auteurs classifient le bloc courant comme faisant partie du même objet que son voisinage si le prédicteur p_i définie par la première étape possède au moins l'une de ses composantes qui est le médian des composantes des vecteurs de son voisinage, la table correspondante est alors adaptée pour le codage de faibles résiduels ε_{vm} : les mots de codes les plus courts sont affectés aux faibles résiduels. De la même manière, deux autres groupes de blocs sont définis et permettent notamment d'associer les plus faibles mots de codes aux résiduels de mouvement potentiellement les plus élevés pour les blocs classifiés parmi les blocs peu corrélés à leur voisinage. Cette approche semble donc prometteuse dans la mesure où la sélection des tables a pour but de s'adapter aux types de résiduels, faibles ou élevés (en d'autres termes, bien ou mal prédits). Des résultats encourageants sont ainsi obtenus en comparaison avec le JM KTA version 1.8 qui inclut l'outil MV-Comp (défini dans 3.2.1), la seconde amélioration apportant en moyenne 0.6% de gains sur le débit dans le profil Baseline. Il serait cependant intéressant de connaître les performances de la méthode dans une configuration incluant le codeur entropique CABAC avec lequel la solution proposée semble sensiblement redondante. Enfin, l'étape de classification des blocs demeure délicate et peut entraîner un surcoût de codage en cas de mauvaise classification d'un bloc.

3.2.3 Autres approches pour le codage du mouvement

Finalement, nous introduisons ici certains travaux en rupture avec les approches conventionnelles. En effet, s'il semble acquis que l'exploitation des redondances temporelles via l'utilisation d'une compensation de mouvement est une brique indispensable des codeurs vidéos actuels la représentation du mouvement et son codage constituent pour leur part une source importante de travaux de recherches davantage en rupture.

Notons que nous n'aborderons pas ici les techniques consistants à exploiter davantage le décodeur en réalisant une estimation de mouvement conjointe au codeur et au décodeur permettant d'économiser la signalisation du vecteur. Ce sujet sera en effet abordé dans le chapitre d'état de l'art de la seconde partie de cette thèse.

Tout d'abord, mentionnons les travaux sur le mode *Merge* que nous avons déjà évoqué lors de la présentation des principaux outils d'HEVC dans le chapitre 1. Dans [Marpe *et al.* 10], les auteurs introduisent cette notion parmi d'autres outils constituant une des réponses de l'appel à propositions HEVC. Inspiré de précédents travaux, notamment [Shukla *et al.* 05], cette approche vise à s'adapter à la structure de partitionnement des blocs en quadtree qui peut donner lieu à une sur-segmentation et elle permet ainsi d'hériter l'ensemble des paramètres de mouvement (vecteur et image de référence) au prix d'une signalisation du bloc voisin optimale à la manière de MV-Comp. Le Merge généralise en ce sens les modes Skip et Direct du standard H.264/AVC et apporte un gain significatif en économisant le coût de ces paramètres pour des zones voisines appartenant à un même ensemble de mouvement. Sélectionné dans 40% des cas en moyenne, ce mode confirme que l'exploitation des fortes corrélations dans le mouvement tant spatiales que temporelles est une source d'amélioration majeure des codeurs vidéo.

Par rapport aux approches proposant une nouvelle représentation du mouvement, citons les récents travaux mis en œuvre pour améliorer le standard H.264/AVC et basés sur une représentation paramétrique [Lakshman *et al.* 10a, Lakshman *et al.* 10b]. Un modèle paramétrique du mouvement permet en effet de représenter plus efficacement les mouvements complexes tandis que les paramètres additionnels engendrent certes un surcoût de codage mais offrent aussi davantage de degrés de libertés pour l'optimisation débit-distorsion. Inspirés de précédentes contributions [Karczewicz *et al.* 97, Wiegand *et al.* 98], les auteurs proposent de mettre en compétition une telle représentation avec le modèle translationnel classique. Au-delà des performances intéressantes obtenues par la méthode qui réduit le débit jusqu'à 12% pour une même qualité sur certaines séquences, cette étude montre que la question du mouvement n'est pas cantonnée au coût de codage des vecteurs. En effet, une meilleure prise en compte de l'ensemble des mouvements d'une séquence a un impact direct sur la prédiction des blocs de texture. On peut observer qu'à l'inverse, dans les codeurs actuels, la combinaison d'un modèle translationnel et d'une prédiction à l'aide d'un ensemble réduit de prédicteurs de mouvement (voir d'un unique prédicteur, dans H.264/AVC) a pour effet de cantonner les mouvements

utilisés aux moins coûteux. Cela peut avoir pour conséquence négative d'augmenter le coût de codage des résiduels de texture lorsqu'un déplacement plus complexe est occulté pour une signalisation moins coûteuse. C'est un point dont nous tiendrons compte dans notre étude présentée dans le chapitre 5.

Finalement, une alternative davantage en rupture en termes de représentation du mouvement est étudiée dans [Urvoy 11] avec l'étude des *tubes de mouvement* qui permettent de représenter la persistance temporelle des textures à travers une description continue le long de l'axe temporel. Malgré différents problèmes de mise en œuvre, cette technique met en avant que les redondances dans le mouvement, plus importantes que celles des blocs uniquement voisins, sont exploitables pour réduire le coût du mouvement.

3.3 Un bref état de l'art sur le tatouage en vidéo

Dans le chapitre suivant, nous introduisons une méthode originale basée sur l'exploitation d'une technique issue du domaine du tatouage vidéo pour améliorer les performances en compression vidéo. Ces domaines du tatouage et de la compression ont traditionnellement deux buts contradictoires : lorsque le premier ajoute des informations de façon non-perceptible afin de masquer des données, le second supprime les informations redondantes afin de réduire la taille du média numérique considéré.

Afin de donner les principales notions nécessaires à notre étude du chapitre 4, nous présentons dans cette section une vue d'ensemble du tatouage, notamment en vidéo, puis nous étudierons les liens avec la compression.

3.3.1 Tatouage

Le tatouage traite de l'habilité à masquer des informations dans un média numérique (image, vidéo, son) avec une quantité minimum de dégradation visible. Les données masquées doivent être invisibles pour la perception humaine, elles constituent la *marque*. Il y a de nombreuses applications du tatouage : protection par droit d'auteur, authentification basée contenu, fingerprinting ou correction d'erreurs (en anglais *error concealment*). Pour chaque application, un délicat compromis entre trois paramètres principaux doit être trouvé :

Capacité représente la quantité d'informations à masquer, aussi appelée le message.

Invisibilité représente la distorsion induite par la marque, elle doit rester imperceptible par l'observateur humain. Elle est aussi appelée fidélité.

Robustesse représente la résistance de la marque face aux attaques (compression, rognage, ajout de bruit...).

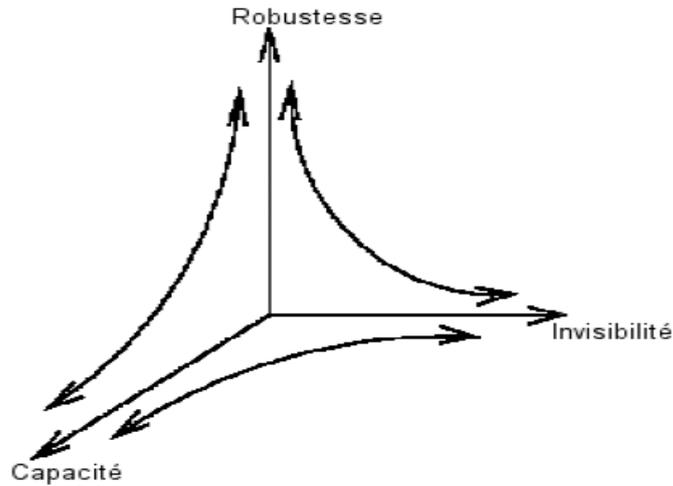


FIGURE 3.1: Problématique du tatouage.

Cette problématique du tatouage est très bien présentée dans les travaux de thèse de Yann Bodo [Yann 04] dont est tirée la figure 3.1.

Parmi les récentes approches de tatouage appliquées à un signal vidéo, nous distinguons trois principales catégories détaillées dans cette section.

La première classe d'approches étend le tatouage d'image fixe en masquant le message dans chaque image indépendamment de la dimension temporelle [Dittmann *et al.* 98]. La méthode d'étalement de spectre peut être ainsi utilisée : le message est réparti parmi un large rang de fréquences du signal de données hôte. Cependant, dans la mesure où cela permet de réduire les dégradations visibles pour la perception humaine, le domaine transformé est généralement préféré. Les domaines transformés tels que le domaine ondelette ou DCT sont ainsi couramment considérés.

La seconde classe de méthodes exploite la dimension temporelle du signal vidéo dans le but d'améliorer les critères d'invisibilité et de robustesse. L'un des premiers travaux à explorer le sujet [Hartung & Girod 98] consiste en une extension de la théorie de l'étalement de spectre aux vidéos par l'utilisation d'un masque à la fois spatial et temporel dans le but d'ajuster la force de la marque dans les différentes parties de la vidéo. Parmi les autres approches, citons l'utilisation d'une décomposition temporelle en ondelettes dans le but de séparer les composantes statique et dynamique d'une vidéo [Swanson *et al.* 98] et une autre contribution [Deguillaume *et al.* 99] qui considère le signal vidéo comme un signal en trois dimensions.

Enfin, la dernière classe d'approches pour le tatouage de vidéo est basée sur les caractéristiques du standard de compression vidéo lui-même. En effet, d'une part les vidéos sont classiquement stockées dans une version compressée et d'autre part le traitement de compression peut être considéré comme une attaque qui pourrait dégrader la marque

si le tatouage est réalisé avant compression. C'est pourquoi il y a de très nombreuses contributions qui exploitent le standard de compression pour réaliser le tatouage. En particulier, le sujet le plus étudié est le masquage d'informations dans les coefficients résiduels transformés par DCT non-nuls d'un flux vidéo. Citons notamment la technique du *Force Even Watermarking* (FEW) proposée dans [Nemethova *et al.* 06] où elle est utilisée pour la détection d'erreurs. Il s'agit d'un tatouage dit fragile qui masque les informations par modification de la parité des coefficients AC. Cependant cette méthode peut augmenter le débit, c'est pourquoi d'autres contributions ont étudié en détail le traitement des codeurs entropiques CAVLC et CABAC du codeur H.264/AVC : dans [Kim *et al.* 07], les auteurs proposent de masquer un bit de la marque dans le signe des Trailing Ones. Ainsi, le débit n'est pas modifié par le tatouage mais cela engendre par ailleurs une dégradation visuelle qui est diffusée par la propagation d'erreur due à la prédiction. Dans le but de préserver la qualité visuelle, le schéma proposé dans [Marvel *et al.* 00] est basé sur la transmission de données par le masquage de chaque bit dans la parité de la somme des coefficients transformés. Cette approche astucieuse permet de sélectionner la modification qui implique la plus faible dégradation de la qualité visuelle, elle est cependant limitée en terme de quantité d'informations à masquer. D'autres contributions de cette classe d'approches ont aussi exploré le tatouage dans les vecteurs de mouvement. Cela a été proposé initialement dans [Kutter *et al.* 97] en modifiant la parité des vecteurs de mouvement afin de masquer les informations. Diverses améliorations de cette méthode ont par la suite été proposées dans le but de diminuer l'impact de la marque sur la vidéo [Kuo *et al.* 08]. Finalement, notons que l'ensemble de ces méthodes de tatouage peuvent être adaptées selon le type d'image (Inter ou Intra) comme proposé dans [Su *et al.* 08].

Plusieurs applications du tatouage diffèrent du but initial de protection par droit d'auteur. Parmi elles, seulement quelques-unes proposent d'utiliser le tatouage dans un but de compression comme proposé dans notre approche. La section suivante introduit ces méthodes.

3.3.2 Tatouage et compression

La notion de tatouage pour la compression (*compressive data hiding* en anglais) a été formalisée dans [Campisi *et al.* 02]. Les auteurs proposent d'exploiter le tatouage afin d'améliorer l'efficacité de codage : les informations de chrominance sont masquées dans le domaine transformé en ondelettes de la composante de luminance avant compression, comme représenté sur la figure 3.2. Cette technique peut être vue comme un pré-traitement dans le but d'améliorer les performances de compression. Ce schéma a été amélioré dans [Giannoula & Hatzinakos 04] par l'addition du masquage de la composante audio dans la composante de luminance. Un schéma similaire est aussi proposé dans [Chaumont & Puech 06] en utilisant la DCT à la place de la transformée en on-

delettes afin de cacher les informations de couleur dans une image JPEG en niveau de gris.

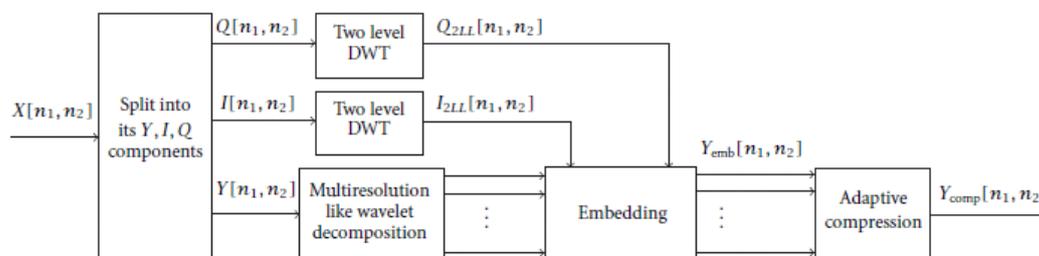


FIGURE 3.2: Schéma de compression basé sur le tatouage (extrait de [Campisi *et al.* 02]).

Ce concept a été présenté dans les travaux plus anciens comme [Swanson *et al.* 97, Zhu & Tewfik 97]. Dans ces articles, le signal à compresser est séparé en une partie d'image hôte et une partie d'image résiduelle. La partie résiduelle est compressée puis masquée par tatouage dans la partie d'image hôte. Finalement, uniquement la partie d'image hôte est compressée. Cette idée est appliquée pour le masquage de la composante audio d'une vidéo dans sa partie visuelle.

Ces schémas confirment que l'utilisation de méthodes de tatouage pour la compression peut être intéressante. Ces approches théoriques demeurent néanmoins difficilement applicables dans le cadre strict de la normalisation. Dans le chapitre suivant, nous introduirons notre travail par une étude des principales techniques de tatouage et une discussion sur leur pertinence dans le cadre d'un standard de compression vidéo tel que H.264/AVC.

3.4 Bilan

Dans ce chapitre, nous avons proposé trois états de l'art servant de base aux travaux présentés dans les chapitres de cette première partie du manuscrit. Concernant, la transmission des informations de signalisation et plus particulièrement des choix en compétition, nous nous sommes concentrés ici sur les méthodes n'impactant pas le décodeur de façon significative : peu de solutions de ce type sont disponibles au-delà de l'amélioration des techniques de codage entropique.

Le sujet du codage des vecteurs de mouvement est davantage abordé. Nous avons en particulier étudié les méthodes basées sur une compétition de prédicteur de mouvement comme cela est mis en œuvre dans le futur standard HEVC. Ces schémas ont l'avantage d'exploiter plusieurs corrélations dans le mouvement mais sont limités par la signalisation générée. Nous proposons dans le chapitre 5 une méthode originale en nous détachant de cette contrainte et en exploitant davantage d'informations de mouvement.

Enfin, en écho à l'état de l'art sur la signalisation, nous avons introduit brièvement le tatouage numérique en vidéo. Le tatouage de manière générale consiste à insérer une marque au sein d'un média numérique de manière invisible. En vidéo, ces techniques tirent habilement profit des standards de compression pour masquer un nombre élevé d'informations. Diverses applications du tatouage existent dont quelques travaux utilisant cette approche dans un but de compression en masquant notamment les données de chrominance dans celles de luminance. En s'inspirant de cette approche, nous proposons dans le chapitre suivant une solution de tatouage intégrée au sein d'un encodeur H.264/AVC et permettant de transmettre des informations de signalisation.

Chapitre 4

Tatouage pour le codage vidéo

Dans les chapitres précédents nous avons mis en lumière un problème classique des codeurs vidéo modernes : le coût de signalisation des paramètres de codage limite l'apport de méthodes pourtant efficaces pour le codage de la texture. Dans le contexte des activités de standardisation il est important d'apporter une solution à ce problème peu traité dans l'état de l'art.

Dans ce chapitre, nous proposons de réduire ce coût de signalisation en utilisant une technique issue du tatouage d'image. Tatouage et compression ayant traditionnellement des buts contradictoires, nous réalisons tout d'abord une étude préliminaire afin de proposer la meilleure solution pour utiliser le tatouage dans un but de compression en s'appuyant sur l'état de l'art proposé dans le chapitre précédent. La méthode est ensuite détaillée avant d'être appliquée à des données issues du codage Inter puis du codage Intra. L'idée principale consiste à masquer des indices de compétition dans les coefficients résiduels de chrominance et de luminance judicieusement sélectionnés. Afin de minimiser l'erreur de prédiction ajoutée par la méthode, la modification des coefficients est réalisée à l'aide d'une optimisation débit-distorsion.

4.1 Choix préliminaires

Dans le but de retenir la meilleure façon d'utiliser le tatouage pour la compression dans le cadre très restrictif des codeurs standardisés nous allons étudier dans cette première section les principales techniques de tatouage introduites dans le chapitre précédent. Les avantages et les inconvénients sont ainsi considérés afin de sélectionner où et comment le tatouage peut être réalisé le plus efficacement.

Dans un second temps, nous précisons les éléments de syntaxe considérés au cours de notre étude à l'aide de l'analyse réalisée dans le chapitre 2.

4.1.1 Technique de tatouage

L'utilisation d'une technique de tatouage dans un but de compression vidéo nécessite de judicieusement sélectionner les données pour réaliser le masquage. En effet, les schémas de compression vidéo sont très poussés, en particulier dans le sens du compromis débit-distorsion. Le cadre de notre étude est donc contraire au cadre des techniques de tatouage utilisées pour des applications conventionnelles où la qualité visuelle ou bien le débit de la séquence peuvent être dans une certaine mesure sacrifiés au profit d'un tatouage plus robuste ou bien d'un tatouage permettant de masquer un nombre plus élevé de données. Dans le cadre de notre travail, le schéma proposé devant être construit pour améliorer les performances de compression, la qualité visuelle et le débit sont par conséquent les paramètres majeurs du schéma. A l'opposé, la contrainte de robustesse n'est clairement pas considérée dans notre schéma. Le problème ainsi posé, nous analysons maintenant les techniques introduites précédemment.

Tout d'abord, notons que les schémas qui s'appliquent avant compression ([Dittmann *et al.* 98, Hartung & Girod 98]) ne sont pas appropriés pour ce travail, en particulier lorsque ces méthodes interviennent en dehors du domaine transformé. En effet, l'information masquée pourrait être affectée par l'étape de compression et rendre le signal non décodable. Les premières et secondes classes d'approches évoquées au chapitre 3 ne sont par conséquent pas envisagées ici.

La troisième classe de techniques de tatouage présentée précédemment apparaît comme la plus intéressante pour notre approche. Ce type de méthode permet en effet de masquer des informations directement dans le flux vidéo compressé tout en tenant compte des caractéristiques du schéma de compression. Cependant, certaines de ces contributions restreignent le procédé de prédiction et ne sont elles aussi pas considérées pour notre travail. Parmi les techniques basées sur une modification des informations de type résiduel, les contributions [Kutter *et al.* 97, Kuo *et al.* 08] qui modifient les vecteurs mouvement avant codage sont intéressantes mais limitées par la faible capacité de masquage. La prédiction du vecteur mouvement dans H.264/AVC, basée sur le median spatial, tend en effet à propager les dégradations résultant du tatouage ce qui rendrait un tel procédé inefficace dans un standard.

C'est donc le tatouage réalisé dans les résiduels de pixels après transformé et quantification qui a été retenu pour notre travail. Cette technique permet en effet de facilement ajuster les paramètres liés au nombre d'informations à masquer et à la fidélité par rapport au signal. Au sein de ces contributions, certaines comme [Nemethova *et al.* 06, Kim *et al.* 07] ne sont pas bien adaptées à notre problème parce qu'elles introduisent une augmentation significative du débit ou bien des détériorations visibles de la qualité de la vidéo. Comme cela a été expliqué précédemment, l'utilisation du tatouage pour le codage vidéo implique des contraintes différentes de celles du tatouage classique. Dans notre cas, une technique de tatouage fixée pour l'ensemble de la séquence à coder (par exemple le masquage dans le bit de poids faible, LSB) n'est pas adaptée du fait des caractéristiques du codeur entropique (communes aux codeurs CAVLC et CABAC), no-

tamment les suites de valeurs égales à un (trailing ones), le codage spécifique des suites de zéros ou bien l'utilisation de contextes particuliers propres au codeur CABAC. Ces spécificités donnent lieu à des variations du coût de codage non-attendues qui peuvent être pénalisantes pour l'efficacité de codage de la séquence. Ajoutons de plus que dans la mesure où le compromis entre le débit et la distorsion est un point clé du problème, la méthode proposée devra idéalement permettre de sélectionner la meilleure modification à l'aide d'une optimisation débit-distorsion. Pour l'ensemble de ces raisons, l'utilisation de la parité de la somme des coefficients résiduels après transformé et quantification, telle que proposée dans [Marvel *et al.* 00], apparaît comme l'approche la mieux appropriée pour notre étude.

Enfin, en accord avec les caractéristiques de perception du système visuel humain, nous proposons d'appliquer le tatouage dans les composantes de chroma (U ou V indifféremment) autant que possible, et dans celle de luma uniquement lorsque cela est nécessaire.

4.1.2 Informations à masquer

Comme expliqué dans le chapitre 3, quelques contributions ont déjà été proposées pour utiliser le tatouage dans un but de compression. Dans les références [Campisi *et al.* 02] et [Giannoula & Hatzinakos 04], les composantes d'audio et de chrominance sont masquées dans la composante de luminance avant compression. Contrairement à ces travaux, le schéma de tatouage proposé dans notre étude est directement introduit pendant le processus de compression. Par conséquent nous nous sommes concentrés sur l'ensemble des indices de compétition générés pour chaque macrobloc tels que le mode, les sous-modes Intra, la taille de la DCT, ou bien d'autres indices de signalisation de codage.

Pour notre étude nous avons choisi de masquer des éléments de syntaxe utilisés pour le codage. Plus précisément, nous nous concentrons sur des indices de signalisations binaires en comptant sur le fait que le ratio entre la proportion de l'indice considéré dans le flux et le nombre de coefficients résiduels disponibles permette de masquer suffisamment d'indices sans générer une dégradation de la prédiction trop élevée. En accord avec l'analyse réalisée dans le chapitre 2, ces indices sont :

- en Inter, l'indice généré par l'outil de compétition des vecteurs mouvement (MV-Comp) qui représente en moyenne 1% du flux en Inter. Précisons à nouveau que le choix de cet indice vient aussi de notre volonté d'améliorer les performances de l'outil MV-Comp.
- en Intra, l'indice *Most Probable Mode* (MPM) qui représente en moyenne 6% du flux Intra.

4.2 Description de la méthode proposée

Nous proposons donc de masquer des indices de signalisations binaires dans les coefficients transformés et quantifiés des résiduels de chrominance et de luminance. Dans le but d'optimiser le compromis à la fois entre la variation du débit et l'impact du tatouage sur la prédiction, la modification des coefficients est réalisée à l'aide d'une optimisation débit-distorsion.

Cette section décrit en détail le schéma proposé.

4.2.1 Schéma global

Notre proposition consiste à utiliser les coefficients transformés et quantifiés des résiduels de chrominance et de luminance afin de masquer des indices de signalisation. Comme cela a été montré dans la section 4.1, le tatouage dans les coefficients transformés est délicat du fait des spécificités du codeur entropique et de la dégradation de la prédiction qui peut se propager aux blocs suivants et aux images suivantes. Afin de résoudre ces problèmes et dans la mesure où il est impossible que la méthode proposée soit systématiquement meilleure que le coût débit-distorsion original du bloc, nous avons défini deux conditions que la méthode de tatouage d'indices doit satisfaire :

1. Contrôle de la variation du débit : cette variation doit être inférieure au coût de l'indice à masquer, soit, $R_k^w - R_k < \zeta(i)$, avec R_k le débit original du $k^{\text{ème}}$ bloc, R_k^w le débit du bloc modifié, $i \in \{0, 1\}$ l'indice de signalisation considéré et $\zeta(i)$ le coût en débit associé.
2. Minimisation de la dégradation de la prédiction : la modification des coefficients transformés doit être invisible et l'impact sur la prédiction des blocs suivants et des images suivantes doit être minimal.

Ces conditions interviennent pour les choix réalisés au niveau de l'encodeur et demeurent invisibles pour le décodeur dans la mesure où tous les choix d'optimisation débit-distorsion sont réalisés au niveau de l'encodeur uniquement. Dans le but de satisfaire ces conditions, nous proposons d'utiliser la parité de la somme des coefficients transformés et quantifiés pour masquer les indices de signalisation. Cette solution permet de tester différentes modifications des coefficients dans le but de sélectionner la façon optimale de modifier les coefficients du bloc considéré lorsque cela est nécessaire. La première condition est respectée en rejetant toutes les modifications qui augmentent le débit d'une valeur supérieure à $\zeta(i)$ tandis qu'une optimisation débit-distorsion est appliquée pour la seconde condition.

Notons a_n les coefficients non-nuls obtenus après transformée et quantification avant transmission, correspondant aux résiduels de chrominance et de luminance du bloc courant, où a_1 correspond à la valeur DC et $n \in \{1, \dots, N\}$. Notons alors S_k la somme des N coefficients non-nuls du $k^{\text{ème}}$ bloc :

$$S_k = \sum_{n=1}^N a_n. \quad (4.1)$$

La somme modifiée après insertion de l'information de codage, notée S_k^w , est obtenue ainsi :

$$S_k^w = \begin{cases} S_k & ; |S_k| \bmod 2 = i \\ S_k + m_k & ; |S_k| \bmod 2 \neq i, \end{cases} \quad (4.2)$$

où i est la valeur de l'indice à transmettre, $i \in \{0, 1\}$ et m_k correspond à la modification des coefficients lorsque cela est nécessaire. Le critère proposé pour déterminer m_k est décrit dans la section 4.2.2.

Dans le but de minimiser l'impact sur la luminance, la composante utilisée pour le tatouage est en priorité la chrominance si des coefficients y sont disponibles, et sinon nous utilisons la luminance. Aussi, afin de limiter la dégradation de la prédiction de chaque composante, les blocs qui contiennent uniquement des coefficients DC ne sont pas utilisés : l'indice de signalisation est alors transmis au décodeur. Dans le cas d'un bloc composé d'un unique coefficient non-nul, nous empêchons toute modification rendant nul ce coefficient afin de se prémunir de la transmission d'un bloc composé de zéros. Cela causerait en effet une erreur de décodage : dans la mesure où aucun coefficient n'y serait disponible, ce bloc ne serait pas considéré comme un bloc contenant un indice masqué au niveau du décodeur tandis que lors de l'encodage un indice aura été transmis à l'aide de notre méthode. Un mauvais décodage du flux serait alors réalisé pour lire l'indice de signalisation.

Il est important de noter que le schéma proposé a aussi la propriété de réduire le coût de codage des coefficients du fait de l'augmentation de la taille des séries de zéro notamment. En effet, comme cela sera analysé dans la section résultats 4.3, le schéma proposé fixe des coefficients à la valeur zéro dans de nombreux cas et le codage de ces coefficients est par conséquent moins coûteux.

Finalement, l'extraction de l'indice au niveau du décodeur est exprimée comme suit :

$$i = |S_k^w| \bmod 2, \quad (4.3)$$

où i représente l'indice extrait du $k^{\text{ème}}$ bloc et S_k^w est la somme des coefficients résiduels décodés a_n s'il y a au moins un coefficient AC différent de zéro.

La valeur de l'indice traité est alors fixé à i . S'il n'y a pas de coefficient AC lu dans le flux, l'indice est décodé de manière classique.

Dans la sous-section suivante, nous décrivons la modification des coefficients transformés au niveau de l'encodeur qui constitue l'étape clé de la méthode.

4.2.2 Modification des coefficients transformés et quantifiés

Nous proposons d'utiliser un critère débit-distorsion pour sélectionner la meilleure modification dans le cas où les coefficients transformés et quantifiés nécessitent d'être modifiés, c'est à dire lorsque :

$$i \neq |S_k| \bmod 2. \quad (4.4)$$

Pour ce faire, nous considérons les N coefficients non-nuls du bloc courant a_n , $a_n \neq 0$ (notons que ce nombre de coefficients N varie d'un bloc à l'autre). Pour chacun des coefficients, 6 couples débit-distorsion $(R_n^{w_j}, D_n^{w_j})$ sont alors calculés après addition d'une valeur impaire m_j :

$$a_n^{w_j} = a_n + m_j. \quad (4.5)$$

Nous avons fait le choix de limiter les valeurs de m_j à $\{-5, -3, -1, 1, 3, 5\}$ car des valeurs plus élevées induisent une modification trop importante du signal original. $R_n^{w_j}$ représente le débit généré par le codage des coefficients modifiés et $D_n^{w_j}$ est la distorsion associée calculée comme la somme des différences au carré (en anglais *sum of squared differences*, SSD) des pixels reconstruits en considérant la modification. Insistons que le fait que les couples $(R_n^{w_j}, D_n^{w_j})$ sont associés à un bloc d'une composante donné (luminance, chrominance U ou V), la distorsion $D_n^{w_j}$ correspond donc à l'une de ces composantes. De plus, notons que les coefficients transformés et quantifiés égaux à zéro ne sont pas considérés afin de ne pas briser les séries de zéro dont le codage est optimisé. En effet, de telles modifications entraînent systématiquement une augmentation du débit $R_n^{w_j}$ combinée, dans la plupart des cas, à une augmentation de la distorsion $D_n^{w_j}$.

Finalement, le problème est de sélectionner le couple (a_n, m_j) correspondant au meilleur compromis débit-distorsion. Nous proposons deux méthodes pour résoudre ce problème.

4.2.2.1 Optimisation débit-distorsion classique

La première méthode est une optimisation débit-distorsion classique avec minimisation du critère suivant :

$$J_n^{w_j} = D_n^{w_j} + \lambda R_n^{w_j}, \quad (4.6)$$

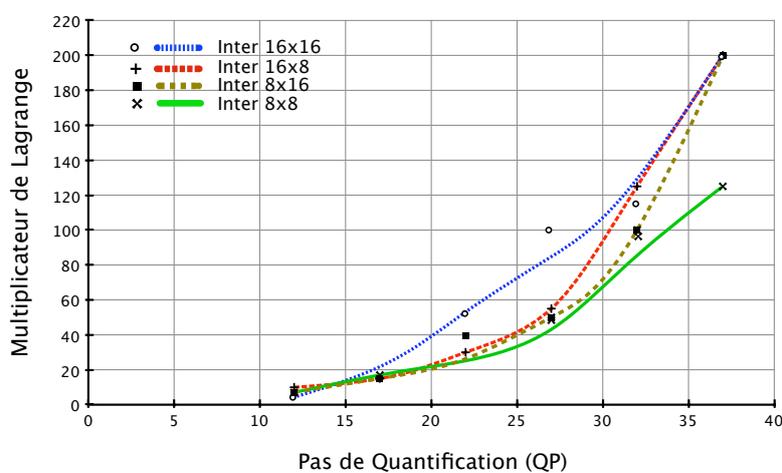
où $(R_n^{w_j}, D_n^{w_j})$ est le couple défini ci-dessus et λ est le multiplicateur de Lagrange.

Nous avons déterminé empiriquement les valeurs des multiplicateurs de Lagrange pour chaque mode et chaque taille de partition à la fois pour les blocs de chrominance et de luminance. En considérant que le format des vidéos n'influence pas les valeurs de λ (observé empiriquement), le protocole de détermination des λ est détaillé ci-dessous :

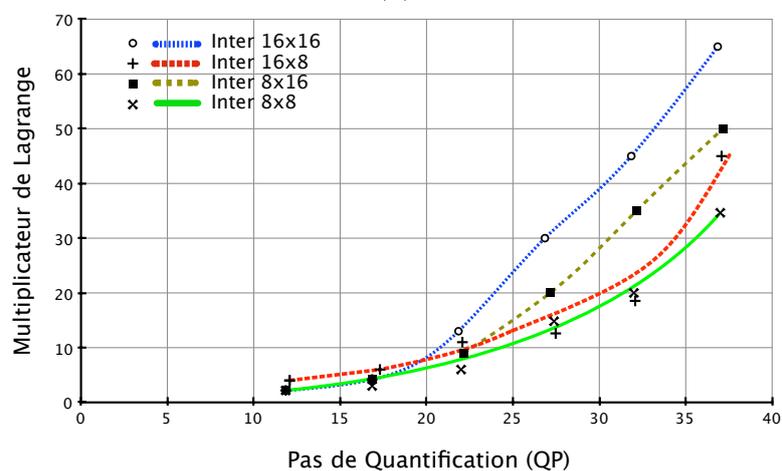
- 5 séquences CIF considérées : Foreman, Mobile, Paris, Tempete, Crew.

- 7 paramètres de quantification (QPs) représentatifs considérés : 12, 17, 22, 27, 32, 37, 42.
- détermination des multiplicateurs de Lagrange optimaux de chacune des 5 séquences pour chaque QP : $\lambda_q(QP)$, $q \in \{1, \dots, 5\}$.
- affectation du médian des valeurs obtenues aux multiplicateurs de Lagrange : $\lambda(QP) = \text{MEDIAN}(\lambda_1(QP), \lambda_2(QP), \lambda_3(QP), \lambda_4(QP), \lambda_5(QP))$.

L'évolution expérimentale des multiplicateurs de Lagrange obtenus est représentée dans les courbes suivantes respectivement pour la chrominance et la luminance :



(a)



(b)

FIGURE 4.1: Valeurs des multiplicateurs de Lagrange déterminées pour chaque QP, pour les modes Inter 16×16 à 8×8 dans le cas d'une modification de la chrominance (a) et de la luminance (b).

Les multiplicateurs de Lagrange obtenus pour chaque mode Inter pour le masquage dans la chrominance et la luminance respectivement ont été déterminés empiriquement à partir de ces courbes. Ainsi, pour le mode Inter 16×16 , les multiplicateurs obtenus sont exprimés comme suit :

$$\lambda_{16 \times 16}(QP) = \gamma \times 2^{\frac{QP-12}{3}},$$

avec :

$$\gamma = \begin{cases} 4.8 & \text{masquage dans la chrominance} \\ 1.4 & \text{masquage dans la luminance} \end{cases}.$$

Des expressions similaires ont été obtenues pour chacune des autres tailles de partitions. Notons que le multiplicateur ainsi déterminé attribue davantage d'importance à la distorsion pour la composante de luminance. En effet, les modifications réalisées dans cette composante sont plus sensibles que celles réalisées dans la chrominance et nécessitent par conséquent d'être réalisées plus subtilement.

Au cours de la détermination expérimentale de ces multiplicateurs de Lagrange, nous avons observé une variance élevée des multiplicateurs optimaux pour chaque séquence. Selon le contenu de la séquence, l'importance de la composante de distorsion dans le critère (4.6) est en effet très variable. Par conséquent, il est difficile de déterminer un multiplicateur optimal parfaitement adapté pour toutes les séquences. Le même comportement est observé au niveau macrobloc.

Suite à cette observation, nous proposons une seconde méthode d'optimisation adaptée pour le schéma de tatouage.

4.2.2.2 Optimisation de Pareto

En alternative nous proposons d'utiliser une méthode de sélection de la meilleure modification basée sur la théorie de l'optimisation multicritère. Cette méthode est adaptative pour chaque bloc. Elle est construite sur la théorie de Pareto [Ehrgott 05] : un couple donné de variables (\bar{a}_n, \bar{m}_j) , est dit Pareto-optimal si et seulement si il n'est pas dominé par un autre couple (a_n, m_j) , où (a_n, m_j) domine (\bar{a}_n, \bar{m}_j) signifie :

$$\text{soit : } \left\{ (R_n^{w_j} \leq \bar{R}_n^{w_j}) \text{ et } (D_n^{w_j} < \bar{D}_n^{w_j}) \right\},$$

$$\text{soit : } \left\{ (R_n^{w_j} < \bar{R}_n^{w_j}) \text{ et } (D_n^{w_j} \leq \bar{D}_n^{w_j}) \right\},$$

où $(\bar{R}_n^{w_j}, \bar{D}_n^{w_j})$ est le couple correspondant aux variables (\bar{a}_n, \bar{m}_j) . Ces couples constituent la frontière de Pareto parmi lesquels la modification optimale va être sélectionnée.

La figure 4.2 montre un exemple de frontière de Pareto. Les points représentent six couples $(R_n^{w_j}, D_n^{w_j})$ générés par chaque modification m_j d'un coefficient transformé non-nul a_n , la courbe rouge constitue la frontière de Pareto qui contient les points Pareto-optimaux.

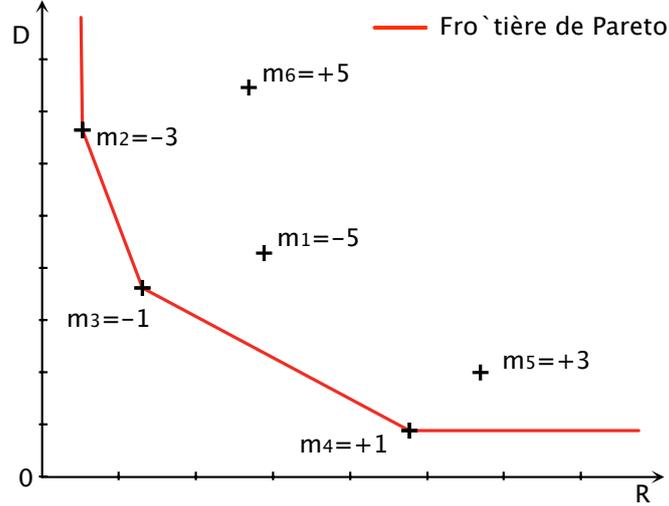


FIGURE 4.2: Frontière de Pareto pour un ensemble donné de couples débit-distorsion.

Nous proposons un schéma en trois étapes qui s'appuie sur cette théorie :

1. Détermination des couples débit-distorsion $(\bar{R}_n^{w_j}, \bar{D}_n^{w_j})$ qui sont Pareto-optimaux.
2. Normalisation de ces couples :

$$\hat{R}_n^{w_j} = \frac{(\bar{R}_n^{w_j} - \bar{R}_{min})}{(\bar{R}_{max} - \bar{R}_{min})}, \quad \hat{D}_n^{w_j} = \frac{(\bar{D}_n^{w_j} - \bar{D}_{min})}{(\bar{D}_{max} - \bar{D}_{min})}$$

où $(\hat{R}_n^{w_j}, \hat{D}_n^{w_j})$ est le couple normalisé, $(\bar{R}_{min}, \bar{R}_{max})$ et $(\bar{D}_{min}, \bar{D}_{max})$ sont les valeurs minimales et maximales du débit et de la distorsion parmi les couples Pareto-optimaux.

3. Sélection du meilleur couple parmi les couples $(\hat{R}_n^{w_j}, \hat{D}_n^{w_j})$ par minimisation de la norme euclidienne :

$$(R_{opt}, D_{opt}) = \min \left\{ \left\| (\hat{R}_n^{w_j}, \hat{D}_n^{w_j}) \right\|_2 \right\}_{n \in [1, N], j \in [1, 6]},$$

où la norme euclidienne du couple (R, D) est définie par :

$$\|(R, D)\|_2 = \sqrt{R^2 + D^2}.$$

Finalement, la modification sélectionnée (a_n, m_j) est celle associée au couple débit-distorsion optimal (R_{opt}, D_{opt}) . Notons que les couples ayant une distorsion trop élevée

sont rejetés par la méthode elle-même afin de ne pas choisir une modification donnant un bloc reconstruit trop différent du bloc original, cela permet de préserver l'efficacité de la prédiction dans la mesure où cette modification a un impact sur la prédiction des blocs suivants. Cette dernière étape de l'optimisation de Pareto affecte le même poids aux composantes de débit et de distorsion afin de préserver l'équilibre, potentiellement optimal, entre ces deux composantes (R_k, D_k) avant l'étape de masquage.

Le principal avantage de ce schéma comparé à l'optimisation débit-distorsion classique est qu'il est indépendant du calcul de multiplicateurs de Lagrange et ainsi adaptatif pour chaque bloc. L'approche par Pareto permet en effet de comparer tous les meilleurs couples avec un ordre de grandeur correct lors de l'étape 3. Cependant, l'inconvénient de ce schéma vient du fait qu'il ne permet pas de tenir compte des paramètres de quantification, contrairement aux multiplicateurs de Lagrange qui sont fonctions du QP. De plus, la complexité de calcul de la méthode est un peu plus élevée pour les hauts débits, mais reste toutefois négligeable pour les moyen et bas débits comme nous le verrons dans la sous-section 4.3.1.1.

4.2.3 Amélioration de la méthode

Un point clé du schéma proposé est de disposer de suffisamment de coefficients transformés et quantifiés pour le masquage afin de plus facilement balancer entre la modification du débit et de la distorsion. Dans ce but, nous proposons de fusionner plusieurs partitions dans le cas où au moins une des partitions ne dispose pas d'indice à masquer. C'est par exemple le cas pour l'indice de compétition sur les prédicteurs de mouvement lorsque ces prédicteurs sont égaux pour une partition P_k et que par conséquent l'indice associé n'est pas transmis au décodeur. Dans ce cas, nous utilisons les coefficients transformés de la partition P_k pour le masquage de l'indice de la partition suivante P_{k+1} . L'indice i' de la $k + 1^{\text{ème}}$ partition est ainsi masqué dans les coefficients transformés des partitions P_k et P_{k+1} . Le même procédé que celui décrit au-dessus est donc appliqué avec davantage de coefficients disponibles.

Au niveau du décodeur, l'extraction de l'indice est exprimée comme suit dans le cas d'une partition courante P_k faisant suite à une partition P_{k-1} sans indice à masquer :

$$i = |S_k + S_{k-1}| \bmod 2,$$

où i représente l'indice de la $k^{\text{ème}}$ partition, S_k et S_{k-1} sont respectivement les sommes des coefficients transformés a_n des partitions P_k et P_{k-1} .

4.2.4 Complexité de la méthode

Nous discutons la complexité de la méthode afin de clore la section description. Pour ce faire, il est important de considérer séparément la partie codeur et la partie décodeur. Le point principal est que la complexité ajoutée au décodeur est très faible et peut être

considérée comme nulle par rapport aux autres étapes réalisées lors du décodage. En effet, le décodeur nécessite uniquement le calcul de la somme des coefficients transformés et quantifiés puis du modulo de cette somme pour chaque partition d'un macrobloc possédant au moins un coefficient non-nul. De plus, ce procédé remplaçant pour ces partitions l'étape de lecture de l'indice dans le flux, le temps de décodage reste stable.

Au niveau de l'encodeur, notons tout d'abord que la complexité ajoutée dépend de choix non-normatif. Par conséquent, le choix de la modification (amplitude et position) des coefficients est libre. Des choix plus ou moins complexes sont ainsi possibles. Dans notre proposition de mise en œuvre, une boucle sur l'ensemble des coefficients est réalisée avec une sous-boucle sur chaque valeur de modification m_j . Pour chaque couple, une évaluation du coût est simulée pour déterminer $R_n^{w_j}$ et la boucle de reconstruction (quantification et transformée inverse) est nécessaire afin de reconstruire le bloc en considérant la modification pour déterminer la distorsion correspondante $D_n^{w_j}$. Une amélioration simple de la rapidité peut alors consister à restreindre l'ensemble des modifications testées de $m_j \in \{-5, -3, -1, 1, 3, 5\}$ à $m_j \in \{-1, 1\}$ par exemple. Nous discuterons plus en détail cette restriction dans la section 4.3.1.1.

La description de la méthode ainsi terminée, nous présentons dans la section suivante les performances obtenues dans le codeur H.264/AVC.

4.3 Résultats expérimentaux

La section résultat est divisée en deux parties : tout d'abord la présentation des résultats obtenus en Inter avec le masquage de l'indice MV-Comp (4.3.1) puis en Intra avec l'indice MPM (4.3.2). Dans les deux cas, une étude objective puis subjective de la méthode est proposée.

4.3.1 Masquage d'informations de mouvement

4.3.1.1 Résultats objectifs

4.3.1.1.1 Configuration de tests

La configuration utilisée pour l'ensemble des tests correspond au profil High d'H.264/AVC avec l'ensemble des outils activés excepté les modes Inter inférieurs à 8×8 (8×4 , 4×8 et 4×4) pour lesquels la méthode n'est pas suffisamment adaptée du fait du faible nombre de coefficients disponibles pour le masquage. Nous avons évalué les performances de notre outil en comparaison avec cette référence pour un ensemble de séquences de différentes résolutions et deux ensembles de pas de quantification (QP) correspondant aux hauts et moyens débits.

Suite à l'analyse réalisée dans le chapitre 2 pour le cas Inter, l'information de signalisation masquée par la méthode correspond à l'indice de compétition des prédicteurs

de mouvement MV-Comp. Nous rappelons que l'un des arguments de ce choix est la volonté d'augmenter les performances de cet outil en diminuant le coût de signalisation qu'il engendre. L'amélioration introduite dans 4.2.3 est activée et les modifications m_j appartiennent à l'ensemble $\{-5, -3, -1, 1, 3, 5\}$ tandis que c'est la méthode d'optimisation classique qui permet de déterminer le masquage optimal (lorsque ce n'est pas précisé).

Cette configuration est résumée ci-dessous :

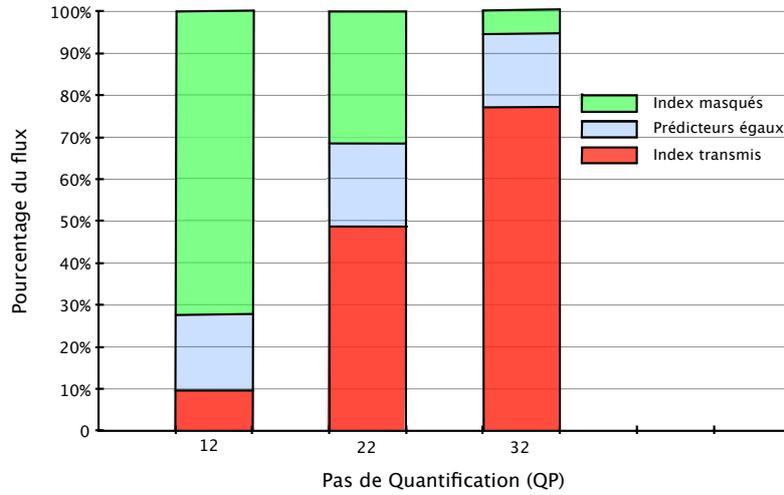
Codeur H.264/AVC, JM KTA version 2.1.
Profil IPPP et IBBP.
Modes Modes Intra 16×16 à 4×4 , Skip et Inter 16×16 à 8×8 .
Transformée DCT 4×4 .
Codeur Entropique CABAC.
QPs deux ensembles testés, Test 1 : $\{12, 17, 22, 27\}$, Test 2 : $\{22, 27, 32, 37\}$.
Séquences 9 séquences de résolutions CIF à 1080p.
Evaluation Pourcentage de réduction de débit moyen par la métrique de Bjontegaard [Bjontegaard 01, Jung & Pateux 07].

4.3.1.1.2 Evaluation débit-distorsion

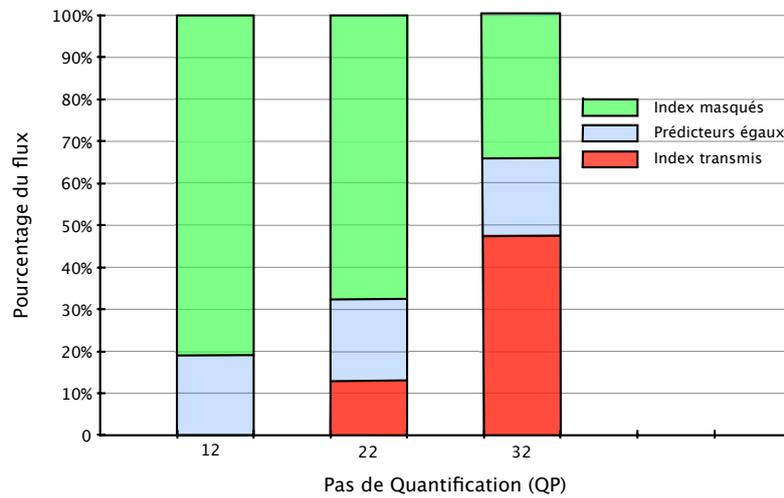
La figure 4.3 donne le pourcentage d'indices de prédiction des vecteurs mouvement (MV-Comp) transmis et non-transmis grâce au test d'égalité des prédicteurs (cf. Chapitre 3, 1.1.3.1) ou à la méthode de masquage d'indice. Ces pourcentages sont présentés pour le masquage dans la chroma uniquement et à la fois dans la chroma et la luma. Nous observons que la proportion de prédicteurs égaux demeure constante quelque soit le QP. Pour le premier schéma (a), le pourcentage d'indices non-transmis est au final supérieur à 30% et atteint 90% à haut débit (QP=12). La méthode proposée est particulièrement efficace pour ce débit où environ 70% des indices sont masqués. En effet, le ratio entre le nombre de coefficients disponibles pour le masquage dans chaque bloc et le nombre d'indices MV-Comp à transmettre est plus favorable à haut débit.

Pour le second schéma qui ajoute le masquage dans la luma, la même évolution est observée avec une augmentation d'indices masqués qui atteint 80% pour le QP=12 et 35% pour le QP=32. En effet, le nombre de coefficients transformés disponibles pour le masquage est logiquement plus élevé dans la composante de luma. Ainsi, à haut débit aucune transmission d'indice n'est nécessaire grâce au schéma proposé et moins de 50% à moyen débit.

La méthode proposée est sensible au nombre de coefficients disponibles, et par conséquent est plus efficace à haut débit.



(a)



(b)

FIGURE 4.3: Pourcentage d'indices MV-Comp transmis ou non-transmis dû au test d'égalité et au masquage d'indice pour les deux schémas : masquage dans la chroma (a) et masquage à la fois dans la chroma et la luma (b), pour trois valeurs de QP.

Notre schéma permet donc d'accomplir le but recherché, particulièrement lors du masquage dans la luma et la chroma à haut débit, regardons maintenant l'impact de la méthode sur l'efficacité de codage. Les résultats présentés dans les tableaux suivants (4.1 et 4.2) donnent le pourcentage de gains sur le débit pour chaque séquence pour les deux masquages. Nous avons de plus souhaité mesurer l'impact du nombre de coefficients disponibles pour le masquage à l'aide d'un test où seul le partitionnement 16×16 est activé parmi les modes Inter. Les résultats sont présentés pour les deux séries de QPs.

En accord avec les conclusions de l'étude du chapitre 2, l'impact sur le codage global est limité du fait de la proportion de l'indice MV-Comp dans le débit total qui varie de 0.3% à 1.5% pour des QPs de 12 et 32 respectivement. Les résultats obtenus ne sont cependant pas négligeables et prouvent l'intérêt de notre approche.

QP	Chroma 16x16		Chroma 16x16 à 8x8	
	Test 1	Test 2	Test 1	Test 2
Foreman CIF	0.4	0.6	0.2	0.1
Bbc_news CIF	0.8	0.9	2.3	2.0
Schumacher CIF	0.1	0.4	0.5	0.2
Drama CIF	0.6	0.4	1.7	0.3
News_car CIF	0.8	1.0	2.2	1.9
Rugby CIF	0.6	0.2	1.3	0.6
Keiba3 WVGA	2.4	-0.4	0.4	-0.6
Crowdrun 720p	0.3	0.2	0.6	0.2
Crowdrun 1080p	0.3	0.2	0.6	0.4
Moyenne	0.7	0.4	1.1	0.6

TABLE 4.1: Pourcentage de réduction du débit moyen pour la méthode de masquage dans la composante de chroma. Modifications sélectionnées à l'aide de l'optimisation débit-distorsion classique.

Le tableau 4.1 donne les gains pour le masquage dans la chroma uniquement. Pour chaque séquence, la méthode permet systématiquement d'obtenir un gain excepté pour *Keiba3 WVGA* à moyen débit. Cette séquence est aussi celle qui a le meilleur gain à haut débit en Inter 16×16 uniquement, avec 2.4% de gains sur le débit. Cette différence de comportement s'explique par le nombre élevé de coefficients transformés disponibles à haut débit, en particulier avec l'Inter 16×16 uniquement lorsque ce mode est utilisé sur des zones texturées difficile à prédire. Le masquage dans les modes Inter 16×16 à 8×8 permet de meilleurs résultats en moyenne, notamment à haut débit avec 1.1%. Comme attendu avec la figure 4.3, les résultats selon le QP sont meilleurs pour le premier ensemble (Test 1) qui correspond aux hauts débits. En effet, du fait de la diminution des coefficients disponibles pour le masquage, l'efficacité du schéma diminue pour les bas débits et devient quasiment négligeable pour les très bas débits. Finalement, comme expliqué dans la section 4.2.1, il est important de noter que la méthode proposée a potentiellement un double impact sur l'efficacité de codage : d'une part, la diminution du coût de codage grâce à la non-transmission des indices et d'autre part, l'amélioration du codage des coefficients résiduels par la suppression des coefficients les moins utiles. Par conséquent, les gains obtenus par la méthode peuvent être parfois supérieurs au coût de l'indice lui-même. Ce cas de figure n'est pas fréquent mais se présente pour la séquence *Bbc_news CIF* par exemple.

L'efficacité de la méthode est liée au nombre de coefficients résiduels disponibles. Dans le but de masquer davantage d'indices, nous avons étendu le schéma en masquant à la fois dans les résiduels de chroma et de luma. Le tableau 4.2 donne le pourcentage de gain sur le débit obtenu avec ce masquage.

QP	Chroma+Luma 16×16		Chroma+Luma 16×16 à 8×8	
	Test 1	Test 2	Test 1	Test 2
Foreman CIF	0.7	1.2	0.8	0.3
Bbc_news CIF	0.9	0.3	2.3	1.4
Schumacher CIF	0.4	0.9	0.9	0.5
Drama CIF	0.7	0.5	1.7	-0.1
News_car CIF	0.6	0.5	1.9	1.2
Rugby CIF	0.5	0.3	1.2	0.5
Keiba3 WVGA	2.4	0.6	0.7	0.5
Crowdrun 720p	0.4	0.5	0.8	0.4
Crowdrun 1080p	0.4	0.5	0.7	0.3
Moyenne	0.8	0.6	1.2	0.5

TABLE 4.2: Pourcentage de réduction du débit moyen pour la méthode de masquage à la fois dans les composantes de chroma et de luma. Modifications sélectionnées à l'aide de l'optimisation débit-distorsion classique.

Les gains sont meilleurs en moyenne comparés aux gains précédents, excepté pour le masquage avec les modes Inter 16×16 à 8×8 à moyen débit, où ils sont équivalents. Le gain le plus élevé est toujours obtenu pour la séquence *Keiba3 WVGA* à haut débit avec le mode Inter 16×16 et pour les séquences *Bbc_news CIF* et *News_car CIF* lorsque les modes Inter 16×16 à 8×8 sont activés avec des gains respectifs de 2.3% et 1.9%.

Cependant, même si le masquage dans la luma permet de cacher davantage d'indices, le gain sur le débit n'augmente pas en proportion. Ceci est un point important de notre étude :

le masquage dans la luma doit être réalisé délicatement dans le but de limiter la propagation aux images suivantes de la dégradation de la prédiction ajoutée.

Le schéma proposé en Inter peut être évalué vis à vis de son impact sur le débit de l'information de mouvement contenant les vecteurs de mouvement, les images de référence et l'indice des prédicteurs MV-Comp. La figure 4.4 donne le pourcentage moyen de réduction du débit de l'information de mouvement comparé à la référence sur l'ensemble des séquences proposées pour le masquage dans la chroma (courbe rouge) et à la fois dans la chroma et la luma (courbe bleue).

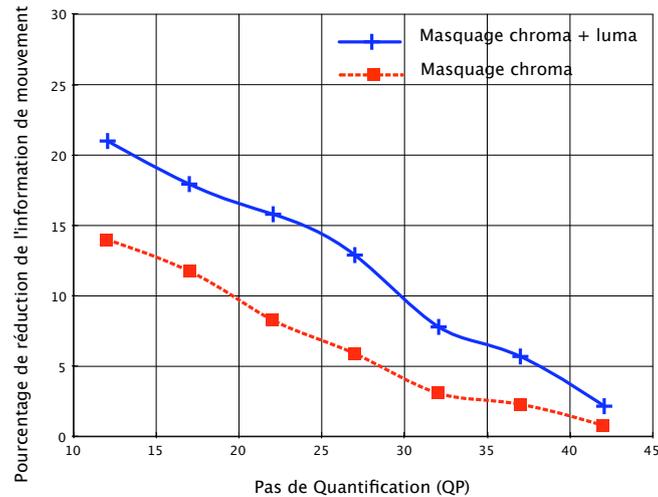


FIGURE 4.4: Pourcentage moyen de réduction de la proportion de l'indice dans le débit de l'information de mouvement comparé à la référence pour l'ensemble des séquences proposés. Masquage dans la chroma (rouge). Masquage à la fois dans la chroma et la luma (bleu).

La réduction atteint 20% pour le masquage à la fois dans la chroma et la luma et elle est d'environ 10% en moyenne. Comme on peut l'observer, la réduction est plus élevée pour les faibles QPs du fait du plus grand nombre de coefficients résiduels disponibles. A propos du masquage dans la chroma uniquement, nous observons toujours la même évolution avec cependant une réduction de l'information de mouvement quasi-nulle à bas débit. Il est important de noter que l'efficacité de la méthode est très liée au choix de l'information à masquer. Pour cette étude, l'indice est lié à l'information de mouvement et permet effectivement une réduction significative du débit de l'information de mouvement. L'impact sur le débit total peut par conséquent être plus bénéfique avec un indice plus coûteux.

Enfin, la méthode a été évaluée dans une configuration IBBP. Les résultats sont rapportés dans le tableau 4.3 pour les deux masquages, les modes Inter 16×16 à 8×8 et les deux ensembles de QPs. La réduction du débit moyen est plus faible que dans un schéma IPPP avec 0.7% et 0.4% pour le masquage dans la chroma et la luma. Ces résultats plus faibles sont expliqués par la meilleure prédiction de la texture obtenue grâce aux images B qui implique une réduction du nombre de coefficients résiduels disponibles pour le masquage. Il y a par conséquent une diminution de la proportion d'indices masqués de la même manière qu'avec l'augmentation du QP. La méthode demeure cependant efficace pour certaines séquences comme *Bbc_news CIF* et *News_car CIF* (jusqu'à 1.2% et 1.8% de gains respectivement) et nous observons de façon générale la même évolution des résultats pour les deux ensembles de QPs.

QP	Chroma		Chroma+Luma	
	Test 1	Test 2	Test 1	Test 2
Foreman CIF	-0.08	-0.04	0.09	0.17
Bbc_news CIF	1.16	0.96	1.10	0.32
Schumacher CIF	0.08	0.28	-0.01	0.64
Drama CIF	1.16	0.30	1.25	0.38
News_car CIF	1.81	1.29	1.82	1.32
Rugby CIF	0.44	0.05	0.42	0.02
Keiba3 WVGA	0.45	0.55	0.14	0.31
Crowdrun 720p	0.44	0.22	0.65	0.31
Crowdrun 1080p	0.52	0.20	0.69	0.28
Moyenne	0.66	0.42	0.68	0.42

TABLE 4.3: Pourcentage de réduction du débit moyen dans une configuration IBBP pour les deux masquages, l'optimisation débit-distorsion classique et les modes Inter 16×16 à 8×8 .

4.3.1.1.3 Comparaison des méthodes d'optimisation

Dans ce paragraphe, nous présentons les résultats des deux méthodes d'optimisation proposées pour la sélection de la modification. Le tableau 4.4 donne le pourcentage de réduction du débit moyen pour l'optimisation classique et l'optimisation à l'aide de Pareto. Les résultats sont présentés pour les deux masquages pour les modes Inter 16×16 à 8×8 et le premier ensemble de QP (Test 1).

	Optimisation classique		Optimisation de Pareto	
	Chroma	Chroma+Luma	Chroma	Chroma+Luma
Foreman CIF	0.2	0.8	0.2	0.4
Bbc_news CIF	2.3	2.3	2.4	2.1
Schumacher CIF	0.5	0.9	0.5	0.7
Drama CIF	1.7	1.6	1.6	1.5
News_car CIF	2.2	1.9	2.1	2.0
Rugby CIF	1.3	1.2	1.0	1.1
Keiba3 WVGA	0.4	0.7	0.6	0.4
Crowdrun 720p	0.6	0.8	0.6	0.6
Crowdrun 1080p	0.6	0.7	0.6	0.6
Moyenne	1.1	1.2	1.1	1.0

TABLE 4.4: Pourcentage de réduction du débit moyen pour les deux méthodes d'optimisation pour l'ensemble de QP du Test 1 et les modes Inter 16×16 à 8×8 .

Les résultats obtenus avec l'optimisation classique sont un peu meilleurs en moyenne, notamment pour le masquage dans la chroma et la luma qui permet respectivement 1.2% contre 1.0% de gains sur le débit. Pour le masquage dans la chroma uniquement, les résultats des deux méthodes sont équivalents. En effet, après calcul des couples Pareto-optimaux, la méthode de Pareto assigne le même poids aux composantes de débit et de distorsion (après normalisation) tandis que l'optimisation classique favorise davantage la composante de distorsion. Par conséquent, l'optimisation classique induit une plus faible altération de la prédiction et les résultats sont en particulier meilleurs lorsque l'on réalise le masquage dans la luma tandis que l'impact sur la prédiction est moins visible avec le masquage dans la chroma uniquement. Les résultats sont cependant assez proches pour considérer que l'optimisation par Pareto constitue une alternative efficace à l'optimisation débit-distorsion classique dans le cadre de notre méthode.

Pour poursuivre la comparaison des deux méthodes d'optimisation, nous étudions leurs temps d'exécutions¹ à l'aide du tableau 4.5. Nous avons mesuré les temps de calcul moyens (en millisecondes) des deux schémas pour chaque bloc pour l'ensemble des séquences proposées. La différence entre les deux optimisations est négligeable, le temps de calcul moyen pour les deux optimisations étant de 0.066ms. En effet, parmi les différentes étapes du schéma proposé, l'étape d'optimisation n'est pas la plus importante en temps de calcul dans la mesure où elle est toujours plus faible comparée à l'évaluation des composantes de débit et de distorsion. De plus, il est intéressant d'observer l'évolution du temps de calcul selon le QP : dans la mesure où il y a davantage de coefficients résiduels à haut débit, les deux méthodes prennent logiquement davantage de temps à haut débit (QP=12) par rapport aux plus bas débits (QP=42). De la même manière, la différence entre les deux méthodes d'optimisation est plus significative pour les débits plus élevés car la détermination de la frontière de Pareto nécessite alors davantage de temps pour les blocs avec de nombreux coefficients résiduels.

	Optimisation classique	Optimisation de Pareto
QP 12	0.09	0.11
QP 17	0.08	0.08
QP 22	0.07	0.07
QP 27	0.06	0.06
QP 32	0.06	0.06
QP 37	0.05	0.05
QP 42	0.05	0.05
Moyenne	0.07	0.07

TABLE 4.5: Temps de calcul moyen (en millisecondes) sur un bloc pour les deux méthodes d'optimisation à différentes valeurs du pas de quantification (QP).

1. Notons que le temps d'exécution en tant que tel ne reflète pas la complexité mais qu'utilisé sous forme de ratio, avec deux morceaux de code écrits de manière similaire, il donne une indication correcte.

En conclusion, le gain sur le débit et le temps de calcul des deux méthodes d'optimisation sont équivalents et on peut par conséquent sélectionner l'une ou l'autre des méthodes selon leurs avantages et inconvénients discutés en section 4.2.2.

En particulier, l'optimisation de Pareto est bien adaptée pour se permettre d'économiser la délicate détermination des multiplicateurs de Lagrange nécessaires à l'optimisation débit-distorsion classique qui pour sa part ne peut pas être efficace sans ces multiplicateurs. Ainsi, c'est la méthode de Pareto que nous utiliserons pour les résultats présentés dans la sous-section suivante (4.3.2) lors de l'application du schéma de masquage à des informations de codage de type Intra.

4.3.1.1.4 Analyse du schéma

Le tableau 4.6 donne la position de la marque dans les différentes composantes (en pourcentage) pour le cas de masquage dans la chroma et à la fois dans la chroma et la luma. Les résultats sont présentés pour trois valeurs de QP.

	Chroma			Chroma+Luma		
	QP 12	QP 22	QP 32	QP 12	QP 22	QP 32
Non-modifié	49.6	50.3	49.7	50.1	49.7	50.2
Chroma DC	3.6	6.9	8.2	3.0	3.2	1.8
Chroma AC	46.8	42.8	42.1	41.0	21.9	7.9
Luma	/	/	/	5.9	25.2	40.1

TABLE 4.6: Position de la marque (en pourcentage) pour le masquage dans la chroma uniquement, et à la fois dans la chroma et la luma, pour 3 valeurs de QP.

La première ligne du tableau correspond aux cas où les blocs ne nécessitent pas d'avoir un coefficient modifié, c'est à dire que la parité de la somme des coefficients résiduels est déjà correcte. Comme attendu, ce cas représente environ 50% du total et par conséquent seulement 50% des cas de transmission d'indices nécessitent la modification d'un coefficient. Par rapport à cette modification, plusieurs observations peuvent être faites. Lors du masquage dans la chroma uniquement, les coefficients modifiés sont principalement les coefficients AC qui représentent plus de 40% des coefficients modifiés. En effet, les coefficients AC ont moins d'importance sur la prédiction que les coefficients DC. De plus, nous observons que cette répartition évolue avec l'augmentation du QP : davantage de coefficients DC sont sélectionnés pour les QPs plus élevés (QP=32) qui correspondent ici à un débit moyen. Ceci peut être expliqué par le plus faible nombre de coefficients AC dans la chroma pour les moyens et bas débits comparés aux hauts débits, les coefficients DC sont alors plus souvent modifiés. Pour le masquage à la fois dans la chroma et la luma, une distribution équivalente est observée. Nous observons de plus un point

intéressant concernant les indices : ils sont majoritairement masqués dans la composante de chroma à haut débit tandis que c'est la luma qui joue le rôle le plus important pour les plus bas débits. Il s'agit d'un résultat important de notre étude qui vient confirmer le comportement de la méthode observée dans les résultats présentés précédemment. L'efficacité du schéma est en fait très liée au nombre de résiduels de chroma disponibles tandis que le masquage dans la composante de luma doit être réalisé avec précaution afin de limiter la propagation d'erreurs de prédiction ajoutées aux images suivantes.

Ainsi, les résultats des tableaux 4.1 et 4.2 peuvent être interprétés par le fait que le masquage dans la chroma et la luma donne de meilleurs gains à haut débit lorsque le nombre de coefficients de luma modifiés n'est pas trop élevé.

Le tableau 4.7 présente, en pourcentage, la sélection de l'amplitude de la modification $m_j \in \{\pm 1, \pm 3, \pm 5\}$ pour différentes valeurs du coefficient résiduel original a_n .

a_n	Chroma			Chroma+Luma		
	m_j					
	± 1	± 3	± 5	± 1	± 3	± 5
± 1	90.0	0	0	86.2	0	0
± 2	9.0	0	0	12.0	0	0
± 3	0.7	0.1	0	1.3	0.1	0
$\pm 4, \pm 5, \dots$	0.2	0	0	0.4	0	0

Table 4.7: Pourcentage de sélection de l'amplitude de la modification m_j relativement à différentes valeurs de coefficients résiduels originaux a_n pour le masquage dans la chroma uniquement, et à la fois dans la chroma et la luma.

Les coefficients transformés qui sont le plus souvent modifiés sont ceux égaux à ± 1 avec 90% et 86% de sélection respectivement pour le masquage dans la chroma uniquement et dans les deux composantes. En effet, c'est la valeur de coefficient la plus représentée, notamment pour les bas débits. La modification de ces valeurs de coefficients entraîne une faible dégradation visuelle dans la mesure où ce sont les valeurs les moins significatives. Comme nous l'observons dans le tableau, ces valeurs de coefficients sont principalement combinées avec une valeur de modification m_j aussi égale à ± 1 qui permet de réduire le coût de codage des résiduels par l'augmentation des valeurs nulles dans les coefficients à coder. A propos de la répartition selon les composantes utilisées pour le masquage, nous pouvons noter que la répartition des modifications sélectionnées est meilleure pour le masquage à la fois dans la chroma et la luma du fait de la plus grande diversité de valeurs de coefficients résiduels dans la luma.

Enfin, notons que la modification ± 5 n'est pas sélectionnée du fait de la trop grande différence engendrée dans la reconstruction du bloc.

Précisons cependant que les pourcentages présentés dans ce tableau sont arrondis et que les modifications ± 5 et ± 3 ont été sélectionnées quelque fois avec des coefficients a_n de fortes amplitudes mais de façon négligeable (0.01% et 0.02% de sélection en moyenne respectivement en chroma et pour les deux composantes). Ces observations confirment par conséquent la proposition de la sous-section 4.2.4 : la complexité de la méthode au niveau de l'encodeur peut être facilement réduite en restreignant les modifications possibles m_j à ± 1 . L'efficacité de la méthode est alors approximativement la même avec une complexité divisée par quatre.

4.3.1.1.5 Evaluation de la Chroma

Pour clore cette étude objective de la méthode appliquée à l'indice MV-Comp, nous nous arrêtons en détail sur l'impact du schéma sur la composante de chrominance. Le tableau 4.8 donne les valeurs de Δ -PSNR moyen (en dB) relativement aux composantes U et V de la chroma et en regard du pourcentage de gain sur le débit présenté dans le tableau 4.1 pour les deux ensembles de QPs (Test 1 et Test 2).

QP	Δ -PSNR Chroma U (dB)		Δ -PSNR Chroma V (dB)		Gains sur le débit (%)	
	Test 1	Test 2	Test 1	Test 2	Test 1	Test 2
Foreman CIF	-0.10	-0.01	-0.06	-0.01	0.2	0.1
Bbc_news CIF	-0.12	-0.01	-0.20	-0.11	2.3	2.0
Schumacher CIF	-0.06	-0.05	-0.16	-0.08	0.5	0.2
Drama CIF	-0.20	-0.07	-0.12	-0.02	1.7	0.3
News_car CIF	-0.13	-0.06	-0.08	-0.02	2.2	1.9
Rugby CIF	-0.22	-0.08	-0.16	-0.07	1.3	0.6
Keiba3 WVGA	-0.11	-0.05	-0.07	-0.03	0.4	-0.6
Crowdrun 720p	-0.17	-0.08	-0.17	-0.10	0.6	0.2
Crowdrun 1080p	-0.15	-0.06	-0.13	-0.03	0.6	0.4
Moyenne	-0.14	-0.05	-0.13	-0.05	1.1	0.6

Table 4.8: Δ -PSNR des composantes de chroma U et V (en dB) ainsi que le pourcentage de gain sur le débit moyen pour le schéma proposé comparé à la méthode de référence. Masquage dans la chroma uniquement pour les modes Inter 16×16 à 8×8 .

Comme attendu, la méthode proposée entraîne une perte systématique sur le PSNR de chroma pour l'ensemble des séquences. Les pertes sont équivalentes en moyenne pour les composantes U et V, confirmant la distribution équivalente de la marque. Les pertes augmentent de 0.1dB pour les hauts débits (Test 1) par rapport aux moyens débits (Test 2). Ceci est expliqué par le nombre élevé de coefficients résiduels modifiés à haut débit comme nous l'avons vu dans la figure 4.3. Les gains moyens sur le débit introduits dans

le tableau 4.1 sont rapportés dans la dernière colonne : les pertes les plus élevées sur le PSNR en chroma correspondent aux séquences avec le gain sur le débit le plus élevé, en particulier pour les séquences *Bbc_news CIF* et *Drama CIF*.

Dans la mesure où il n'y a pas de métrique globalement acceptée par la communauté afin d'évaluer conjointement chroma et luma, nous ne proposons pas d'évaluation objective incluant directement la chroma. Cependant, la dégradation de la chroma est une conséquence naturelle de notre méthode qui ne peut pas être négligée comme le confirme ce tableau. La question la plus importante est par conséquent de savoir si cette dégradation est perçue par l'observateur humain. Dans ce but, une évaluation subjective est proposée dans la sous-section suivante afin d'évaluer les artefacts éventuels de chroma.

4.3.1.2 Résultats subjectifs

Protocole Expérimental

Une évaluation subjective est une méthode expérimentale visant à déterminer des modifications dans la qualité perçue. Chaque test consiste en un couple de stimuli, comprenant la séquence de référence et la séquence encodée avec la méthode proposée, affichées dans un ordre aléatoire. Cinq des séquences avec la plus grande perte de chroma calculée objectivement (tableau 4.8) ont été affichées deux fois : le premier test est pour une évaluation globale et le second pour une évaluation spécifique de la qualité de la couleur. Dans ce second test, on demande explicitement aux évaluateurs de se concentrer sur la distorsion de la chroma. La figure 4.5 détaille l'échelle utilisée pour l'évaluation de la qualité.

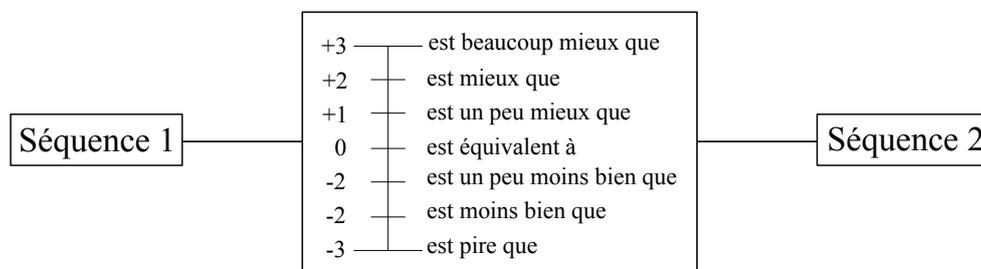
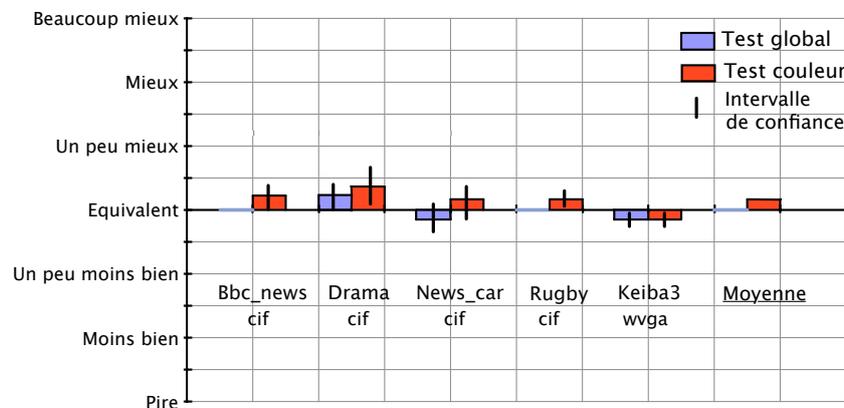


FIGURE 4.5: Echelle d'évaluation de la qualité.

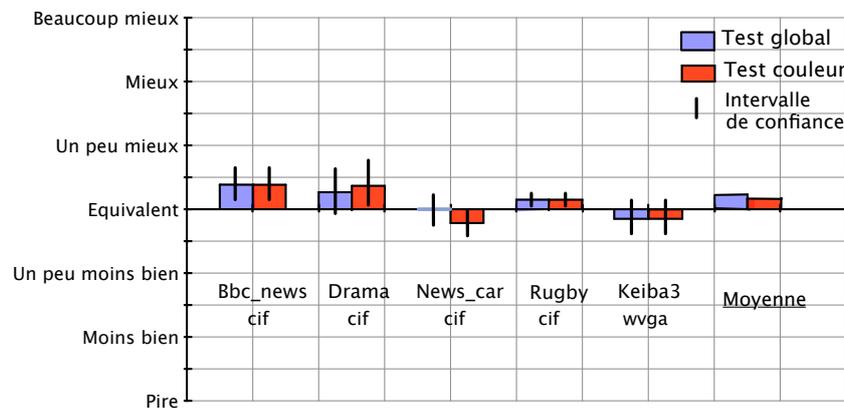
Interprétation

La figure 4.6 présente les résultats obtenus pour chaque séquence lors des deux phases du test, ainsi que pour deux débits différents : haut débit (a) et moyen débit (b). Les MOS (en anglais, *mean opinion score*) et les intervalles de confiance associés sont affichés pour chaque séquence. La conclusion majeure est que la qualité des séquences (globale et du point de vue de la couleur) n'est pas visuellement affectée par la méthode proposée.

Ainsi, aucun évaluateur n'a noté une séquence avec un score supérieur à 1 (« un peu mieux que »). Nous pouvons par conséquent considérer que les dégradations visuelles sont invisibles. De plus, nous observons que l'impact est globalement plus perceptible à haut débit qu'à moyen débit, ce qui confirme les résultats observés précédemment (4.3.1.1). Bien que la taille des intervalles de confiance nécessite d'être prudent pour tirer des conclusions de ces résultats, nous pouvons aussi noter que certains évaluateurs ont mieux noté les séquences encodées avec le schéma proposé, en particulier *Keiba3*, ce qui confirme que la qualité des séquences est très proche. Enfin, d'après ces résultats la modification de la qualité est davantage perceptible pour le second test lorsque les évaluateurs ont eu la consigne de se concentrer sur la dégradation de la couleur, elle demeure cependant faible.



(a)



(b)

FIGURE 4.6: Évaluation subjective des séquences à haut (a) et moyen (b) débit. Comparaison de la référence avec notre méthode.

A titre d'exemple, la figure 4.7 présente des images tirées de la séquence *Bbc_news CIF* qui est à priori l'une des séquences les plus dégradées en terme de chroma. La

figure montre une image de la séquence compressée avec l'encodeur de référence (a), compressée avec le schéma proposé (b), la différence entre les deux images (c) et enfin un zoom sur cette image de différence (d). Le test est réalisé à haut débit afin de se placer dans le cas extrême où un maximum de coefficients sont modifiés. Nous pouvons distinguer quelques blocs de chroma de moins bonne qualité dans la figure (d) mais la différence demeure peu perceptible visuellement entre les images (a) et (b).

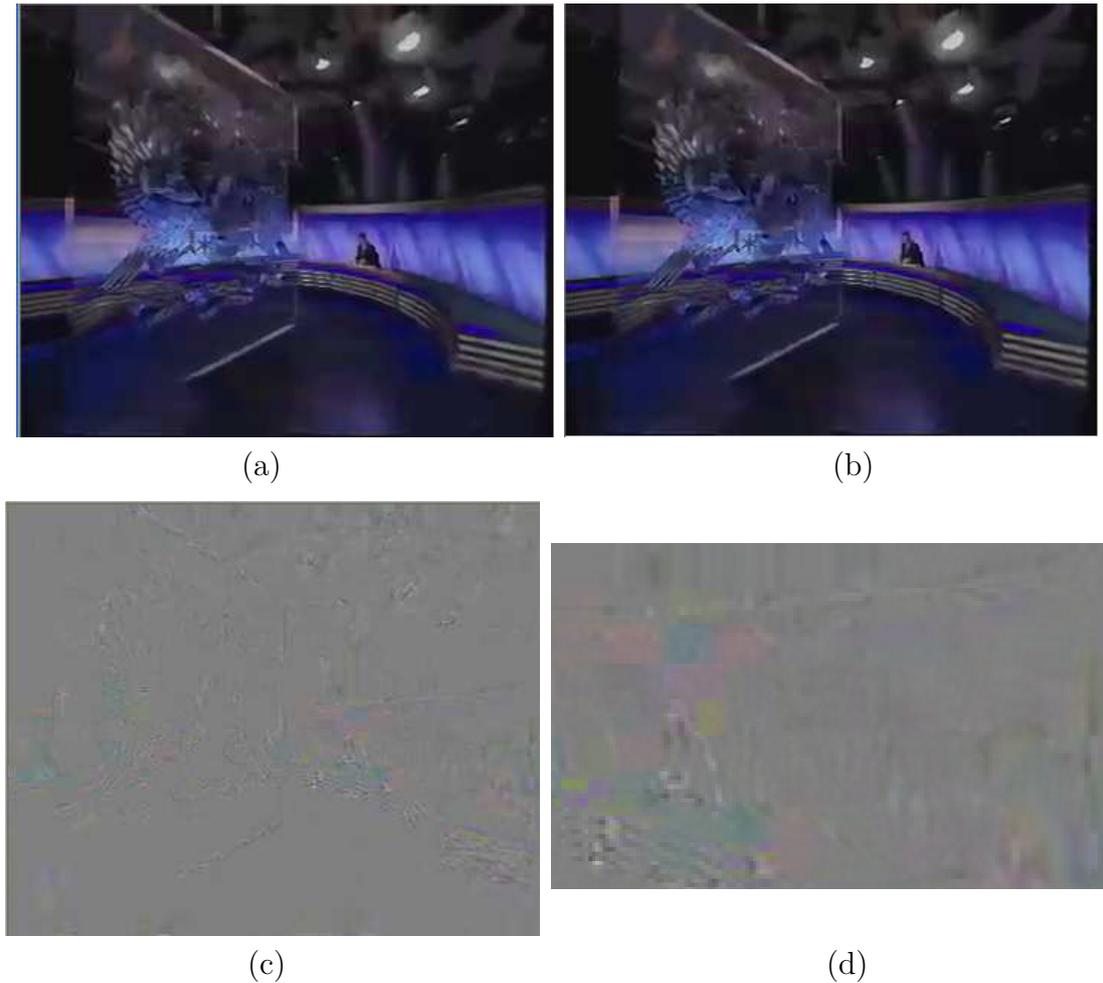


FIGURE 4.7: Séquence *Bbc_news CIF* : (a) compressée avec le codeur de référence ; (b) avec le masquage dans la chroma ; (c) différence entre (a) et (b) ; (d) zoom de (c). Test réalisé à haut débit.

La méthode proposée n'entraîne pas d'altération subjective majeure, ni de la séquence complète, ni de la chroma.

4.3.2 Masquage d'informations Intra

4.3.2.1 Résultats objectifs

4.3.2.1.1 Configuration de tests

La configuration utilisée pour l'ensemble des tests correspond au profil High Intra d'H.264/AVC avec l'ensemble des outils activés excepté le mode Intra 4×4 pour lequel la méthode n'est pas suffisamment adaptée du fait du faible nombre de coefficients disponibles pour le masquage. De la même manière que dans les tests précédents (4.3.1.1), nous avons évalué les performances de notre outil en comparaison avec cette référence pour un ensemble de séquences de différentes résolutions au format YUV 4 :2 :0 et deux ensemble de pas de quantification (QP) correspondants aux hauts et moyens débits. Nous proposons de plus des tests réalisés sur des séquences au format YUV 4 :4 :4 (4.3.2.1.3) afin de mesurer l'impact de composantes de chroma supplémentaires sur notre schéma.

Suite à l'analyse réalisée dans le chapitre 2 pour le cas Intra, l'information de signalisation masquée par la méthode correspond ici à l'indice *Most Probable Mode* (MPM). L'amélioration introduite dans 4.2.3 est activée et les modifications m_j appartiennent à l'ensemble $\{-5, -3, -1, 1, 3, 5\}$ tandis que c'est la méthode d'optimisation de Pareto qui permet de déterminer le masquage optimal. Enfin, point important, le masquage est appliqué uniquement dans la chroma lors de cette expérimentation.

Cette configuration est résumée ci-dessous :

Codeur H.264/AVC, JM KTA version 2.1.
Profil High Intra.
Modes Modes Intra 16×16 à 8×8.
Transformée DCT 8×8.
Codeur_Entropique CABAC.
QPs deux ensembles testés, Test 1 : {12, 17, 22, 27}, Test 2 : {22, 27, 32, 37}.
Séquences 8 séquences YUV 4 :2 :0 de résolutions qwvga à 1080p ; 12 séquences disponibles à la fois en YUV 4 :2 :0 et YUV 4 :4 :4.
Evaluation Pourcentage de réduction de débit moyen par la métrique de Bjontegaard [Bjontegaard 01, Jung & Pateux 07].

4.3.2.1.2 Résultats en YUV 4 :2 :0

La figure 4.8 présente le pourcentage d'indices MPM transmis et non transmis selon qu'ils aient été masqués dans la composante DC ou AC de la Chroma. Ces chiffres sont proposés pour trois valeurs de QP. On observe qu'un peu plus de 80% des indices sont masqués pour le QP=12 et que cette proportion diminue pour le moyen débit avec un peu moins de 20% des indices masqués. De la même manière que pour l'expérimentation en Inter, cette évolution s'explique par la variation du nombre de coefficients disponibles

pour le masquage. Notons cependant que davantage d'indices sont masqués dans ce test.

Par rapport à la position de la marque, nous observons que les indices sont majoritairement masqués dans la composante AC à haut débit tandis que la composante DC joue un rôle plus important à moyen débit. Par rapport au cas Inter (voir tableau 4.6), la composante DC est en effet davantage exploitée car elle présente des amplitudes plus importante en Intra qu'en Inter.

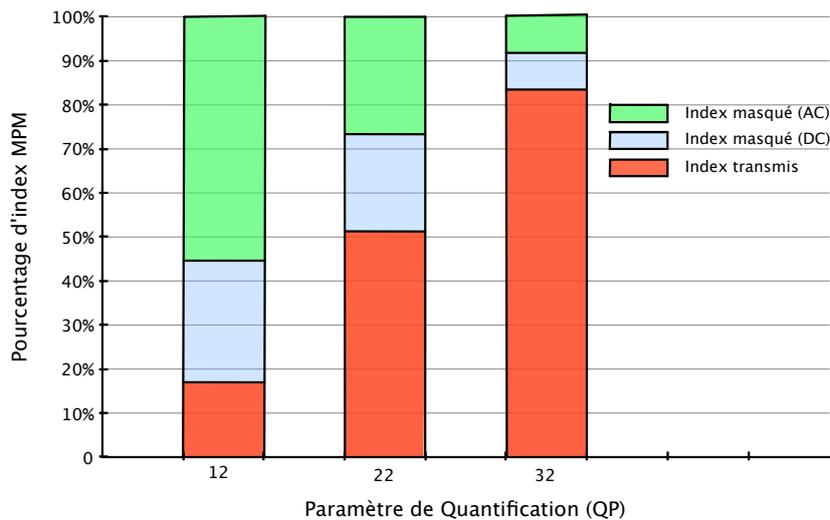


FIGURE 4.8: Pourcentage d'indices MPM transmis ou masqués dans les coefficients AC ou DC, pour trois valeurs de QP.

Le tableau 4.9 donne le pourcentage de gains sur le débit pour chaque séquence et pour les deux ensembles de QPs. La méthode proposée donne systématiquement un gain pour toutes les séquences étudiées qui atteint 2.8% et 2.1% pour les séquences *SunFlower 1080p* et *BlueSky 1080p*. Contrairement au précédent schéma proposé en codage Inter, l'efficacité de la méthode est préservée à moyen débit où le nombre de coefficients transformés est pourtant moins élevé qu'à haut débit avec respectivement 1.5% et 1.6% de gains sur le débit en moyenne. Ce gain est significatif compte tenu du fait que la méthode traite seulement un indice de signalisation dont la proportion dans le flux reste limitée bien qu'étant l'indice le plus coûteux (6% du débit Intra total à moyen débit). En considérant de plus que la méthode demeure bornée en Intra par le nombre de blocs ayant des coefficients non nuls, ces gains confirment l'intérêt d'une telle approche pour réduire le coût de transmission d'indices de signalisation.

Ensembles de QP	Test 1	Test 2
Mobisode2 qwvga	2.3	1.3
RaceHorses qwvga	0.9	1.1
RaceHorses wvga	1.0	1.0
DucksTakeOff 720p	1.4	1.7
BlueSky 1080p	1.3	2.1
DucksTakeOff 1080p	1.6	1.8
SunFlower 1080p	2.8	1.1
Tractor 1080p	1.9	1.5
Moyenne	1.6	1.5

TABLE 4.9: Pourcentage de gains sur le débit pour les deux ensembles de QPs.

En résumé, cette expérimentation en Intra apporte des gains plus élevés qu'en Inter du fait du plus grand nombre de coefficients disponibles pour le masquage.

4.3.2.1.3 Résultats en YUV 4 :4 :4

Dans ce paragraphe, le schéma est appliqué au format de couleur 4 :4 :4 qui est un format important pour l'amélioration future de la qualité des services vidéos. Dans le cadre de notre étude, il permet logiquement de masquer davantage d'indices dans la chroma qu'en YUV 4 :2 :0 comme illustré sur la figure 4.9. Chaque index étant caché dans plus de coefficients, il y a aussi un impact réduit sur les pixels reconstruits car la marque est mieux répartie.

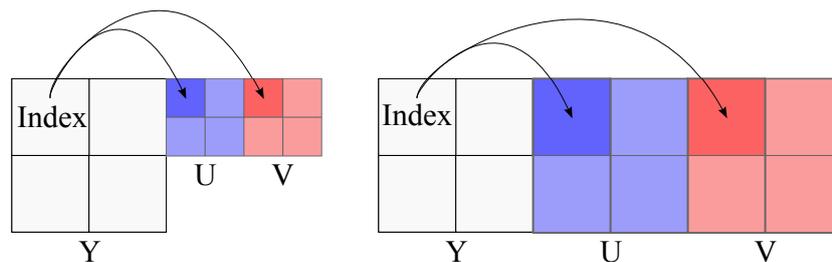


FIGURE 4.9: Schéma de masquage en YUV 4 :2 :0 (à droite) et en YUV 4 :4 :4 (à gauche).

Nous avons testé la méthode sur un ensemble de douze séquences disponibles à la fois au format de couleur YUV 4 :2 :0 et YUV 4 :4 :4. La figure 4.10 donne le pourcentage moyen d'indices cachés avec le schéma appliqué pour les deux formats pour des QPs de 12 à 37. Le masquage en YUV 4 :4 :4 permet systématiquement de réduire le nombre d'indices transmis comparé au masquage pour les séquences en YUV 4 :2 :0. Cette

réduction varie entre 10% et 20% et est plus élevée pour les QPs 17 et 22 où le nombre de coefficients transformés commence à décroître pour les séquences YUV 4 :2 :0. Pour la méthode étendue au format de couleur 4 :4 :4, le pourcentage d'indices masqués atteint 93% à haut débit. Comme attendu, moins d'indices sont transmis pour ce format. Cela est confirmé par une augmentation du gain sur le débit qui atteint 2.60% pour la séquence *DucksTakeOff* 720p pour l'ensemble de QPs du Test 2 en YUV 4 :4 :4 (contre 1.70% en YUV 4 :2 :0).

Notre schéma apporte une amélioration plus significative de l'efficacité de codage en YUV 4 :4 :4 par rapport au YUV 4 :2 :0 avec une augmentation de 36% et 47% des gains sur le débit en moyenne pour les QPs des tests 1 et 2 respectivement.

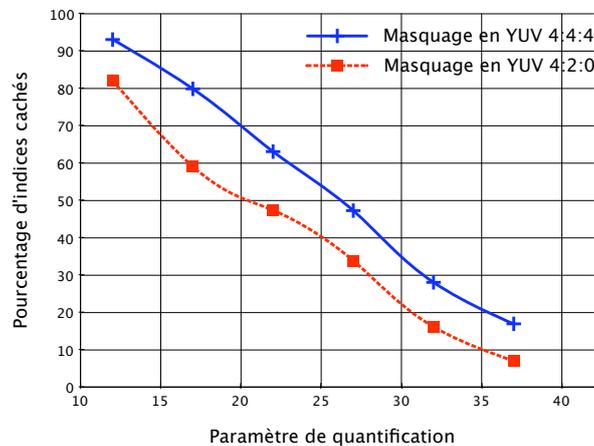
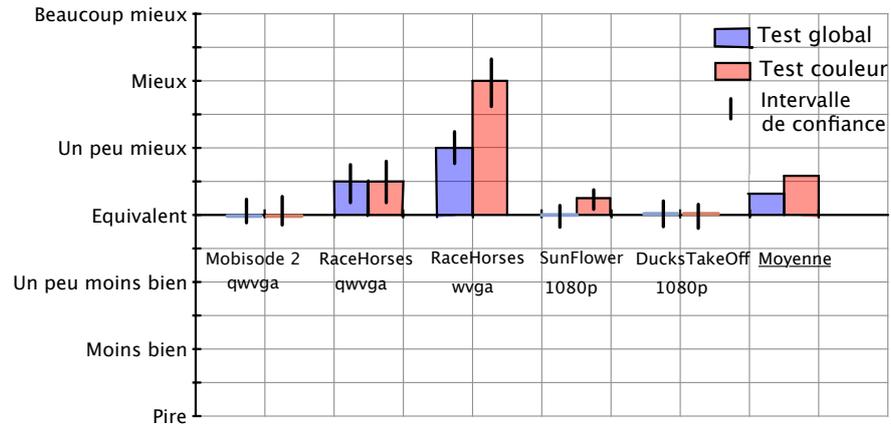


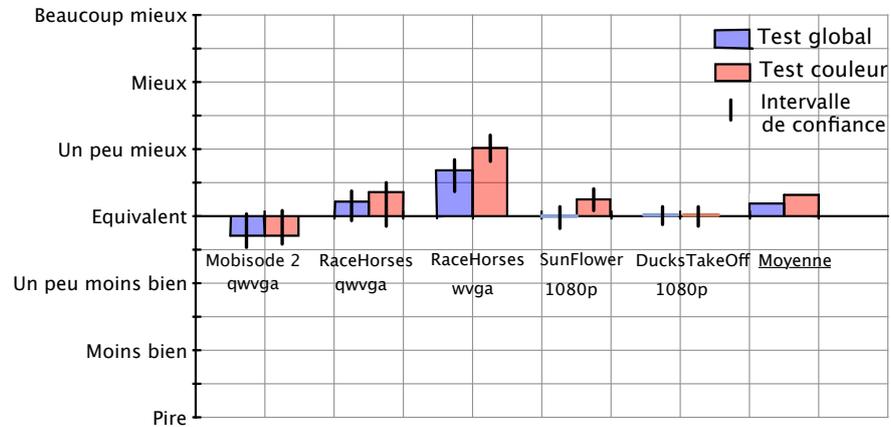
FIGURE 4.10: Pourcentage d'indices masqués en YUV 4 :2 :0 et YUV 4 :4 :4 selon le QP.

4.3.2.2 Résultats subjectifs

De la même manière que pour le cas Inter (4.3.1.2), la figure 4.11 résume les résultats subjectifs obtenus pour chaque séquence pour les deux phases de tests, à haut (a) et moyen (b) débits. La conclusion majeure est toujours que la qualité des séquences (globale et couleur) n'est pas visuellement affectée par la méthode proposée mais elle l'est cependant un peu plus qu'en Inter du fait du plus grand nombre d'indices masqués. En dehors de la séquence *RaceHorses* dont les couleurs sont très vives, aucun score supérieur à 1 ("un peu meilleur") n'a été utilisé par les évaluateurs. La dégradation visuelle peut donc être considérée comme limitée. On peut ici aussi noter que l'impact de la méthode est davantage perçue à haut débit qu'à moyen débit, ce qui confirme les résultats observés précédemment. Enfin, la modification de la qualité est davantage perçue dans le test axé sur la couleur mais reste faible.



(a)



(b)

Figure 4.11: Evaluation subjective des séquences à haut (a) et moyen (b) débit. Comparaison de la référence avec notre méthode.

La figure 4.12 illustre cette analyse avec la séquence *RaceHorses WVGA* qui correspond à la perte la plus élevée subjectivement. La figure présente une capture de la séquence compressée pour un haut débit avec le codeur de référence (a), avec la méthode proposée (b) et la différence des deux séquences (c).

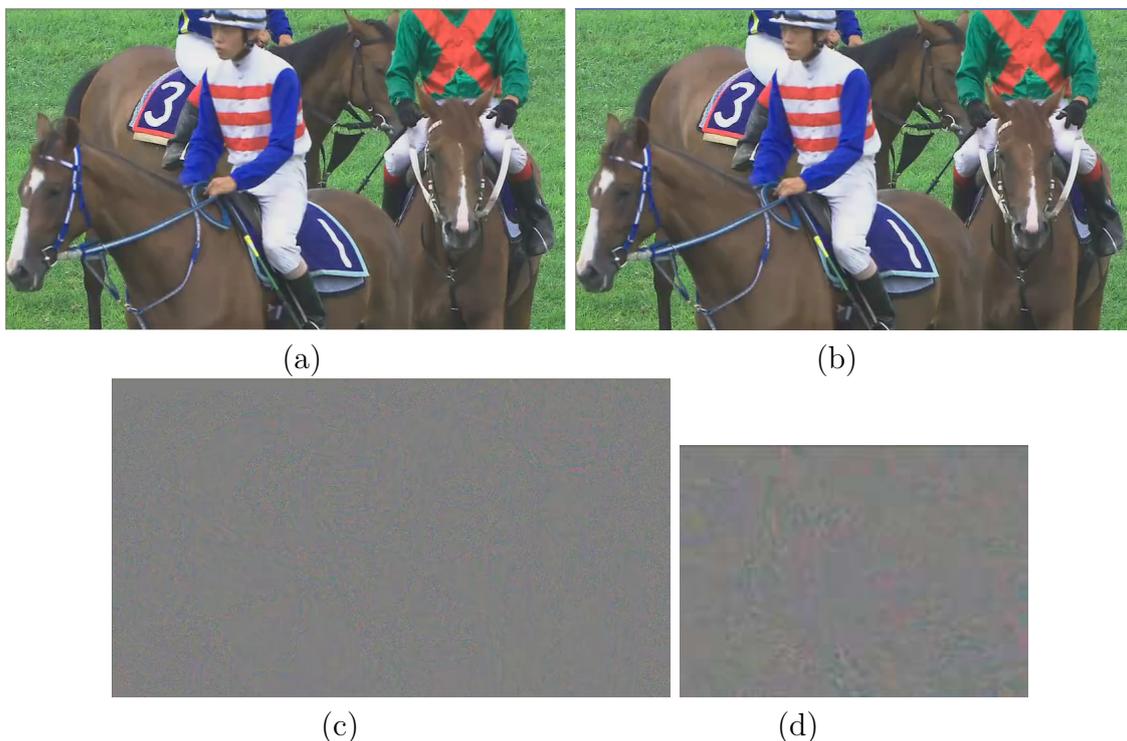


FIGURE 4.12: Séquence *RaceHorses* WVGA : (a) compressée avec le codeur de référence ; (b) avec le masquage dans la chroma ; (c) différence entre (a) et (b) ; (d) zoom de (c). Test réalisé à haut débit.

4.4 Bilan

Dans ce chapitre nous avons présenté une application non-conventionnelle du tatouage : il est utilisé au sein d'un codeur vidéo pour réduire les coûts de signalisation. Nous masquons des éléments de syntaxe dans les coefficients résiduels de chroma et de luma à l'aide d'une optimisation débit-distorsion permettant de minimiser la dégradation de la prédiction engendrée. Nous avons mis en lumière et résolu les principaux problèmes liés à une telle approche dans le but d'obtenir un bon compromis entre le gain en coût de codage, la propagation des erreurs de prédiction dans la luma et les aspects visuels liés à la chroma. En particulier, les sujets de la quantité de données à masquer, de la position de la marque ainsi que l'intensité de l'altération des coefficients, ont été abordés.

En moyenne, les gains présentés ne sont pas négligeables avec respectivement, pour le masquage d'indices en Inter et en Intra, 1.3% et 1.6% de gains sur le débit pour une évaluation basée sur le PSNR-Y associée néanmoins à une dégradation de la chroma. Ces résultats sont très liés à la proportion de l'indice traité dans le débit total, qui est relativement faible dans les deux cas, mais aussi à l'impact de la méthode sur la

dégradation de la prédiction. Les gains observés ont cependant un réel intérêt (jusqu'à 2.3% en Inter et 2.8% en Intra) et surtout le but recherché est atteint avec 50% d'indices non-transmis en moyenne. La méthode proposée apparait par conséquent comme une manière astucieuse et efficace de réduire le coût d'autres informations de codage plus coûteuses tout en prenant bien soin de faire attention à la dégradation de la composante de chroma. En particulier, ceci peut être particulièrement utile pour de nouveaux modes de codage qui apportent une nette amélioration sur les performances en distorsion mais demeurent limités par le coût de signalisation engendré.

A ce jour, ces travaux ont donné lieu à un stage d'ingénieur [Raad Cisa 11] pour étudier la mise en œuvre de cette méthode dans le futur standard HEVC. Les résultats très encourageants obtenus ont abouti à plusieurs contributions de notre équipe [Clare *et al.* 12] en standardisation repris par d'autres compagnies comme RIM (Research In Motion). Finalement, la solution proposée par RIM [Yu *et al.* 12] a été jugée suffisamment efficace et peu complexe pour être intégrée dans la future norme HEVC [Bross *et al.* 12] au sein du profil principal.

Chapitre 5

Codage des vecteurs de mouvement par redistribution des résiduels basée sur un pronostic du mouvement

Le codage à moindre coût des vecteurs de mouvement est un enjeu majeur des codeurs vidéo modernes qui sont basés sur une exploitation très efficace des redondances temporelles. La combinaison d'une estimation de mouvement précise avec une compensation de mouvement réalisée pour différentes tailles de partitions résulte en effet en une réduction significative du coût de codage de la texture. Par conséquent, la proportion de l'information de mouvement dans le flux total a augmenté en parallèle et constitue maintenant un sujet largement traité dans les récentes contributions. Dans ce chapitre nous traitons ce sujet délicat avec une approche en rupture par rapport à la plupart des travaux visités dans le chapitre 3, bâtis principalement sur une compétition des prédicteurs de mouvement.

L'idée de départ de notre méthode¹ est l'observation de la redondance des vecteurs de mouvement entre des blocs voisins tant spatialement que temporellement. Après avoir collecté cette information de mouvement propre à chaque bloc, nous proposons de réaliser un pronostic précis de chaque mouvement à l'aide d'un modèle probabiliste avant de modifier la distribution des vecteurs résiduels à coder afin de privilégier les mouvements les plus attendus. Après la présentation de notre méthode, nous réalisons une comparaison avec l'outil MV-Comp avant d'envisager une manière efficace de combiner les deux approches.

1. Ce travail a été initié au cours du stage de Julien Le Tanou [Le Tanou 10].

5.1 Description de la méthode proposée

5.1.1 Schéma global

Dans le chapitre 2 nous avons mis en avant les limites du codeur vidéo H.264/AVC liées au coût de transmission des vecteurs de mouvement. Ils constituent l'information la plus importante dès les moyens débits en profil hiérarchique ou bi-prédictif et représentent en moyenne un quart du flux pour les différents profils Inter. Le sujet du codage des vecteurs de mouvement est par conséquent un point clé des schémas de compression vidéo comme le confirme l'inventaire des principaux travaux réalisé au chapitre 3 ainsi que les récentes activités de standardisation. En particulier, dans le cadre du projet HEVC, le sujet du codage des vecteurs de mouvement anime de nombreuses études et entraîne de nouvelles problématiques liées à la robustesse ou au décodage des indices de compétition.

Notre méthode, notée MV-FMap pour *Motion Vector Forecast and Mapping*, se veut en rupture des méthodes basées sur la mise en compétition de prédicteurs de mouvement, notamment du fait des difficultés énoncées précédemment. L'idée principale est basée sur l'observation, présentée dans le chapitre 2, de la forte corrélation entre les vecteurs de mouvement d'images ou de blocs successifs. Ces corrélations sont certes exploitées par les méthodes de compétitions de prédicteurs mais cela reste limité du fait du coût de signalisation du choix de prédicteurs. Par exemple, uniquement deux prédicteurs (médian spatial et vecteur du bloc colocalisé) sont utilisés dans la configuration optimale de l'outil MV-Comp [Laroche *et al.* 08]. En opposition, notre méthode est construite de telle manière qu'aucun indice de signalisation additionnel ne soit transmis : l'erreur de prédiction des vecteurs est corrigée a posteriori.

En s'appuyant sur une étape additionnelle d'observation et de pronostic de l'information de mouvement, MV-FMap modifie efficacement la distribution des vecteurs de mouvement résiduels avant codage entropique dans le but d'améliorer leur codage. Ceci est réalisé pour chaque bloc afin de tirer profit au maximum des corrélations entre vecteurs de mouvement. Nous divisons la méthode en trois étapes représentées dans la figure 5.1 :

1. Création de l'information de mouvement à partir de la collecte des vecteurs des images et blocs précédents.
2. Détermination d'une densité de probabilité de l'ensemble des vecteurs de mouvement résiduels permettant d'affiner l'information de mouvement collectée.
3. Modification de la distribution originale des vecteurs résiduels : les résiduels les plus probables sont associés à des positions de résiduels de plus faible coût.

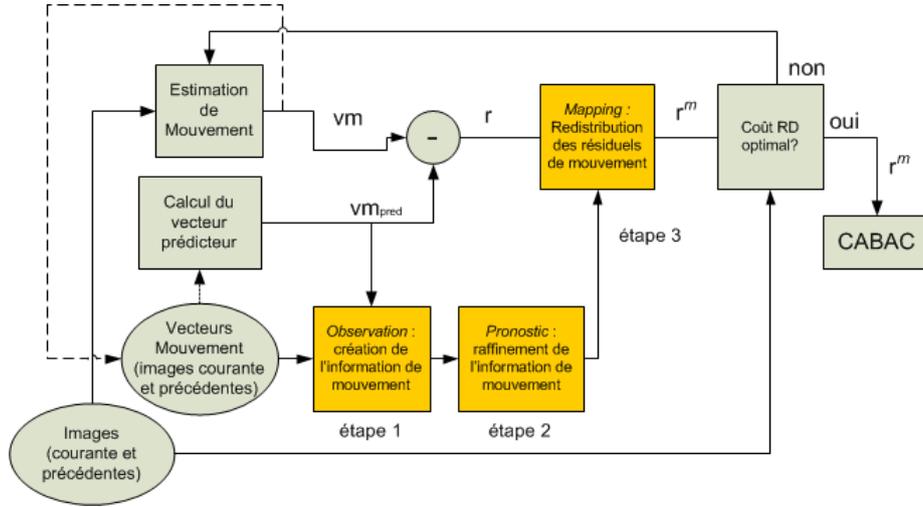


Figure 5.1: Vue globale de la méthode proposée.

Cette redistribution est réalisée au cours de l'estimation de mouvement et elle est reproductible au niveau du décodeur. Par conséquent, un point majeur de notre méthode est qu'aucune information additionnelle n'est nécessaire à la différence de plusieurs méthodes introduites dans la chapitre 3. Les trois étapes constituant notre solution sont détaillées dans les sous-sections suivantes.

5.1.2 Etape 1 : Création de l'information de mouvement

Idéalement, nous souhaitons connaître le vecteur de mouvement optimal d'un bloc courant ou plus généralement la liste des vecteurs de mouvement pour la prédiction de l'image courante. Cette information n'est évidemment pas disponible et nous choisissons de l'approcher à partir des vecteurs de mouvement précédemment codés : ils constituent l'information de mouvement, dite information observée. Appelons $D_{vm} \subset \mathbb{Z}^2$ (respectivement $D_r \subset \mathbb{Z}^2$) le domaine de définition des vecteurs de mouvement (respectivement des vecteurs résiduel).

L'histogramme (h_{vm}^i) des vecteurs de mouvement précédemment codés est défini comme suit pour chaque bloc b^i :

$$\begin{aligned} h_{vm}^i : D_{vm} &\longrightarrow \mathbb{N} \\ vm &\longmapsto n, \end{aligned} \quad (5.1)$$

où n représente le nombre d'occurrences du vecteur vm dans un ensemble prédéfini de blocs précédemment codés, noté E_b^i .

Formellement, après initialisation de l'histogramme, appelons $vm^j, \forall j < i$ le vecteur de mouvement associé au bloc $b^j \in E_b^i$, alors (h_{vm}^i) est mis à jour selon la règle :

$$h_{vm}^i(vm^j) \leftarrow h_{vm}^i(vm^j) + 1. \quad (5.2)$$

Typiquement, l'ensemble E_b^i peut être constitué de tous les blocs de l'image courante, déjà codés, ainsi que tous ceux de l'image précédente pour une information de mouvement dite *globale*, ou bien uniquement des blocs voisins spatialement et temporellement pour une information de mouvement dite *locale*. Nous comparons ces deux configurations dans la sous-section 5.3.1.

Afin de s'intégrer dans un schéma de codage vidéo, rappelons que l'encodeur ne transmet pas les vecteurs directement mais utilise une prédiction et traite par conséquent des vecteurs résiduels. Ainsi, pour chaque bloc b^i à coder en Inter le vecteur de mouvement vm^i , déterminé par l'estimation de mouvement, est prédit à l'aide d'un vecteur prédicteur vm_{pred}^i fixé (médian spatial dans H.264/AVC). C'est le vecteur de mouvement résiduel $r^i = (r_x^i, r_y^i) \in D_r$ qui est alors codé, où r^i est défini par :

$$r^i = vm^i - vm_{pred}^i. \quad (5.3)$$

Dans le cadre de MV-FMap, on déduit l'histogramme des vecteurs résiduels (h_r^i) de l'histogramme (h_{vm}^i) en effectuant le changement de variable bi-univoque :

$$r = vm - vm_{pred}^i, \forall vm \in D_{vm}, \quad (5.4)$$

qui équivaut à $vm = r + vm_{pred}^i, \forall r \in D_r$. Il en résulte :

$$h_r^i(r) = h_{vm}^i(r + vm_{pred}^i), \forall r \in D_r = D_{vm - vm_{pred}^i}. \quad (5.5)$$

C'est ce dernier histogramme que nous utiliserons dans la seconde étape afin de pronostiquer les résiduels attendus pour le bloc courant. La figure 5.2 donne un exemple de distribution de ces résiduels pour un bloc donné. Dans l'image courante, nous avons fait apparaître les quatre vecteurs associés aux blocs voisins du bloc courant ; ces vecteurs dits *spatiaux* sont reportés sur la partie droite de la figure à partir de l'origine du référentiel des vecteurs de mouvement (en noir), classiquement positionnée dans le coin haut et gauche du bloc courant. En complément des positions de ces vecteurs attendus (points rouges ou verts) figurent aussi celles des vecteurs collectés dans l'image de référence, dits *temporels*. Le référentiel des résiduels, centré sur vm_{pred}^i , est pour sa part représenté en vert et permet de visualiser le principe du changement de variable bi-univoque de l'équation 5.5. Enfin, la fenêtre rouge, elle aussi centrée sur vm_{pred}^i , représente la zone de recherche de l'estimation de mouvement $FR \subset D_r$ à laquelle nous limitons les résiduels utilisés pour notre étude. Les résiduels utilisés pour MV-FMap sont à l'intérieur de FR et sont représentés en rouge.

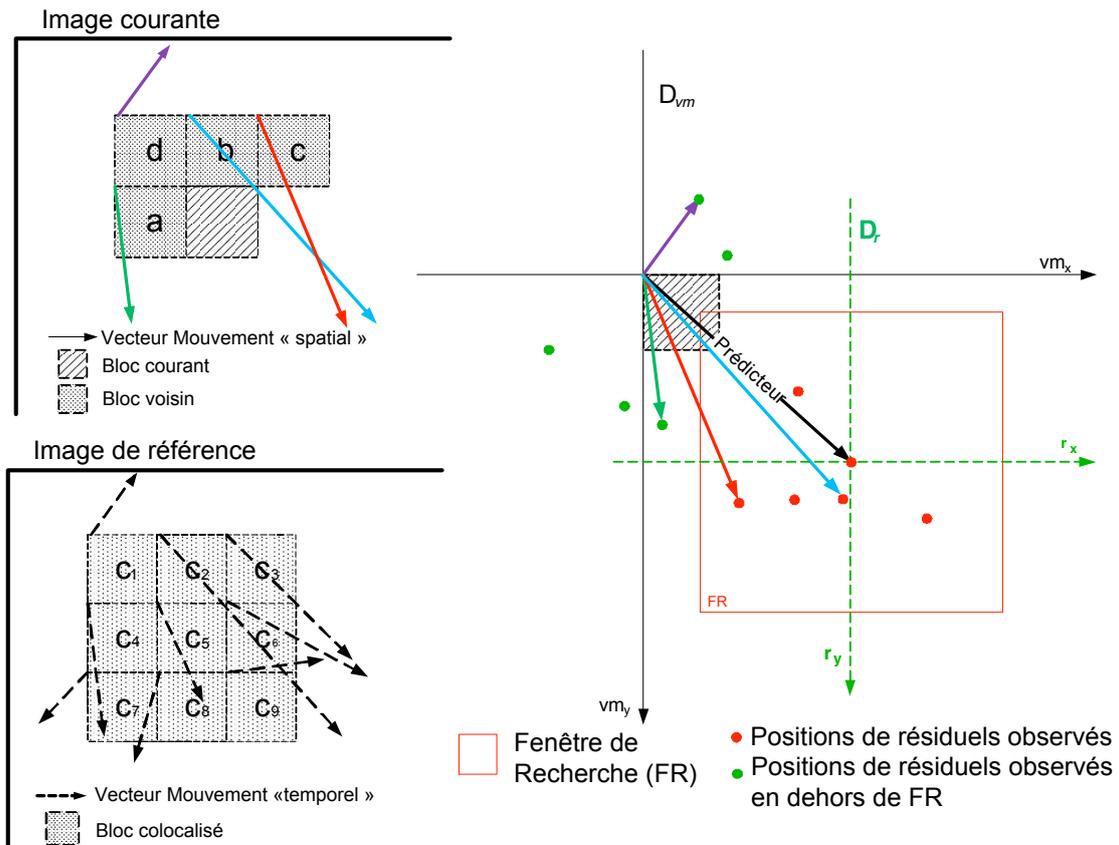


FIGURE 5.2: Exemple de distribution des résiduels observés pour un bloc donné.

En complément de ces vecteurs de mouvement précédemment codés, il est intéressant de noter que l'outil proposé dans ce chapitre permet l'ajout aisé de tout vecteur prédicteur suffisamment pertinent. Ainsi, certains prédicteurs proposés dans les travaux détaillés au chapitre 3 ne sont pas assez efficaces dans un cadre MV-Comp. En effet, même s'ils permettent de coder à moindre coût les vecteurs de blocs spécifiques, ils restent en moyenne moins intéressants que les prédicteurs classiques comme le médian spatial et le vecteur du bloc colocalisé du fait du surcoût de signalisation. A l'inverse, tous ces vecteurs trouvent leur place dans le cadre de l'outil MV-FMap dans la mesure où ils sont très utiles pour coder certains mouvements. C'est pourquoi nous ajoutons dans l'histogramme, deux vecteurs particuliers :

Vecteur translaté qui correspond au vecteur appartenant au bloc de l'image précédente pointé par le médian spatial des vecteurs voisins du bloc courant. Comme représenté sur la figure 5.3, le vecteur médian permet de déterminer le bloc translaté dont on utilise le vecteur mouvement (en rouge). Dans le cas d'un bloc appartenant à un ensemble (objet, personnage, ...) en mouvement, il permet de retrouver le vecteur du bloc correspondant au même ensemble au sein de l'image de référence.

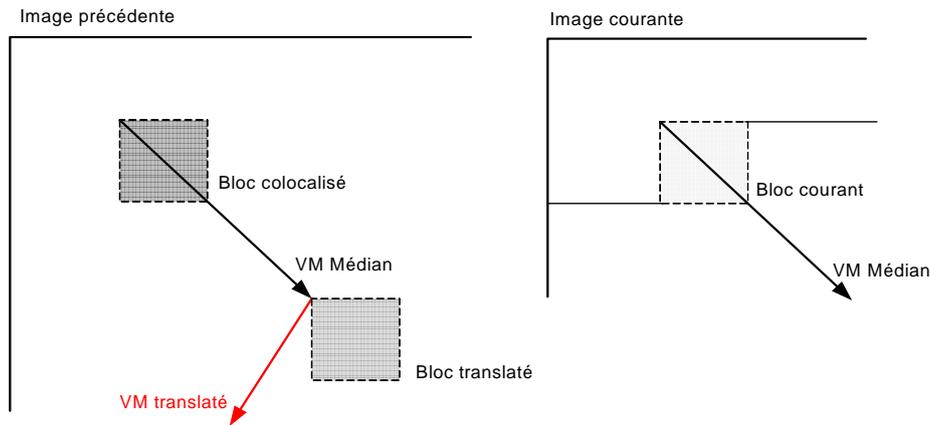


FIGURE 5.3: Représentation du vecteur traduit.

Vecteur colocalisé déplacé qui correspond au vecteur du bloc colocalisé prolongé selon sa propre direction d'un nombre prédéfini de positions δ à la résolution pixellique définie pour la séquence. Ce vecteur, en rouge sur la figure 5.4, est particulièrement pertinent pour anticiper un mouvement qui aurait débuté dans l'image précédente mais qui ne se retrouve pas encore parmi les blocs voisins. La valeur du δ est ajustée proportionnellement à la résolution de la séquence, une résolution élevée sera ainsi associée à une valeur élevée de δ .

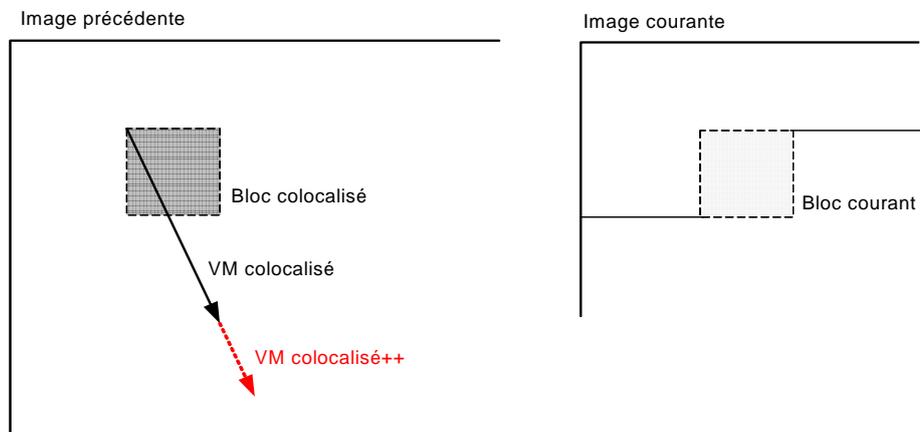


FIGURE 5.4: Représentation du vecteur colocalisé déplacé.

Enfin, insistons sur l'importance de prendre des précautions pour ajouter de nouveaux vecteurs dans l'histogramme. Comme nous le verrons dans les sections suivantes, la clé de la réussite de la méthode réside dans le compromis entre pertinence des vecteurs et nombre de vecteurs déplacés par rapport à la distribution originale. Un nombre trop élevé de vecteurs peu pertinents entraîne une modification trop importante de la distribution,

qui peut être fatale dans le cas d'un nouveau mouvement non pronostiqué.

A ce stade du schéma, l'information de mouvement est composée uniquement d'observations brutes des vecteurs précédemment codés. Dans le but de rendre cette information plus représentative du mouvement attendu, nous réalisons un pronostic précis de la probabilité d'apparition de l'ensemble des vecteurs résiduels. C'est l'objet de la sous-section qui suit.

5.1.3 Etape 2 : Affinage de l'information de mouvement

A partir de l'histogramme construit dans l'étape 1, le but de cette étape 2 est d'évaluer la probabilité d'apparition de chaque vecteur résiduel afin de privilégier les résiduels les plus probables au cours de l'étape 3. L'information brute que constitue l'histogramme est ainsi affinée pour obtenir un pronostic précis de l'ensemble des résiduels de la fenêtre de recherche FR . Dans cette optique, à tout bloc à coder b^i , nous associons la fonction de densité p^i qui attribue à chaque vecteur résiduel r , une probabilité d'apparition ($p = p^i(r)$) pour le bloc b^i , soit :

$$\begin{aligned} p^i : \quad FR &\longrightarrow [0, 1] \\ r(r_x, r_y) &\longmapsto p. \end{aligned} \quad (5.6)$$

5.1.3.1 Théorie de Parzen

Il s'agit à présent d'estimer (au sens statistique) p^i , à partir d'un échantillon de n mesures $\{x_1, x_2, \dots, x_n\}$ supposées indépendantes et identiquement distribuées, d'une variable aléatoire continue X , en dimension d . Nous choisissons l'estimateur de Parzen défini par (en respectant les notations de l'équation (11), page 64 de [Duda & Hart 73], à l'exception du paramètre h_n que nous notons l_n pour ne pas le confondre avec l'histogramme des vecteurs) :

$$p(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{(l_n)^d} \varphi\left(\frac{x - x_i}{l_n}\right), \quad (5.7)$$

où φ est la fonction *noyau*, appelée fenêtre de Parzen et l_n est donc un paramètre défini par $l_n = \frac{l_1}{\sqrt[n]{n}}$, où l_1 est le paramètre de *lissage*, définissant la taille de la fenêtre. Notons que pour le cas ($d = 1$) on retrouve bien l'équation (1.7) page 1066 de [Parzen 62].

Les degrés de liberté de cet estimateur sont donc φ et l_1 :

1. φ détermine la forme de l'influence des mesures. Remarquons que pour $\varphi = \chi_{\left\{B_{l_{inf}}(0, \frac{1}{2})\right\}}$, la fonction caractéristique de la boule fermée (en norme infinie) de rayon $\frac{1}{2}$, centrée à l'origine, on retrouve l'estimateur basique associé à l'histo-

gramme. Cet estimateur n'est pas continu, ce qui constitue un handicap rédhibitoire dans le cas de MV-FMap. En l'absence de connaissances a priori, on utilise classiquement la fonction de densité de la loi normale, centrée réduite, à savoir :

$$\varphi(u) = \left(\frac{1}{\sqrt{2\pi}} \right)^d \exp \left[-\frac{\|u\|^2}{2} \right]. \quad (5.8)$$

C'est notre choix pour MV-FMap où le domaine des vecteurs étant en deux dimensions, $d = 2$.

2. l_1 détermine le diamètre de la zone d'influence des mesures. Si l_1 est petit, chaque mesure force localement le résultat, indépendamment des autres. Si l_1 est grand, le résultat devient une moyenne pondérée par la distance de la contribution de l'ensemble des mesures.

La figure 5.5 (tirée de [Duda & Hart 73], page 171) illustre la théorie pour :

- $d = 1$.
- φ , la loi normale $\varphi(u) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{u^2}{2} \right]$.
- $l_1 \in \{0.2, 0.5, 1\}$ (noté h_1 sur la figure) .
- $n \in \{1, 16, 256, \infty\}$.

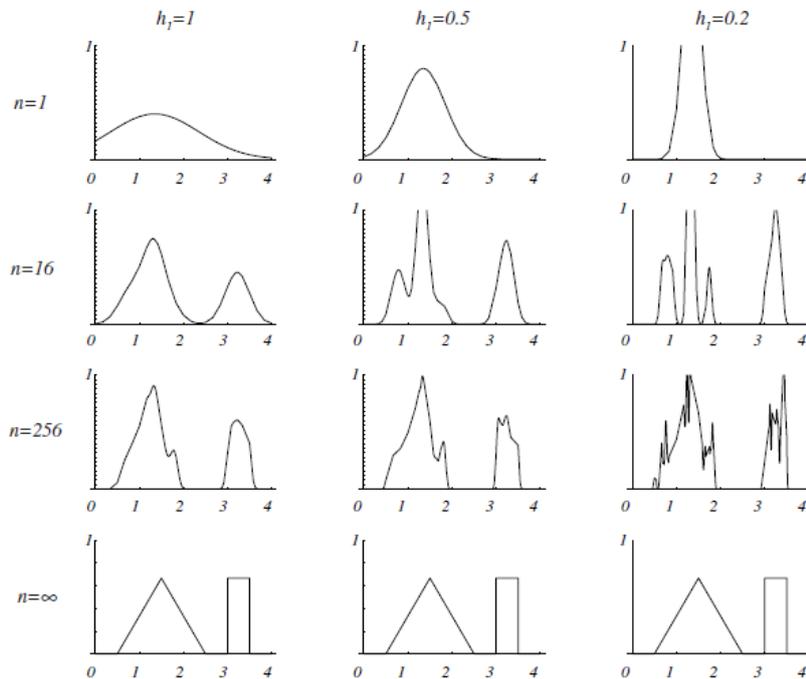


FIGURE 5.5: Exemple de densités de probabilité estimées par noyau pour différentes valeurs du paramètre de lissage (h_1) et différents nombres de mesures (n).

On y voit l'effet du paramètre de lissage qui, pour une valeur élevée ($h_1 = 1$) donne effectivement des courbes très étalées autour des mesures alors qu'avec une valeur faible ($h_1 = 0.2$), les courbes sont davantage centrées sur les mesures. Ce constat est atténué avec l'augmentation du nombre de mesures, la densité de probabilité estimée convergeant logiquement vers la densité de probabilité réelle lorsque n tend vers l'infini.

Nous prenons en compte ces observations, afin d'obtenir un pronostic très efficace des vecteurs résiduels. Concernant h_1 , nous privilégierons les valeurs faibles qui induisent des courbes multi-lobes plutôt étroites. Concernant n , contrairement au cas classique, un nombre élevé de vecteurs observés peu pertinents fait diverger l'estimateur de façon non-désirée pour MV-FMap ; c'est notamment le cas avec l'information de mouvement *globale* introduite précédemment. Ainsi, $n \approx 15$ semble être un bon compromis au vu de nos nombreuses expériences.

5.1.3.2 Application à notre méthode

En pratique, dans le cadre de MV-FMap, nous effectuons tout d'abord un changement des notations de [Duda & Hart 73] afin d'éviter toute confusion, à savoir :

- $p \longleftrightarrow p^i$, la densité de probabilité estimée.
- $x \longleftrightarrow r$, les résiduels.
- $n \longleftrightarrow N$, le nombre de mesures (on oublie l'exposant i pour alléger les notations).
- $l_1 \longleftrightarrow l$, le paramètre de lissage, soit : $l_N = \frac{l}{\sqrt{N}}$ (on oublie l'exposant i pour alléger les notations).
- $i \longleftrightarrow n$, l'indice pour la somme.

A partir de l'histogramme des résiduels (h_r^i) obtenu à l'étape 1, nous formons un échantillon de mesures $E_r^i = \{r \in FR \mid h_r^i(r) > 0\} = \{r_1, r_2, \dots, r_n\}$ (après renumérotation et par abus de notations). En appliquant directement la méthode de Parzen en dimension $d = 2$, nous définissons alors la densité de probabilité de l'ensemble des vecteurs résiduels $r \in FR$ comme suit :

$$p^i(r) = \frac{1}{2\pi l^2} \sum_{n=1}^N \exp \left[-N \frac{\|r - r_n\|^2}{2l^2} \right]. \quad (5.9)$$

Nous prenons finalement en compte le nombre d'occurrences des vecteurs résiduel pour donner plus de poids à ceux qui figurent dans plusieurs blocs voisins, et obtenons :

$$p^i(r) = \frac{1}{2\pi l^2} \sum_{n=1}^N \frac{h_r^i(r_n)}{S_{h_r^i}} \exp \left[-N \frac{\|r - r_n\|^2}{2l^2} \right], \quad (5.10)$$

où $S_{h_r^i} = \sum_{n=1}^N h_r^i(r_n)$ est un terme de normalisation.

Concernant le paramètre de lissage l (h_1 , sur la figure 5.5) qui est essentiel dans notre méthode, nous proposons de le déterminer adaptativement en le reliant à la distribution des résiduels observés lors de l'étape 1 à l'aide de la formule :

$$l = 0.3 + \frac{d_{moy}(h_r^i)}{\frac{L_{FR}}{2}}, \quad (5.11)$$

où L_{FR} représente la taille de la fenêtre de recherche et d_{moy} la distance moyenne des résiduels ($r \in E_r$) au centre de la fenêtre de recherche FR :

$$d_{moy}(h_r^i) = \frac{1}{N} \sum_{n=1}^N \|r_n\|_2. \quad (5.12)$$

Typiquement, pour une distribution des résiduels observés dite *serrée* (d_{moy} petit), l est faible et on obtient une distribution, centrée sur les mesures, aboutissant à un pronostic qui privilégie davantage les résiduels observés. Inversement, pour une distribution des résiduels observés dite *étendue* (d_{moy} grand), l est élevé et résulte en une distribution étalée des densités de probabilités qui favorise l'anticipation de nouveaux mouvements.

En conclusion, notre pronostic des résiduels les plus probables pour le bloc courant s'appuie sur la formule (5.10) avec l donné par (5.11). Triés par ordre décroissant de probabilité, les positions de résiduels sont alors modifiées conformément à l'étape 3 décrite ci-après.

5.1.4 Etape 3 : Redistribution des résiduels de vecteurs de mouvement

Le but de la troisième étape de MV-FMap est de définir une nouvelle distribution des résiduels de vecteurs de mouvement, attendus et non-attendus. Mathématiquement parlant il s'agit de définir une application bijective, dénotée Map^i qui, à une position de vecteur résiduel originale r associe une position résiduelle cible r^m , à savoir :

$$\begin{aligned} Map^i : \quad FR &\longrightarrow FR \\ r(r_x, r_y) &\longmapsto r^m(r_x^m, r_y^m). \end{aligned} \quad (5.13)$$

A l'issue des deux étapes précédentes nous définissons les sous-ensembles L_1^i , L_2^i , L_m^i de FR tels que :

- $L_1^i \cup L_2^i = FR$.
- le sous-ensemble L_1^i des vecteurs résiduel attendus est initialisé comme suit :

$$L_1^i = \{r \in FR \mid p^i(r) > \tau_p\},$$

où $\tau_p \in [0, 1]$ définit la valeur de probabilité à partir de laquelle un vecteur résiduel est considéré comme attendu. ($\tau_p = 0$) est une valeur par défaut, tandis que ($\tau_p > 0$) permet de limiter le nombre d'éléments de L_1^i et ainsi de modifier modérément la

distribution originale des résiduels. En choisissant $(\tau_p \mid \text{Card}(L_1^i) = N_a)$ on peut réguler le nombre N_a de vecteurs résiduel attendus à déplacer.

- le sous-ensemble L_2^i des vecteurs résiduel non-attendus est initialisé comme suit :

$$L_2^i = \{r \in FR \mid p^i(r) \leq \tau_p\} = \mathcal{C}_{FR}(L_1^i).$$

- le sous-ensemble L_m^i des vecteurs résiduel cibles est initialisé comme suit :

$$L_m^i = \{r^m \in FR\}.$$

Naturellement, à chaque bloc b_i , est associé une bijection Map^i qui diffère selon les résultats des étapes d'observation et de pronostic.

Cette initialisation (Phase 0) est suivie de deux phases successives de mise à jour :

Phase 1. Les résiduels attendus (L_1^i) sont associés à des positions cibles libres (L_m^i) de plus faible coût.

Phase 2. Les résiduels non-attendus (L_2^i) sont associés à des positions cibles libres (L_m^i), de coût, a priori, plus élevé. Un principe de prévoyance est appliqué ici pour minimiser les dégradations de la distribution originale afin qu'un nouveau mouvement non-pronostiqué soit traité au mieux malgré l'intervention de MV-FMap.

Ces deux phases s'articulent autour de l'ordre de parcours de L_1^i , L_2^i et des contraintes d'associations \mathcal{C}_j , $j \in \{1, 3\}$.

1. L'ordre de parcours.

La liste des vecteurs attendus L_1^i est triée pour satisfaire le préordre \succ ainsi défini : $r \succ r'$ (qui se lit « le résidu r est prioritaire sur le résidu ») si et seulement si $(p^i(r) > p^i(r'))$ ou alors $((p^i(r) = p^i(r'))$ et $(\|r\|_1 \leq \|r'\|_1)$.

La norme 1 étant définie par $\|r\|_1 = |r_x| + |r_y|$. Ce préordre n'est pas anti-symétrique pour en faire un ordre et il permet de définir des résidus de même priorité qui ne sont alors pas déplacés. En pratique, les vecteurs résiduels de plus fortes probabilités sont traités en premier pour être associés à des positions de très faible coût et lorsqu'ils sont équiprobables, c'est la norme qui les départage (si possible).

Pour la liste des vecteurs non-attendus L_2^i , seul le tri selon la norme entre en compte.

2. Les contraintes d'association.

De manière informelle, le but de ces contraintes est d'assurer l'efficacité de la méthode en répondant à deux questions :

- une association au cours de la phase 1 (position originale \leftrightarrow position cible de coût inférieur) est-elle pertinente ?
- une association au cours de la phase 2 (position originale \leftrightarrow position cible de coût minimal) est-elle pertinente ?

De manière formelle, trois contraintes se dégagent :

- Les contraintes $\mathcal{C}_j, j \in \{1, 2\}$:

$$\mathcal{C}_j : \begin{cases} \forall r \in L_j^i, \text{Map}^i(r) = (r^m \in L_m^i) \\ \Rightarrow (r^m = \text{arg min}_{\forall r^{m'} \in L_m^i} (\|r^{m'}\|_1)) \end{cases} \quad (5.14)$$

Ces deux contraintes assurent une distance minimale entre le vecteur résiduel cible et le centre de la fenêtre de recherche.

- La contrainte \mathcal{C}_3 :

$$\mathcal{C}_3 : \begin{cases} \forall r \in L_1^i, \text{Map}^i(r) = (r^m \in L_m^i) \\ \Rightarrow (|r_x| - |r_x^m| \geq 0) \text{ et } (|r_y| - |r_y^m| \geq 0) \end{cases} \quad (5.15)$$

\mathcal{C}_3 prend en compte les caractéristiques du codeur entropique CABAC en empêchant les associations d'un résiduel attendu à un résiduel cible dont l'une des composantes est supérieure à la composante correspondante du résiduel attendu. Cette contrainte est très restrictive mais permet d'empêcher un codage plus coûteux du résiduel par rapport à sa position initiale du fait du modèle de probabilité sélectionné par l'encodeur.

Au niveau algorithmique, Map^i est une table, L_1^i, L_2^i, L_m^i sont des listes et le processus global de création de Map^i est résumé comme suit :

Initialisation les listes L_1^i et L_2^i sont triées selon les ordres de parcours prédéfinis. La table d'association Map^i est vide.

Phase 1 en suivant l'ordre de parcours prédéfini, chaque résiduel attendu de la première liste $r \in L_1^i$ est associé à une position de résiduel cible disponible $r^m \in L_m^i$ qui vérifie les conditions suivantes :

$$\text{Map}^i(r) = r^m \Leftrightarrow \begin{cases} \mathcal{C}_1 \\ \mathcal{C}_3 \end{cases} \quad (5.16)$$

Après chaque association de résiduel, les deux listes sont mises à jour :

$$L_1^i \leftarrow L_1^i \setminus \{r\} \text{ et } L_m^i \leftarrow L_m^i \setminus \{r^m\}. \quad (5.17)$$

Phase 2 de la même manière, chaque résiduel non attendu $r \in L_2^i$ est associé à une position de résiduel cible restante qui vérifie la condition suivante :

$$\text{Map}^i(r) = r^m \Leftrightarrow \mathcal{C}_2 \quad (5.18)$$

Comme (5.17), les listes sont mises à jour après chaque association de résiduel :

$$L_2^i \leftarrow L_2^i \setminus r \text{ et } L_m^i \leftarrow L_m^i \setminus \{r^m\}. \quad (5.19)$$

La figure 5.6 présente un exemple d'association : les résiduels des listes L_1^i (en rouge) et L_2^i (en cyan) sont représentés sur la colonne de gauche et les positions de résiduels cibles de la liste L_m sont représentés sur la colonne de droite, les positions libres étant en bleu tandis que les positions occupées sont entourées en rouge. Les listes sont parcourues selon une forme en losange (cf. sphère en norme 1) centrée sur la position de plus faible coût (0,0). Enfin, les flèches rouges et cyans correspondent respectivement aux première et seconde phases d'associations.

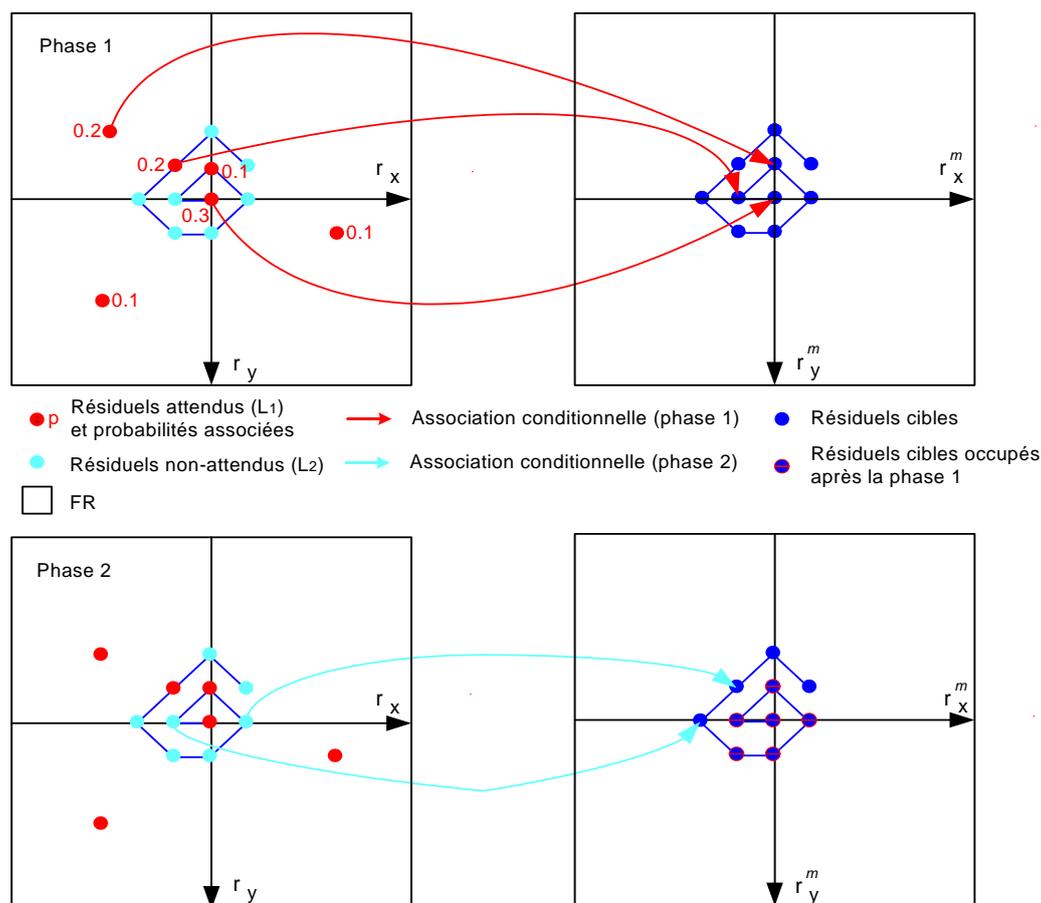


FIGURE 5.6: Exemple type de mise en œuvre des deux phases du processus d'association.

Avant de conclure la présentation de notre méthode, évoquons la question de complexité de la méthode. En ce qui concerne les deux premières étapes de collecte et de pronostic, leur complexité est clairement dépendante du nombre de vecteurs collectés. Il est ici nécessaire de trouver un compromis entre efficacité de la méthode et complexité. La dernière étape de modification de la distribution des vecteurs est pour sa part plus gênante dans la mesure où il faut parcourir lors des deux phases, les listes L_1^i puis L_2^i en parallèle de la liste L_m^i . Ceci est d'autant plus complexe que la fenêtre de recherche est

grande, en particulier pour les séquences de hautes résolutions. Nous avons par conséquent proposé de mettre à jour la table d'association Map^i au cours de l'estimation de mouvement afin de construire l'association uniquement pour les résiduels nécessaires.

5.2 Discussion sur l'intégration dans un encodeur HEVC

Avant de présenter les performances de la méthode proposée que nous avons évaluée dans un encodeur H.264/AVC, nous souhaitons discuter de son intégration au sein du schéma de codage qui se dessine dans HEVC. Comme cela a été introduit dans le premier chapitre, le codage du mouvement y est notamment constitué d'une compétition entre de multiples prédicteurs, du mode Merge, et finalement du codage du résiduel de mouvement à l'aide d'un codeur entropique CABAC. Nous évoquons ces trois points ci-dessous en nous concentrant tout d'abord plus précisément sur MV-Comp pour lequel nous avons proposé une combinaison qui sera évaluée dans la section résultats de ce chapitre.

5.2.1 Interaction avec la méthode de compétition de prédicteurs de mouvement (MV-Comp)

5.2.1.1 Comparaison des deux méthodes

Comme nous l'avons vu dans l'état de l'art, l'environnement proposé par l'outil MV-Comp est de plus en plus utilisé et étudié, notamment pour introduire de nouveaux prédicteurs de mouvement ou bien réaliser une adaptation des ensembles de prédicteurs. En particulier, le codage du mouvement dans le projet HEVC est basé sur un tel modèle. Un ensemble fini de prédicteurs de mouvement en compétition est ainsi considéré et nécessite ensuite de signaler le choix optimal au décodeur à l'aide d'un indice dont le coût dépend directement du nombre de prédicteurs en compétition. Soit vm_{pred_1} et vm_{pred_2} deux vecteurs de mouvement prédicteurs différents et vm le vecteur à coder pour le bloc courant, chacun de ces vecteurs correspond alors à un point potentiel sur lequel la distribution des résiduels de mouvement à coder va être centrée. A l'opposée, l'outil MV-FMap prend acte du prédicteur défini et vient corriger la possible erreur de prédiction en déplaçant les mouvements les plus attendus aux positions les moins coûteuses. Ainsi, si l'un des deux mouvements correspondants à vm_{pred_1} et vm_{pred_2} correspond effectivement au mouvement du bloc courant, les coûts de codage respectifs avec MV-Comp et MV-FMap seront équivalents. Autrement, MV-FMap tirera profit d'une grande variété de mouvements codables à moindre coût tandis que MV-Comp sera plus efficace pour coder un vecteur proche d'un des prédicteurs. On peut donc assimiler le déplacement d'un résiduel dans MV-FMap à un changement de prédicteur dans MV-Comp. Une vue d'ensemble de la comparaison des deux outils est présentée dans le tableau 5.1 pour résumer cette étude.

	MV-Comp	MV-FMap
Avantages	outil « simple » et peu complexe, exploitation des redondances spatiales et temporelles	pas de coût de signalisation additionnel, anticipation de nouveaux mouvements
Limites	coût de la signalisation additionnel	coût très élevé si erreur du pronostic, complexité non négligeable

Table 5.1: Résumé de la comparaison MV-Comp/MV-FMap.

A la lumière de ces observations, MV-Comp et MV-FMap semblent pouvoir cohabiter pour tirer profit des avantages des deux approches et c'est pourquoi nous proposons une combinaison détaillée ci-après.

5.2.1.2 Proposition de combinaison des méthodes

Notons tout d'abord qu'une combinaison partielle des deux outils consiste à exploiter les deux premières étapes de MV-FMap pour générer les N_{pred} meilleurs vecteurs prédicteurs à mettre en compétition dans le cadre de MV-Comp. Les prédicteurs correspondent alors aux vecteurs ayant le meilleur pronostic à l'issue de la seconde étape de MV-FMap. Cette combinaison n'utilise cependant pas la redistribution des résiduels qui apporte une réelle efficacité de codage.

Une première solution intuitive pour combiner effectivement les deux outils et profiter à la fois d'une prédiction pertinente et d'un codage efficace des résiduels consiste à définir un histogramme des vecteurs résiduels ($h_r^{i,j}$) pour chaque vecteur prédicteur vm_{pred_j} , $j \in \{1, \dots, N_{pred}\}$, N_{pred} étant le nombre de prédicteurs en compétition. Ces histogrammes ($h_r^{i,j}$) sont construits à partir d'un unique histogramme des vecteurs de mouvement (h_{vm}^i) comme défini dans (5.1), en effet celui-ci est indépendant des vecteurs prédicteur. Le changement de variable bi-équivoque de l'équation (5.5) devient $vm = r + vm_{pred_j}^i$, $\forall r \in D_r$ et il en résulte :

$$h_r^{i,j}(r) = h_{vm}^i(r + vm_{pred_j}^i), \forall r \in D_r = D_{vm - vm_{pred_j}^i}. \quad (5.20)$$

Les deux étapes suivantes de MV-FMap sont alors appliquées pour chacun des histogrammes.

Afin d'améliorer cette solution est de tirer davantage profit des caractéristiques des deux méthodes, nous proposons d'adapter les histogrammes des vecteurs de mouvement selon l'origine des prédicteurs vm_{pred_j} . Définissons $E_b^{i,j}$, $j \in \{1, \dots, N_{pred}\}$ l'ensemble de blocs précédemment codés liés au prédicteur vm_{pred_j} pour le bloc b_i . Typiquement, $E_b^{i,j}$ correspond à l'ensemble des blocs voisins spatiaux pour le vecteur prédicteur médian tandis qu'il correspond à l'ensemble des blocs voisins temporellement (dans l'image de référence), pour le vecteur prédicteur colocalisé. La mise à jour de l'histogramme, en référence à (5.2), devient alors :

$$h_{vm}^{i,j}(vm^k) \leftarrow h_{vm}^{i,j}(vm^k) + 1, \quad (5.21)$$

$\forall vm^k, k < i$, associé au bloc $b^k \in E_b^{i,j}$. La suite du schéma est alors identique.

Cette solution constitue la seconde combinaison que nous évaluerons dans la sous-section 5.3.2. Avant cela, nous évoquons dans la suite de cette section les interactions de MV-FMap avec le mode Merge et le codeur entropique CABAC.

5.2.2 Interaction avec le mode Merge

Le mode Merge [Marpe *et al.* 10] est un outil majeur pour le codage du mouvement du futur standard HEVC. Comme nous l'avons introduit précédemment (3.2.3), il consiste à hériter l'ensemble des paramètres de mouvement d'un blocItemize lorsque cela est pertinent. C'est une solution intuitive au problème inhérent à la représentation en arbre quaternaire de HEVC qui donne lieu à un surpartitionnement des images et donc au codage répété d'un même mouvement. Il étend le concept du mode Direct d'H.264/AVC et permet donc de ne pas transmettre de vecteurs en tant que tel au prix de la signalisation de la partition dont le mouvement est hérité. Dans la mesure où aucun résiduel de mouvement n'est transmis, il n'est pas possible d'envisager une combinaison avec notre méthode autre que l'exploitation de nos deux premières étapes de collecte et pronostic du mouvement qui permettent de fournir des candidats au mode Merge. Celui-ci est par contre un sérieux concurrent de MV-FMap en termes d'efficacité de codage car il permet de capter quasiment toutes les redondances dans le mouvement parmi l'ensemble des partitions voisines spatialement mais aussi temporellement. Ainsi, d'après les premiers tests réalisés, plus de 50% des partitions Inter utilisent le mode Merge dans HEVC et les résiduels de mouvement qui sont alors codés pour d'autres partitions correspondent principalement à de nouveaux mouvements pour lesquels MV-FMap n'est pas systématiquement efficace. On visualise sur la figure 5.7 les régions codées avec la même information de mouvement qui sont nombreuses et réduisent en effet le nombre de partitions pour lesquelles un vecteur est effectivement codé.

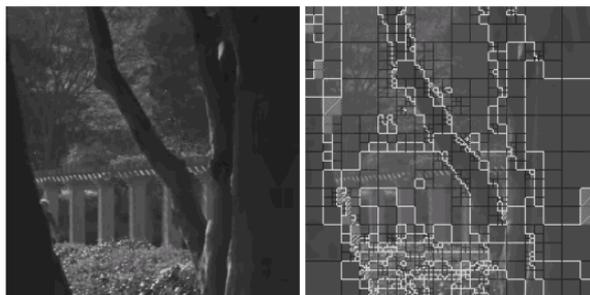


FIGURE 5.7: Sélection du mode Merge pour la séquence ParkScene tirée de [Marpe *et al.* 10]. Sur la figure de droite, les bordures blanches délimitent l'ensemble des partitions codées avec la même information de mouvement.

La cohabitation du mode Merge avec notre méthode semble par conséquent délicate mais il faut cependant noter qu'en biaisant ainsi le codage du mouvement, le mode Merge entraîne un manque de diversité dans le mouvement qui peut avoir un impact subjectif négatif, en particulier à bas débits. D'un point de vue débit-distorsion le mode Merge peut en effet être utilisé très fréquemment du fait de son faible coût alors que le mouvement hérité n'est pas le plus pertinent. Au contraire, MV-FMap qui autorise bien évidemment l'apparition de nouveaux mouvements, permet aussi de coder à moindre coût des faibles variations autour d'un mouvement principal pour capter ainsi plus efficacement les redondances temporelles du signal.

5.2.3 Interaction avec le codeur entropique CABAC

Le codeur entropique CABAC (en anglais, *Context Adaptive Binary Arithmetic Coder*) est déjà présent dans H.264/AVC [Marpe *et al.* 03] et nous en tenons compte lors de la dernière étape de MV-FMap (5.1.4) qui inclut une règle pour la redistribution des résiduels avec la contrainte \mathcal{C}_3 (5.15). Le même codeur entropique étant utilisé dans HEVC, nous revenons ici sur l'interaction entre CABAC et notre méthode. En effet, on peut voir une certaine redondance entre CABAC et notre méthode dans la mesure où le but de MV-FMap est de capter le mouvement potentiel du bloc courant à partir des blocs précédemment codés pour coder à moindre coût les mouvements les plus probables. CABAC exploite pour sa part les redondances entre les symboles à l'aide de multiples modèles de probabilités qui sont sélectionnés adaptativement en fonction des symboles déjà codés dans le voisinage puis mis à jour au fil du codage. Ainsi, le codeur entropique s'adapte pour coder à moindre coût les symboles les plus probables dans un certain voisinage et permet donc d'effectivement coder sur peu de bits les mouvements les plus probables dès lors qu'il choisit le bon modèle. Cependant, ce choix de modèle est limité et a tendance à biaiser la décision tandis qu'un choix de modèle non-optimal entraîne un surcoût de codage pénalisant.

Nous rappelons plus en détails comment les résiduels de mouvements sont traités par le codeur entropique CABAC qui considère séparément chaque composante (r_x, r_y) et on encode leur valeur absolue après avoir transmis le signe du résiduel selon les étapes suivantes décrites ici pour r_x :

1. Binarisation de $|r_x|$ qui utilise un mode code de type Exp-Golomb pour $|r_x| \geq 9$. Chaque bit du mode de code représente alors respectivement le bin 1, 2, ...
2. Sélection d'un contexte pour chaque bin à coder en fonction des vecteurs résiduels r^A et r^B des blocs déjà codés de gauche et du dessus. Trois modèles catégorisant le type de mouvement sont ainsi possibles pour le premier bin selon la valeur de $e_k = \|r^A\|_1 + \|r^B\|_1$:
 - Modèle 0 si $0 \leq e_k < 3$.
 - Modèle 1 si $3 \leq e_k < 33$.
 - Modèle 2 si $33 \leq e_k$.

Les bins suivants se voient pour leur part affecter un unique contexte parmi quatre autres modèles de façon indépendante du voisinage.

3. Encodage des bins en fonction du contexte sélectionné pour chaque bin qui fournit la probabilité de la valeur « 0 » ou « 1 » utilisée par le codeur arithmétique lui-même.
4. Mise à jour des contextes.

Ainsi, dans CABAC, les vecteurs précédemment codés interviennent d'une part dans la sélection du modèle pour le codage du premier bin et d'autre part dans l'estimation des probabilités de chaque contexte qui sont mises à jour au fil du codage. Les probabilités étant conservées tout au long du codage d'un slice, cela correspondrait lors de la première étape de MV-FMap à réaliser l'histogramme des vecteurs pour l'ensemble d'une slice donc potentiellement pour plusieurs images, ce qui n'est pas le plus efficace comme nous le verrons dans les premiers résultats de la section suivante. En effet, tous les mouvements au sein d'une slice ne sont pas nécessairement corrélés et peuvent par conséquent interférer entre eux et limiter l'efficacité du codeur entropique CABAC. Au contraire, MV-FMap est particulièrement pertinent dans la mesure où il permet de capter un mouvement très local tout en étant efficace pour le codage d'un mouvement plus global. De plus, l'étape de redistribution des résiduels qui permet d'exploiter les données observées de façon très précise constitue elle aussi un point clé qui est difficilement exploitable dans CABAC même en augmentant le nombre de contextes.

A ce titre, nous présenterons dans la section suivante les gains significatifs obtenus par notre méthode avec le codeur entropique CABAC qui confirment ces observations. Il serait cependant nécessaire d'envisager une prise en compte plus judicieuse du comportement des deux outils pour optimiser les gains. Par exemple, un changement inapproprié de contexte pour le codage du premier bin des composantes du vecteur résiduel peut en effet résulter en une perte d'efficacité, c'est ce que nous cherchons à limiter avec la règle de mapping \mathcal{C}_3 mais qui peut néanmoins se produire.

5.3 Résultats expérimentaux

5.3.1 Evaluation de notre méthode

5.3.1.1 Configuration de tests

La configuration utilisée pour l'ensemble des tests correspond au profil High IPPP d'H.264/AVC avec l'ensemble des outils activés. Nous avons évalué les performances de MV-FMap en comparaison avec cette référence dans le cadre de l'appel à proposition (CfP) HEVC constitué de séquences de résolutions et contenus très variés. En outre, nous avons testé la méthode pour 4 séquences additionnelles présentant des caractéristiques de mouvement intéressantes pour notre étude. Cette configuration est résumée ci-dessous :

<p>Codeur H.264/AVC, JM KTA version 2.1.</p> <p>Profil High IPPP .</p> <p>Modes Intra, Skip et Inter 16×16 à 4×4.</p> <p>Mouvement Estimation de mouvement rapide EPZS, prédicteur Median spatial, représentation du mouvement au quart de pixel.</p> <p>Codeur Entropique CABAC.</p> <p>QPs Débit moyen si non-indiqué et haut, moyen et bas débits sinon.</p> <p>Séquences Séquences de l'appel à proposition (CfP) MPEG HEVC (QWVGA à 1080p) et 4 séquences caractérisées par un mouvement complexe.</p> <p>Evaluation Pourcentage de réduction de débit moyen calculé par la métrique de Bjontegaard [Bjontegaard 01, Jung & Pateux 07].</p>
--

5.3.1.2 Analyse préliminaire

Dans un premier temps, nous proposons d'évaluer l'impact des différents éléments composants les trois étapes de la méthode définis précédemment : 1) collecte de l'information de mouvement, 2) affinage de l'information de mouvement et 3) modification de la distribution des résiduels de mouvement. Six configurations détaillées ci-dessous ont été testées de façon incrémentale (i.e. : la configuration du Test 2 correspond à celle du Test 1 enrichie d'une modification). Nous pouvons ainsi percevoir l'apport de chaque modification sur l'efficacité de la méthode de manière globale ou bien pour chaque type de résolution de QWVGA à 1080p. Les résultats sont rapportés dans le tableau 5.2 pour l'ensemble des configurations suivantes :

- Test 1** configuration basique de MV-FMap où les vecteurs observés par l'histogramme correspondent à l'ensemble des vecteurs précédemment codés dans l'image courante et dans l'image précédente. La seconde étape n'est pour sa part pas réalisée, et les vecteurs observés lors de la première étape sont traités indépendamment de leur fréquence d'apparition lors de la redistribution des résiduels (étape 3). Les vecteurs résiduels (tels que $h_i(r) > 0$) les plus proches du centre sont ainsi déplacés en premier.
- Test 2** configuration modifiant la première étape de collecte en restreignant les vecteurs observés dans l'histogramme aux vecteurs des quatre blocs voisins spatiaux et des neuf blocs colocalisés.
- Test 3** configuration ajoutant la prise en compte de la fréquence d'apparition lors de l'étape de redistribution des résiduels. A fréquence d'apparition égale, les résiduels les plus proches du centre de la fenêtre de recherche sont toujours traités en premier.
- Test 4** configuration activant la seconde étape de la méthode permettant d'affiner l'information de mouvement collectée. Une densité de probabilité est affectée à chaque

résiduel de la fenêtre de recherche à partir de l’observation d’un nombre restreint de vecteurs selon la méthode de Parzen. Le paramètre de lissage est ici fixé à $l = 0.7$.

Test 5 configuration modifiant le calcul du paramètre de lissage l utilisé pour la méthode de Parzen dans la seconde étape. Le paramètre est déterminé adaptativement selon une loi dépendant de la distribution des résiduels observés et plus précisément de la distance entre les vecteurs comme définie dans la formule 5.11 dans la sous-section 5.1.3.

Test 6 configuration ajoutant deux nouveaux vecteurs dans l’histogramme des vecteurs observés lors de la première étape : le vecteur translaté et le vecteur colocalisé déplacé (introduits dans la sous-section 5.1.2).

De manière générale on observe, dans le tableau 5.2, des gains significatifs pour l’ensemble des tests qui évoluent de 0.9% de réduction du débit moyen pour le Test 1 à 3.0% pour le Test 6. Notons de plus que les gains pour les séquences de résolution QWVGA sont sensiblement inférieurs aux gains des autres résolutions du fait de la petite taille des séquences qui implique des mouvements moins amples pour lesquels notre méthode permet potentiellement des gains moins élevés. Ces derniers atteignent cependant un gain significatif de 2.1% pour la configuration du Test 6.

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
QWVGA	0.2	1.2	1.5	1.8	2.0	2.1
WVGA	0.8	2.8	2.7	3.1	3.4	3.7
720p	1.1	2.2	2.4	2.5	2.9	3.2
1080p	1.4	2.0	2.2	2.8	3.0	3.1
Moyenne	0.9	2.0	2.2	2.6	2.8	3.0

TABLE 5.2: Comparaison de différentes configurations de notre méthode pour les différentes résolutions des séquences du CfP HEVC. Pourcentage de gain moyen sur le débit.

Au cas par cas, nous observons tout d’abord l’écart important (plus de 1%) entre le Test 1 et les tests suivants qui confirme l’intérêt de collecter les vecteurs des blocs les plus proches spatialement et temporellement plutôt que l’ensemble des vecteurs précédemment codés. L’histogramme local utilisé à partir du Test 2 permet ainsi d’avoir des vecteurs à déplacer plus pertinent pour le bloc courant qui permettent de réduire significativement le coût du vecteur résiduel à coder.

L’importance du compromis entre pertinence et quantité de vecteurs déplacés est clairement mis en avant ici dans la mesure où un nombre très élevé de vecteurs peu pertinents comme dans le Test 1 apporte clairement moins de gains que les autres tests.

Notons cependant le cas particulier d'une vidéo parmi les séquences de résolution 1080p : BQTerrace atteint un gain de 3.5% avec la configuration du Test 1 qui est un gain significativement supérieur aux gains dans les autres configurations et en particulier dans la dernière configuration qui contient l'ensemble des outils (Test 6) et où l'efficacité de la méthode est seulement de 2.5%. Ce cas particulier s'explique par le contenu de la séquence qui correspond à un lent mouvement de caméra temporellement continue qui est adapté à une approche plus globale. Cette observation ouvre ainsi la voie à une adaptation de l'ensemble des vecteurs observés selon le contenu de la séquence. Entre le Test 2 et le Test 3, on s'aperçoit que la prise en compte de la fréquence d'apparition lors de la redistribution des résiduels apporte une amélioration moyenne de 0.2% qui confirme qu'il faut donner plus d'importance à certains vecteurs très attendus. Le Test 4 permet de poursuivre sur cette piste avec l'intégration de l'étape 2 basée sur la méthode de Parzen qui permet d'affecter une densité de probabilité à l'ensemble des vecteurs de la fenêtre pour ensuite redistribuer les résiduels par probabilité décroissante. Un gain significatif est ainsi obtenu par rapport au Test 3 avec une amélioration de 0.4% encore plus notable sur les séquences de haute résolution 1080p qui sont particulièrement adaptées à la méthode de Parzen du fait de la plus grande diversité de vecteurs qui permettent par conséquent une bonne estimation de la densité de probabilité des résiduels de la fenêtre de recherche. Le perfectionnement de cette méthode de Parzen dans le Test 5 améliore de 0.2% le gain du test précédent.

L'adaptation du paramètre d'étalement des vecteurs en fonction de la distribution est en effet très important dans notre recherche du meilleur compromis entre quantité et pertinence des vecteurs.

En particulier, les gains pour les séquences de résolution 720p augmentent en moyenne de 0.4% grâce à cette modification. Ces séquences correspondent exclusivement à des contenus de type vidéo-conférence qui contiennent un fond fixe avec des personnes en mouvement, il est donc important que MV-FMap modifie peu la distribution des vecteurs pour les zones fixes ou ayant un mouvement faible et inversement qu'un nombre élevé de vecteurs soient anticipés pour les blocs avec du mouvement. Un tel résultat est obtenu avec l'adaptation du paramètre d'étalement alors qu'au contraire ce paramètre restant fixe dans le Test 4 le type de bloc n'était pas pris en compte. Une perspective de perfectionnement de la méthode serait d'ailleurs d'améliorer encore la qualité de ce paramètre en réalisant un apprentissage permettant d'adapter encore plus précisément sa valeur selon le contenu de la séquence et de chaque bloc.

Finalement, nous avons ajouté de nouveaux vecteurs dans l'histogramme avec le Test 6 ce qui permet d'augmenter encore l'efficacité de la méthode qui atteint un gain moyen de 3.0% sur l'ensemble des séquences.

Cela confirme un avantage majeur de la méthode proposée qui permet d'ajouter des vecteurs candidats pertinents sans intervenir sur la signalisation comme le nécessite l'outil MV-Comp.

Ces vecteurs peuvent ainsi réduire le coût de codage du résiduel de façon ponctuel en anticipant efficacement un nouveau mouvement tandis que dans les autres cas où leur effet serait négatif s'il fallait signaler le prédicteur choisi (augmentation significative du coût de signalisation), l'effet est ici atténué par le biais de l'étape 2 qui affecte une faible probabilité aux résiduels correspondants qui sont par conséquent déplacés suffisamment loin lors de la redistribution.

5.3.1.3 Performances de la méthode

Dans la suite de l'étude nous avons retenu la configuration du Test 6 pour présenter et analyser en détail les résultats dans la mesure où c'est la configuration la plus efficace en moyenne. Le tableau 5.3 donne les pourcentages de réduction du débit de chaque séquence du CfP ainsi que d'autres séquences sélectionnées pour leur mouvement difficile à coder. Les résultats sont présentés pour trois types de débits différents afin d'évaluer le comportement de la méthode proposée selon le débit. Relevons tout d'abord que MV-FMap apporte systématiquement un gain par rapport au codeur de référence H.264/AVC.

En moyenne, des gains significatifs sont obtenus et confirment la pertinence de la méthode avec respectivement 1.9%, 3.0% et 4.3% de réduction du débit obtenus pour les hauts, moyens et bas débits.

Les gains plus élevés pour les bas débits sont expliqués par la proportion plus importante de l'information de mouvement dans le flux total. En particulier, le meilleur résultat est obtenu à ce débit pour la séquence *Vidyo4* 720p avec 6.4% de réduction du débit moyen tandis que le gain le plus faible est obtenu à haut débit pour la séquence *Raven* 720p avec 0.2% de gains. Nous analyserons plus en détail le comportement de notre outil sur cette séquence particulière dans la sous-section suivante où l'on se compare à l'outil MV-Comp. Pour le gain maximal obtenu avec la séquence *Vidyo4* 720p, précisons qu'il s'agit d'une séquence type vidéo-conférence qui possède par conséquent un fond fixe et où le mouvement de la personne est complexe et donc coûteux, MV-FMap est alors particulièrement efficace pour anticiper les mouvements futurs et en réduire leur coût. Par rapport aux gains selon la résolution des séquences, nous observons seulement que pour les petites résolutions (QWVGA et CIF) l'efficacité de la méthode est logiquement plus limitée du fait de la proportion moindre du mouvement. Les gains n'augmentent cependant pas avec les résolutions et les meilleurs sont obtenus ici en WVGA avec 3.7% et 4.8% en moyenne respectivement à moyen et bas débits. Cela est en partie lié au contenu des séquences mais s'explique aussi par le fameux compromis délicat entre quantité et pertinence des vecteurs récoltés qui semble optimal en WVGA.

	Haut débit	Moyen débit	Bas débit
CfP HEVC			
QWVGA			
BasketBallPass	1.7	2.4	3.0
BlowingBubbles	1.5	2.2	2.7
BQSquare	1.1	1.3	1.5
RaceHorses	1.4	2.4	3.6
WVGA			
BasketBallDrill	2.7	4.0	5.8
BQMall	3.3	4.6	5.3
PartyScene	2.0	2.7	2.7
RaceHorses	1.7	3.4	5.5
720p			
Vidyo1	2.6	3.9	4.8
Vidyo3	0.7	1.4	2.1
Vidyo4	2.2	4.2	6.4
1080p			
BasketBallDrive	1.7	3.7	5.9
BQTerrace	2.0	2.5	3.2
Cactus	2.7	4.0	5,2
Kimono1	1.3	2.8	4.6
ParkScene	1.2	2.3	3.7
Moyenne CfP	1.9	3.0	4.1
Séquences additionnelles			
Foreman CIF	1.9	3.3	4.4
Ice CIF	2.9	3.9	5.0
Nuts5 WVGA	2.7	4.2	6.1
Raven 720p	0.2	1.2	3.5
Moyenne Totale	1.9	3.0	4.3

TABLE 5.3: Pourcentage de gains sur le débit de notre méthode MV-FMap dans la configuration optimale (Test 6) pour trois catégories de débits cibles. Application à l'ensemble des séquences du CfP HEVC et à 4 séquences additionnelles.

5.3.1.4 Analyse

Afin d'analyser les résultats précédents, nous proposons tout d'abord d'étudier la classification des vecteurs codés à l'aide de MV-FMap. Nous définissons quatre catégories de résiduels codés en fonction de leur appartenance ou non à l'information de mouvement

(IM) et le cas échéant selon leur origine :

IM temporelle le vecteur codé fait partie des vecteurs pronostiqués et il provient de l'un des blocs voisins temporellement (blocs de l'image de référence).

IM spatiale le vecteur codé fait partie des vecteurs pronostiqués et il provient de l'un des blocs voisins spatialement.

IM spatiale et temporelle le vecteur codé fait partie des deux catégories précédentes.

Non pronostiqué le vecteur codé ne fait pas partie des vecteurs pronostiqués.

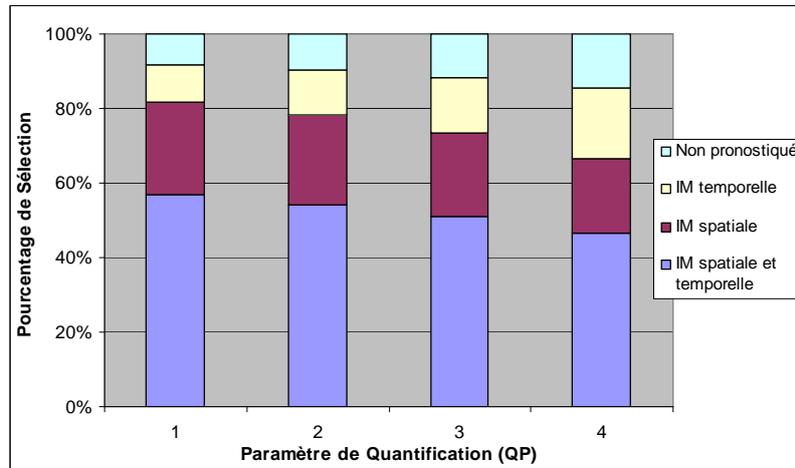
La figure 2.6 donne l'appartenance des résiduels à ces quatre catégories en pourcentage de l'ensemble vecteurs résiduel codés. Nous présentons tout d'abord la répartition moyenne pour l'ensemble des séquences du CfP sur la figure (a), puis le détail pour deux séquences particulières (b et c). Enfin, les pourcentages sont fournis pour quatre niveaux de quantification allant d'une faible (« QP 1 ») à une forte (« QP 4 ») quantification.

La première observation est que le vecteur codé a en moyenne été pronostiqué par MV-FMap dans 89% des cas comme on le voit sur la figure (a). Ce résultat est très important et confirme l'intérêt de la méthode :

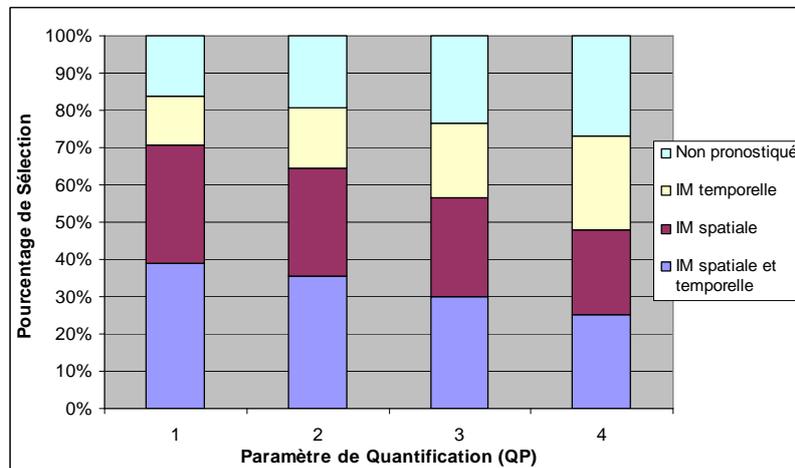
dans 89% des cas, le vecteur est codé à moindre coût ou à coût équivalent lorsque la prédiction du vecteur elle-même est suffisamment efficace.

Ce chiffre évolue selon le pas de quantification et atteint notamment 92% pour des faibles quantifications (« QP 1 ») car les redondances du mouvement sont alors mieux exploitées. Pour les deux autres figures obtenues pour deux séquences particulières, le pronostic est globalement plus délicat pour la séquence *BasketBallDrill* (b) notamment à bas débit (« QP 4 ») avec seulement 73% de vecteurs correctement pronostiqués du fait du contenu de la séquence qui contient de nombreux nouveaux mouvements difficilement prévisibles. Inversement, la séquence *BQSquare*, dont nous reparlerons après, contient peu de mouvement et les vecteurs y sont pronostiqués quasi-systématiquement. Notons cependant que les gains présentés dans le tableau 5.3 ne sont pas directement liés à ces chiffres, il est en effet nécessaire de considérer le coût global du mouvement dans chaque séquence pour avoir une bonne interprétation. Il faut de plus tenir compte des cas où notre méthode se combine mal avec le codeur entropique pour conclure sur cette analyse mais les chiffres observés ici demeurent très significatifs.

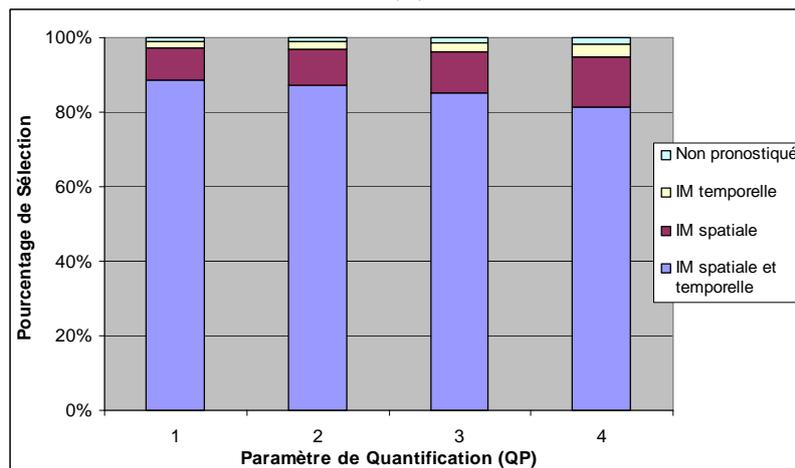
Nous observons maintenant la répartition des vecteurs correctement pronostiqués selon leur bloc d'origine. Ainsi, en moyenne 52% des vecteurs proviennent à la fois de blocs voisins spatialement et temporellement ce qui correspond aux nombreuses zones d'une séquence où le mouvement est très faible. Ceci est confirmé par la figure (c) obtenue avec la séquence *BQSquare* correspondant à un lent zoom arrière et une activité temporelle quasi-nulle, plus de 80% des vecteurs pronostiqués proviennent alors de cette catégorie et ont peu d'impact sur l'efficacité de codage puisque c'est la séquence avec les gains les plus faibles observés dans le tableau 5.3.



(a)



(b)



(c)

FIGURE 5.8: Classification de l'origine des vecteurs résiduels codés (en pourcentage) selon le niveau de quantification. Evaluation moyenne pour l'ensemble des séquences du CfP (a) ; pour la séquence *BasketBallDrill* WVGA (b) ; et la séquence *BQSquare* QWVGA (c).

La seconde catégorie la plus représentée (23% environ) correspond aux vecteurs provenant de blocs voisins uniquement spatialement. Ce chiffre, additionné aux 52% précédents, permet de « valider » l'utilisation de prédicteurs de mouvement spatiaux en priorité tel que le font les codeurs H.264/AVC et HEVC, les redondances spatiales dans le mouvement sont en effet les mieux exploitables dans le sens où, contrairement au bloc colocalisé, les blocs voisins sont très fréquemment associés au bloc courant. La proportion de vecteurs codés issus de l'IM temporelle n'est cependant pas négligeable avec 14% en moyenne et finit de valider notre approche. Observons d'ailleurs que pour la séquence *BasketBallDrill* (b) qui est l'une des séquences où MV-FMap est le plus performant, la proportion de l'IM spatiale et temporelle s'inverse à bas débit où l'IM temporelle prend alors une place plus importante car les blocs voisins spatialement ne fournissent plus forcément assez de vecteurs diversifiés et pertinents pour ce QP.

Ainsi, un très large ensemble de redondances dans le mouvement est capté par notre outil et dès lors qu'un mouvement un peu complexe nécessitant une fine utilisation des vecteurs de l'image de référence apparaît, MV-FMap permet d'améliorer significativement le codage des vecteurs correspondants.

Lorsque la séquence contient de nombreuses zones de ce type, l'efficacité de codage globale de la séquence en est d'autant plus augmentée.

Finalement, la figure 5.9 montre l'évolution du gain en bits pour le codage d'un vecteur en fonction du déplacement moyen au quart de pixel effectué à l'étape 3. Pour évaluer le gain nous avons simulé le codage des vecteurs avec la méthode de référence et MV-FMap tandis que le déplacement moyen correspond à la distance en norme 1 entre le vecteur original et le vecteur pronostiqué par MV-FMap, le tout évalué en moyenne pour le codage d'une séquence. Nous observons principalement que l'amélioration du codage augmente avec le déplacement réalisé, un déplacement d'une amplitude de 70 positions au quart de pixel faisant ainsi gagner quasiment 20 bits tandis que les faibles déplacements (<10) font gagner entre 1 et 5 bits en moyenne. Cette évolution n'est cependant pas linéaire du fait du comportement du codeur entropique CABAC et cela confirme l'importance des conditions définies pour l'étape 3 qui permettent de limiter les effets non-désirés. Enfin, alors que les courbes des deux séquences ont la même évolution jusqu'à un déplacement moyen de 20, on observe que les déplacements élevés apportent davantage de gains (2 bits environ) pour la séquence *PartyScene*. En effet cette séquence est plutôt composée de mouvements de faibles amplitudes. Par conséquent l'apparition d'un nouveau mouvement différent à un coût élevé avec le codage classique tandis que MV-FMap permet effectivement de réduire ce coût. La séquence *BasketBallDrill* étant davantage composée de mouvements amples, le gain apporté par MV-FMap est moindre mais reste significatif.

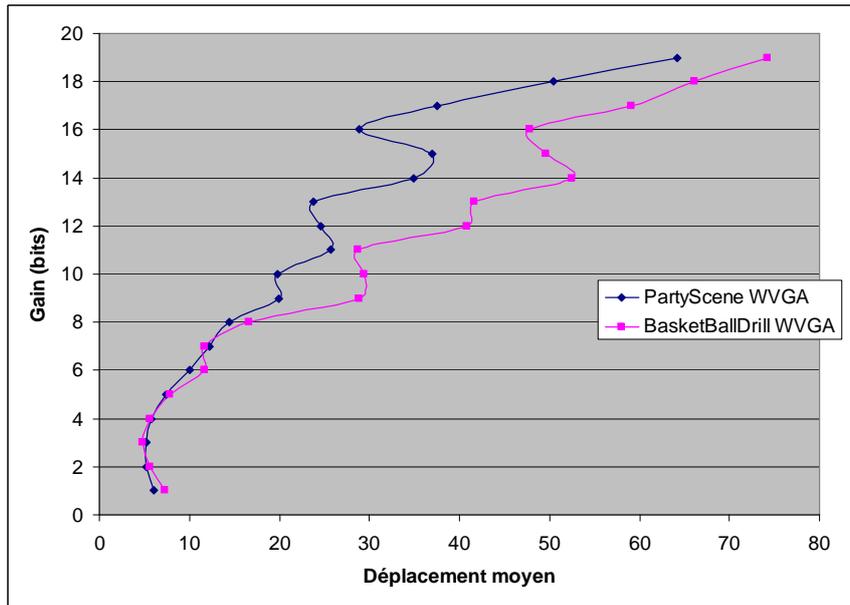


FIGURE 5.9: Evolution du gain de codage des vecteurs résiduels (en bits) en fonction du déplacement moyen au quart de pixel réalisé lors de l'étape 3 (évalué par la norme 2). Résultats obtenus avec les séquences *PartyScene* (bleu) et *BasketBallDrill* (rose) WVGA pour le QP=32.

L'outil MV-FMap a donc prouvé son efficacité par rapport au schéma de codage de référence tel que défini dans H.264/AVC. Dans la sous-section suivante, nous présentons les résultats de l'étude de l'outil par rapport à MV-Comp.

5.3.2 Evaluation de notre méthode et l'outil MV-Comp

La configuration utilisée pour les tests présentés dans cette sous-section est identique à celle des tests précédents pour l'ensemble de QP correspondant aux moyens débits.

Dans le tableau 5.4 nous avons rapporté les résultats de notre outil MV-FMap qui est tout d'abord comparé avec l'outil MV-Comp. Pour cet outil, nous utilisons la configuration optimale définie par les auteurs : pour les vecteurs, une compétition entre deux prédicteurs, le médian spatial et le vecteur colocalisé, ainsi qu'une compétition pour le mode Skip entre le médian spatial étendu et le vecteur colocalisé. Précisons en effet que l'outil MV-Comp tel qu'introduit dans [Jung & Laroche 06] modifie le mode Skip tandis que MV-FMap ne peut pas être appliqué en tant que tel au mode Skip dans la mesure où aucun vecteur de mouvement résiduel n'y est codé. Par conséquent, nous avons testé MV-FMap combiné à l'outil MV-Comp lorsque ce dernier est appliqué uniquement au mode Skip (Combinaison 1). Les deux schémas peuvent alors être comparé directement avant d'essayer de combiner les deux outils pour le codage des vecteurs de mouvement comme proposé dans 5.2.1.2 (Combinaison 2).

	MV-Comp	MV-FMap	Combinaison 1	Combinaison 2
CfP HEVC				
QWVGA				
BasketBallPass	1.5	2.4	3.1	2.0
BlowingBubbles	1.1	2.2	2.2	1.8
BQSquare	0.2	1.3	1.0	0.8
RaceHorses	1.2	2.4	2.3	2.2
WVGA				
BasketBallDrill	0.3	4.0	4.0	1.9
BQMall	3.4	4.6	5.7	4.3
PartyScene	1.9	2.7	2.5	2.9
RaceHorses	3.3	3.4	3.9	4.7
720p				
Vidyo1	4.0	3.9	6.1	5.0
Vidyo3	-0.6	1.4	1.7	0.3
Vidyo4	5.3	4.2	3.7	6.3
1080p				
BasketBallDrive	4.9	3.7	4.4	5.5
BQTerrace	3.5	2.5	3.4	4.4
Cactus	1.9	4.0	4.1	2.8
Kimono1	2.3	2.8	2.8	3.4
ParkScene	4.0	2.3	3.6	4.8
Moyenne CfP	2.4	3.0	3.4	3.3
Séquences additionnelles				
Foreman CIF	3.3	3.3	3.1	4.6
Ice CIF	2.1	3.9	4.0	2.8
Nuts5 WVGA	1.4	4.2	1.0	2.2
Raven 720p	9.4	1.2	2.7	10.5
Moyenne Totale	2.7	3.0	3.3	3.6

TABLE 5.4: Pourcentage de gains sur le débit des méthodes MV-Comp et MV-FMap ainsi que de deux combinaisons des méthodes. Tests réalisés à moyens débits.

Commentons tout d'abord les résultats moyens obtenus sur les séquences du CfP avant d'analyser plus en détails les séquences additionnelles qui présentent un intérêt particulier pour cette comparaison des deux outils.

MV-FMap est plus efficace que MV-Comp de 0.6% en moyenne pour les séquences du CfP et il est très performant pour les faibles résolutions (QWVGA et WVGA) où des gains de plus de 1% sont obtenus.

On observe globalement que les performances de MV-FMap sont plus stables (pas de pertes ni de gains très forts) alors qu'au contraire l'outil MV-Comp est de plus en plus efficace avec l'augmentation de la résolution et s'avère d'ailleurs meilleur que notre méthode de 0.2% pour les séquences 1080p. Ceci s'explique par le délicat compromis entre pertinence et nombre de vecteurs dans MV-FMap qui se complexifie d'autant plus lorsque la variété des vecteurs augmente comme c'est le cas pour les hautes résolutions. On observe le même type de comportement général sur la première combinaison qui inclut la compétition de prédicteurs pour le mode Skip. Dans ce test où l'on peut alors réellement comparer les deux méthodes, l'efficacité de MV-FMap est particulièrement mise en avant puisque seulement deux séquences (*Vidyo4* 720p et *BQTerrace* 1080p) obtiennent de meilleurs gains avec l'outil MV-Comp. La compétition sur les prédicteurs du mode Skip apporte ainsi jusqu'à 2.0% de gains additionnels combinée avec MV-FMap pour les vecteurs mais n'est pas systématiquement pertinente.

Enfin, la combinaison des deux outils pour le codage des vecteurs améliore les gains de MV-FMap de 0.3% en moyenne ce qui est plutôt décevant lorsqu'on considère la complexité qui est alors quasiment doublée par rapport à la complexité de MV-FMap seul.

Des gains particulièrement importants sont cependant atteints avec cette combinaison pour certaines séquences comme *Vidyo4* qui fait état d'une réduction du débit moyen s'élevant à 6.3%.

Attardons nous maintenant sur les autres séquences testées et en particulier au cas de la séquence *Raven* 720p pour laquelle MV-Comp est particulièrement efficace avec 9.4% de gains alors que MV-FMap apporte seulement 1.2% seul et 2.7% dans la combinaison 1. Ceci s'explique par le contenu de la séquence dont le mouvement est double avec le déplacement de la caméra d'une part et celui d'un oiseau d'autre part. Dans la mesure où ces deux mouvements sont nettement identifiables, MV-Comp est particulièrement adapté à ce contenu car la prédiction est alors quasiment parfaite. Inversement, bien que notre solution pronostique correctement l'ensemble des mouvements, ils ne sont pas clairement dissociés et la redistribution n'apporte par conséquent pas une amélioration aussi significative du codage. La seconde combinaison tire logiquement profit de l'outil MV-Comp tandis que MV-FMap permet de corriger certaines difficultés de prédiction du vecteur et cela apporte finalement un gain très élevé de 10.5% par rapport à la référence.

En conclusion, la combinaison des deux outils semble pertinente pour certains contenus au mouvement difficile mais nécessite d'être davantage étudiée pour exploiter de façon plus efficace les atouts des deux approches.

5.4 Bilan

Dans ce chapitre, nous avons proposé une approche originale pour améliorer le codage des vecteurs de mouvement qui constitue une partie essentielle des codeurs vidéo

modernes. Le but de l'approche est d'exploiter au maximum les corrélations entre les vecteurs de mouvement d'images et de blocs successifs afin de corriger a posteriori l'erreur de prédiction des vecteurs. Cela est réalisé sans transmettre d'information de signalisation additionnelle, en modifiant la distribution des résiduels afin de favoriser les mouvements les plus probables. Le schéma est divisé en trois étapes de collecte, pronostic et enfin redistribution des résiduels.

Chaque étape a été configurée afin d'obtenir les meilleurs résultats en moyenne sur un large ensemble de séquences de test. La méthode permet ainsi d'obtenir un gain moyen de 3.0% ce qui prouve son efficacité par rapport au schéma de codage du mouvement d'H.264/AVC basé sur une prédiction du vecteur par médian spatial. Grâce à l'exploitation des redondances à la fois spatiales et temporelles dans le mouvement, quasiment 90% des vecteurs codés sont ainsi correctement pronostiqués et donc potentiellement déplacés à une position moins coûteuse. En comparaison avec la méthode par compétition de prédicteurs qui a été intégrée dans le futur standard HEVC, notre méthode est de nouveau plus efficace avec un gain moyen de plus de 0.6% mais accompagné d'une complexité plus élevée. La combinaison des deux méthodes, bien que prometteuse, est limitée avec un gain moyen de 3.3% qui s'élève cependant à plus de 4.0% en moyenne pour les séquences de grandes résolutions où le mouvement peut alors être très efficacement traité.

Deuxième partie

Approches en rupture basées sur l'exploitation d'un décodeur intelligent

Chapitre 6

Transfert d'intelligence au décodeur et échantillonnage compressé : un état de l'art

En rupture avec les schémas de codage conventionnels, l'approche que nous explorons dans la seconde partie de cette thèse permet de transférer une partie de l'intelligence au niveau du décodeur afin d'économiser, en accord avec l'encodeur, le codage de certaines informations. L'idée du transfert de la complexité au niveau du décodeur n'est pas nouvelle : elle est mise en œuvre dans le cadre du codage vidéo distribué qui constitue un paradigme de compression. Certains travaux abordent aussi cette approche dans un cadre de codage vidéo plus classique. Ainsi, dans la première section de cet état de l'art, nous regroupons une introduction au codage vidéo distribué et l'étude plus détaillée de travaux s'en inspirant dans un cadre de codage classique.

Le chapitre est ensuite scindé en deux à la manière du chapitre d'introduction de la première partie avec pour seconde section une présentation de la théorie émergente de l'échantillonnage compressé. Ce bref état de l'art constitue surtout un pré-requis pour notre étude introduite dans le chapitre 8 où nous aborderons l'intégration d'une méthode d'échantillonnage compressé dans un schéma de compression vidéo. Une telle mise en œuvre nécessitant un nombre élevé de paramètres, nous les traiterons à l'aide de notre approche décodeur intelligent.

6.1 Etat de l'art sur l'exploitation d'un décodeur intelligent

6.1.1 Codage vidéo distribué

Dans les schémas de codage, dits classiques, étudiés jusqu'à présent dans cette thèse, l'encodeur réalise l'ensemble des choix de codage en exploitant les redondances spatiales et temporelles puis il transmet un flux au décodeur. Celui-ci se contente alors de lire séquentiellement les informations de codage et d'appliquer les consignes pour décoder le signal vidéo. Toute la complexité calculatoire (possiblement très élevée en fonction des choix d'implémentation) se trouve ainsi concentrée au niveau de l'encodeur tandis que le décodeur, normalisé, est très léger en termes de calcul et on peut alors parler de décodeur passif. A contre-courant de ces schémas classiques, un nouveau paradigme de compression, le *codage vidéo distribué*, se développe depuis plus d'une décennie. La motivation derrière ce paradigme est justement de transférer toute la complexité calculatoire au niveau du décodeur. Il trouve sa place en termes applicatifs avec le développement récent de systèmes légers (tels que les téléphones portables ou les caméras de surveillance) nécessitant d'encoder des vidéos avec une complexité restreinte tout en assurant une qualité satisfaisante lors du décodage.

La mise en œuvre de cette approche date du début des années 2000 mais repose sur des activités de recherche en théorie de l'information des années 1970 avec d'une part les travaux de Slepian et Wolf [Slepian & Wolf 73] et d'autre part ceux de Wyner et Ziv [Wyner & Ziv 76]. En résumé, ces travaux (théorèmes à l'appui) prouvent que sous certaines conditions, l'encodage de deux sources corrélées peut être réalisé de façon identique, conjointement ou indépendamment, sans entraîner de perte d'efficacité du point de vue de la transmission, dans la mesure où le décodage, lui, est effectué de façon conjointe.

La complexité calculatoire des encodeurs conventionnels venant principalement de l'estimation de mouvement, ces résultats théoriques apportent une solution prometteuse permettant de déporter cette opération au niveau du décodage dans le cadre du codage vidéo distribué. Celui-ci considère ainsi la compression de multiples sources corrélées mais ne communiquant pas les unes avec les autres, les corrélations inter-images étant alors exploitées uniquement au niveau du décodeur. Les premières réalisations pratiques ont été proposées respectivement dans [Puri & Ramchandran 02] avec le codeur PRISM et dans [Aaron & Girod 02, Aaron *et al.* 03] avec le codeur de Stanford. Ce dernier, représenté sur la figure 6.1, traite alternativement deux types d'images :

- les *images clefs*, codées de façon indépendante à l'aide d'un codeur Intra traditionnel.
- les *images Wyner-Ziv*, traitées à l'aide d'une transformée suivie d'une quantification avant d'être transmises à l'aide d'un codeur canal (turbo-code).

Au niveau du décodage, les images clefs sont décodées indépendamment alors que pour le

décodage des images Wyner-Zyvv, une information adjacente obtenue à l'aide des images voisines (clefs ou Wyner-Zyvv) est exploitée. Cette information adjacente est typiquement générée par extrapolation à partir d'images précédentes ou bien par interpolation en utilisant une image clef dans le futur.

Malgré des performances théoriques prometteuses, cette technique tarde cependant à s'approcher de l'efficacité des schémas de codage conventionnels. Ceci est dû au non-respect de certaines des conditions des théorèmes et aux nombreuses années d'avance de la compression classique et cela sera encore plus vrai avec l'arrivée de HEVC. Le codage vidéo distribué trouve néanmoins sa place dans le cadre de certaines applications et continue de mobiliser une large partie de la communauté [Dufaux *et al.* 09] qui y voit notamment un fort intérêt pour le codage multivues [Maugey 10].

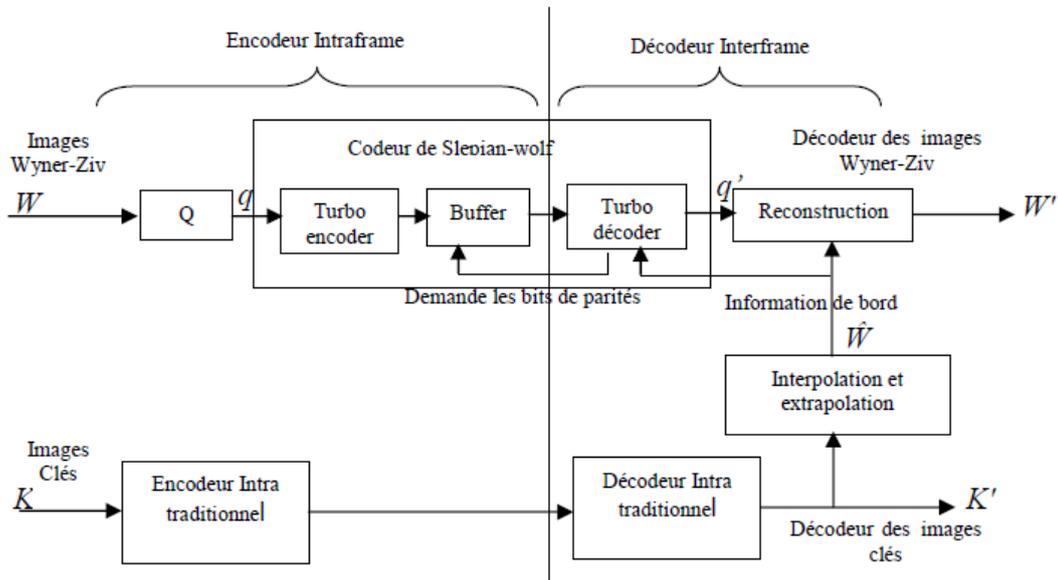


FIGURE 6.1: Schéma de codage vidéo distribué.

Du fait du transfert intégral de la complexité au décodeur, le codage vidéo distribué diffère de l'approche que nous proposons dans le chapitre suivant. Motivés par l'amélioration de l'efficacité de codage, nous proposons de faire évoluer les schémas de codage conventionnels vers un partage plus équitable des ressources entre codeur et décodeur. Les activités en codage vidéo distribué constituent néanmoins une source d'inspiration potentielle pour notre étude. Dans la sous-section qui suit, nous « visitons » plusieurs travaux pionniers de cette approche.

6.1.2 Applications au schéma de codage classique

Les travaux réalisés sur le codage vidéo distribué ont eu d'une certaine manière un écho au sein de la communauté du codage vidéo classique : diverses études ont vu le jour au cours des dernières années pour exploiter davantage le décodeur. Ce transfert d'opérations au niveau du décodeur, rendu possible par l'évolution des performances matérielles, constitue une solution majeure pour surmonter les limites probables des schémas de codage actuels. Ces derniers maximisent en effet la qualité de la prédiction de la texture par l'exploitation des redondances spatiales et temporelles à l'aide de multiples choix de codage. Cette compétition nécessitant d'être de plus en plus poussée, les performances de codage globales sont intrinsèquement bornées du fait des coûts de signalisation engendrés et de la marge de gain de plus en plus faible sur le codage de la texture.

Examinons à présent les principaux travaux réalisés sur ce sujet du transfert d'intelligence au niveau du décodeur.

6.1.2.1 Approches par « Template Matching »

Abordons tout d'abord les nombreuses études menées sur la base du *template matching* que ce soit dans un cadre de prédiction Intra ou Inter. Le principe commun à ces approches est l'utilisation des pixels de la zone causale proches du bloc à coder qui sont disponibles à la fois lors du codage et du décodage.

En Intra, l'idée de départ, telle qu'introduite dans [Tan *et al.* 06], consiste à proposer en tant que prédicteur, un bloc dont la forme constituée de pixels causaux voisins est la plus proche des pixels situés à la même position par rapport au bloc courant. Formellement, soit ψ_c la forme cible constituée de pixels causaux voisins de la partition courante, on détermine alors au sein de la fenêtre de recherche causale Ψ , la forme source ψ_s^* minimisant un critère d'appariement entre les deux formes, par exemple la somme des carrés des différences (SSD) pixel par pixel :

$$\psi_s^* = \arg \min_{\psi_s \in \Psi} SSD(\psi_s, \psi_c). \quad (6.1)$$

Cette méthode est issue de travaux de synthèse et restauration d'images qui exploitent les similarités au sein de la même image. De nombreuses autres variantes ont fait suite pour améliorer la qualité de cette première solution de prédiction, alors imparfaite. Citons [Tan *et al.* 07] des mêmes auteurs où une recherche des N formes sources $\psi_{s_i}^*$, $i \in \{1, \dots, N\}$ optimales selon le critère (6.1) est réalisée avant de générer une prédiction sous forme de moyenne des $\psi_{s_i}^*$, $i \in \{1, \dots, N\}$. Sur le même principe, il est proposé dans [Moinard *et al.* 10, Moinard 11] de transmettre un indice signalant la forme optimale au sein d'un ensemble. Cette solution apporte des gains en Intra de plus de 4% en moyenne malgré l'ajout d'une signalisation supplémentaire. Cela confirme que la similarité entre les formes $\psi_{s_i}^*$ et ψ_c n'est pas suffisante pour assurer que les blocs associés sont de bons

prédicteurs du bloc courant mais il existe néanmoins de très bon prédicteurs potentiels dans la zone causale. Notre approche, présentée au chapitre suivant, s'appuie sur cette conclusion.

En Inter, les solutions d'estimation de mouvement, conjointes au codeur et au décodeur, constituent une voie de recherche importante pour la réduction du coût de codage des vecteurs de mouvement comme nous l'avons mentionné dans le chapitre 3 (3.2). Nous complétons cette analyse en re-visitant ces méthodes sous l'angle du transfert de l'intelligence au niveau du décodeur. L'idée forte de ces schémas, notamment introduits dans [Sugimoto *et al.* 04, Kamp *et al.* 08, Kamp *et al.* 09], est de réaliser une estimation de mouvement à l'identique au niveau de l'encodeur et du décodeur pour ne plus transmettre d'information de mouvement lors de l'encodage. A l'aide de pixels de la zone causale suffisamment proches du bloc à prédire, un appariement est réalisé de manière identique à l'encodeur et au décodeur pour fournir un bloc prédictor dans l'image de référence. La figure 6.2 (extraite des travaux de thèse de G. Laroche [Laroche 09]) illustre une réalisation d'une telle méthode avec une zone causale utilisée pour le critère d'appariement définie par les quatre lignes de pixels au-dessus et les quatre colonnes de pixels à gauche du bloc courant. Les résultats encourageants d'une telle méthode, intégrée au codeur H.264/AVC, en compétition avec une transmission classique des vecteurs de mouvement (1% de réduction du débit moyen pour l'ensemble de QP 22-37) ont notamment amené le groupe JCT-VC à réaliser une étude approfondie de l'intégration d'un outil basé sur une estimation de mouvement au niveau du décodeur dans le futur standard HEVC. Cependant, sans aborder la question délicate de l'implémentation matérielle, ces travaux montrent pour l'instant leurs limites du fait de la complexité élevée ajoutée au niveau du décodeur pour des résultats insuffisamment significatifs, comparés à d'autres outils moins complexes. En particulier, notons que les blocs pour lesquels la méthode est la plus efficace correspondent aux blocs très corrélés avec leurs voisins. Or dans cette même configuration la prédiction du vecteur de mouvement est alors également souvent efficace. Loin de fermer la porte, de tels travaux prouvent que le recours à des opérations plus complexes au niveau du décodeur demeure une voie de recherche essentielle pour les futures améliorations des codeurs vidéo.

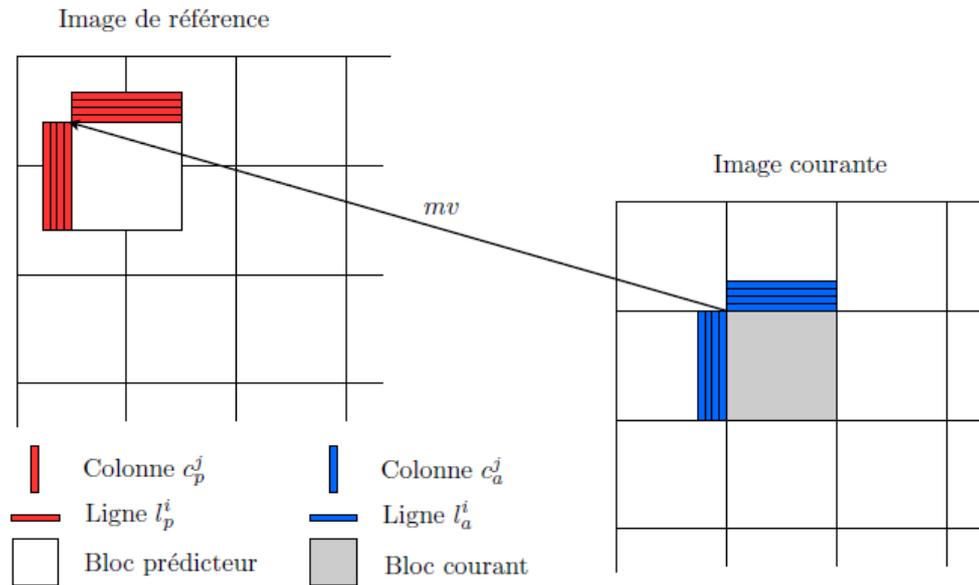


FIGURE 6.2: Exemple de représentation de l'estimation de mouvement au niveau du décodeur.

6.1.2.2 Traitement des informations de compétition

Au-delà de cette activité importante autour du *template matching*, d'autres travaux plus isolés ont abordé la problématique de la réduction du coût des informations de compétition. Dans [Laroche *et al.* 10], les auteurs proposent de réduire le coût des prédicteurs Intra en définissant adaptativement la taille de l'ensemble de prédicteurs en compétition pour chaque macrobloc. Ce traitement est réalisé à la fois au codeur et au décodeur et permet de supprimer astucieusement les prédicteurs redondants de manière à conserver uniquement les prédicteurs les plus complémentaires. Le prédicteur optimal étant alors signalé parmi un nombre réduit de modes, son coût est efficacement diminué et cela impact directement sur la qualité de la prédiction Intra puisque le codeur peut alors allouer davantage de ressources aux résiduels de texture. En pratique, cette suppression de prédicteurs est réalisée en étudiant leur proximité dans le domaine transformé en prenant en compte le pas de quantification afin de s'adapter aux différents débits. Des gains significatifs sont présentés comparés au profil Baseline d'H.264/AVC avec 3.4% de réduction du débit moyen mais ces gains sont moindres avec le meilleur profil d'H.264/AVC qui utilise des tailles de bloc 8×8 très sélectionnés alors que la méthode est plus efficace pour les blocs de petites tailles (4×4) où les redondances dans les prédicteurs sont en effet nombreuses. Malgré la disparition possible des plus petites tailles de blocs, cette approche conserve toutefois un intérêt avec la multiplication des prédicteurs Intra en compétition dans le futur standard HEVC (34 au lieu de 9 dans H.264/AVC). Surtout, les gains entrevus confirment le potentiel de méthodes réduisant

le coût de signalisation des informations de compétition tout en limitant la diminution de la qualité de la prédiction. Ce travail est une généralisation astucieuse de la méthode introduite dans [Kim *et al.* 08] où il est décidé adaptativement d'utiliser un seul prédicteur (DC) ou bien l'ensemble des modes de prédiction d'H.264/AVC selon un critère pouvant être évalué au codeur et au décodeur. Ce critère reflète simplement que lorsque les pixels du voisinage utilisés pour la prédiction sont quasiment similaires, il est inutile de mettre en compétition l'ensemble des prédicteurs Intra puisqu'ils seront sensiblement équivalents à l'utilisation du prédicteur DC définie par la valeur moyenne des pixels voisins.

Toujours par rapport aux modes Intra qui présentent un fort intérêt en termes de gain, le travail réalisé dans [Negusse 08] cherche à économiser le coût du prédicteur Intra en proposant une détermination adaptative du prédicteur optimal uniquement à partir des pixels causaux. En réalisant une prédiction entre différents niveaux de voisinages, la méthode permet en effet d'économiser la signalisation du prédicteur. Même si le prédicteur ainsi déterminé n'est pas systématiquement optimal, le gain en signalisation libère suffisamment de ressources pour le codage des résiduels de prédiction additionnels et améliore de façon non négligeable l'efficacité de codage globale.

6.1.2.3 Applications en codage multivues

Enfin, ce type de technique peut particulièrement trouver sa place dans le codage multivues qui permet d'exploiter davantage de corrélations. Ainsi, dans le cadre du codage vidéo multivue plus profondeur, les auteurs de [Shimizu *et al.* 11] ont très récemment proposé de déterminer différents paramètres de codage en utilisant en tant qu'image originale une vue synthétisée prédite générée par un outil tel qu'introduit dans [Yea & Vetro 08]. Cette vue étant générée identiquement lors du codage et du décodage, il n'est alors plus nécessaire de transmettre les informations de signalisations (mode, partitionnement, informations de mouvement) et cela engendre une économie significative. Mise en compétition avec les modes de codage classique, cette technique apporte des gains de plus de 10% par rapport au standard MVC. Elle est cependant limitée à bas débit par la qualité de la vue synthétisée qui est alors sensiblement éloignée de l'image originale et entraîne par conséquent des erreurs lors de la détermination des paramètres de codage combinés alors avec des résiduels plus élevés pour compenser la prédiction de moindre qualité. En termes de complexité, codeur et décodeur sont alors comparables puisque lors du décodage il faut répéter l'ensemble de la compétition pour chaque bloc codé avec ce mode en plus de la génération de la vue synthétisée par rapport à laquelle les auteurs proposent aussi de se positionner au mieux relativement au voisinage causal du bloc courant.

Ces contributions dévoilent une vaste voie de recherche très prometteuse que nous avons explorée au cours de ces travaux de thèse et qui seront introduits dans le chapitre suivant. L'approche que nous proposons sera dans un second temps appliquée à l'échantillonnage compressé et c'est pourquoi un bref état de l'art est consacré à cette théorie

émergente dans la section suivante.

6.2 Echantillonnage compressé : un bref état de l'art

6.2.1 Théorie

L'échantillonnage compressé (en anglais Compressed Sensing ou Compressive Sampling, noté CS) est une récente théorie du traitement du signal développée par E.J. Candès [Candès & Wakin 08, Candès 06]. Elle offre la possibilité de capturer un signal à de hautes résolutions à partir d'un nombre limité de mesures, très en-dessous du nombre défini par la théorie de Nyquist/Shannon [Nyquist 24, Shannon 01]. Le théorème de Shannon, à la base du traitement du signal depuis de nombreuses années, impose que l'acquisition des échantillons d'un signal à bande limitée soit réalisé à la cadence d'au moins le débit défini par Nyquist, équivalent à deux fois la fréquence maximale du signal considéré. Après cette acquisition qui est utilisée par la majorité des systèmes numériques modernes, les algorithmes de compression, qui exploitent les nombreuses redondances de ces signaux, interviennent afin de faciliter le stockage et la transmission des sons, images et vidéos. Tout ceci représente un temps de traitement significatif d'une quantité élevée de données brutes dont la majorité sera supprimée par l'étape de compression. L'échantillonnage compressé propose une alternative à ces deux limites : acquisition et compression sont réalisées simultanément grâce à une connaissance a priori sur la propriété de parcimonie du signal (définie ci-après).

Le problème considéré est la reconstruction d'un vecteur $f \in \mathbb{R}^N$ à partir d'un vecteur $y \in \mathbb{R}^M$ de mesures de f obtenu par projection du signal original sur un ensemble de fonctions de base $\{\varphi_m\}$, $m \in \{1, 2, \dots, M\}$:

$$y_m = \langle f, \varphi_m \rangle. \quad (6.2)$$

Soit, en notation matricielle :

$$y = \Phi f,$$

où Φ est la matrice de la base de mesures, ou base incomplète d'échantillonnage, de taille $M \times N$, où les M lignes de Φ sont $\varphi_i^T = [\varphi_i(1)\varphi_i(2)\dots\varphi_i(N)]$.

Nous nous intéressons aux cas de sous-échantillonnages où le nombre de mesures disponibles M est plus petit que la dimension N du signal f . Ce cas de figure est très fréquent en pratique, par exemple lorsque le nombre de capteurs est limité ou lorsque les mesures sont très coûteuses comme dans les techniques d'acquisition par diffusion de neutrons. Pour toutes ces applications, la question est :

*est-il possible de reconstruire correctement le signal à partir de $M \ll N$ mesures
uniquement ?*

La réponse est en général non, dans la mesure où ce problème de reconstruction est mal-posé. Cependant, avec les hypothèses additionnelles que le signal f est suffisamment parcimonieux et que la base Φ et la base utilisée pour la reconstruction, notée Ψ , sont incohérentes, une reconstruction parfaite est possible : c'est le principe fondamental de l'échantillonnage compressé.

Nous détaillons ces deux propriétés de départ ci-dessous :

Parcimonie Un signal $f \in \mathbb{R}^N$ est parcimonieux dans une base Ψ de taille $S \times S$, $\Psi = [\psi_1 \psi_2 \dots \psi_S]$, s'il peut être représenté par un petit nombre $S \ll N$ d'éléments x_i résultats de la projection de f sur Ψ :

$$f = \Psi x = \sum_{i=1}^S x_i \psi_i, \text{ avec } S \ll N.$$

On parle de signal S -parcimonieux.

Incohérence Nous avons défini la base Φ utilisée pour l'acquisition du signal f et Ψ la base utilisée pour la représentation de f . La cohérence entre ces deux bases est définie par :

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \leq k, j \leq n} |\langle \varphi_k, \psi_j \rangle|.$$

La cohérence mesure donc la plus grande corrélation entre deux des éléments de Φ et de Ψ . Logiquement, nous avons $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$, avec une valeur élevée si les deux bases ont des éléments corrélés, faible sinon. Dans le cadre de l'échantillonnage compressé les deux bases doivent être incohérentes, en d'autres termes les vecteurs de mesures $\{\varphi_m\}$ constituent une représentation dense des éléments $\{\psi_m\}$ de la base Ψ . Ceci est vérifié par plusieurs combinaisons de bases couramment utilisées en échantillonnage compressé. Citons notamment, les combinaisons fonctions de Dirac et base de Fourier, base d'ondelettes et noiselettes. Enfin, notons un résultat important qui est qu'une base orthonormée Φ créée à partir d'une matrice aléatoire est incohérente avec n'importe laquelle des bases Ψ précédemment fixée.

Ces définitions ainsi posées, la réponse à la question énoncée précédemment revient à résoudre le problème suivant :

$$\hat{f} = \arg \min_f \|\Psi f\|_{l_0}, \text{ tel que } y = \Phi f, \quad (6.3)$$

où la norme l_0 est définie par $\|f\|_{l_0} = \text{Card}\{i, \text{ tel que } f_i \neq 0\}$.

Ainsi on cherche la reconstruction \hat{f} qui a la plus petite norme l_0 dans le domaine Ψ tout en étant consistante avec la mesure y . Cependant, ce problème est très délicat à résoudre d'un point de vu combinatoire et c'est donc sa relaxation convexe qui est utilisée [Donoho 06] :

$$\hat{f} = \arg \min_f \|\Psi f\|_{l_1}, \text{ tel que } y = \Phi f, \quad (6.4)$$

où la norme l_1 est définie par $\|f\|_{l_1} = \sum_{i=1}^N |f_i|$.

Ce problème est résolu par des algorithmes de poursuite de base (en anglais *Basis Pursuit*, BP) qui demeurent complexe en terme de calcul en pratique.

A ce stade, nous pouvons résumer le fonctionnement de l'échantillonnage compressé à l'aide de la figure 6.3 : le signal f est échantillonné par l'opérateur d'acquisition Φ , le vecteur de mesures y est ensuite transmis puis décodé (l'ajout d'un bruit de quantification peut résulter de ces deux étapes), le vecteur de mesures y est finalement utilisé par l'algorithme de résolution du problème 6.3 permettant de reconstruire le signal \hat{f} dans une base Ψ .

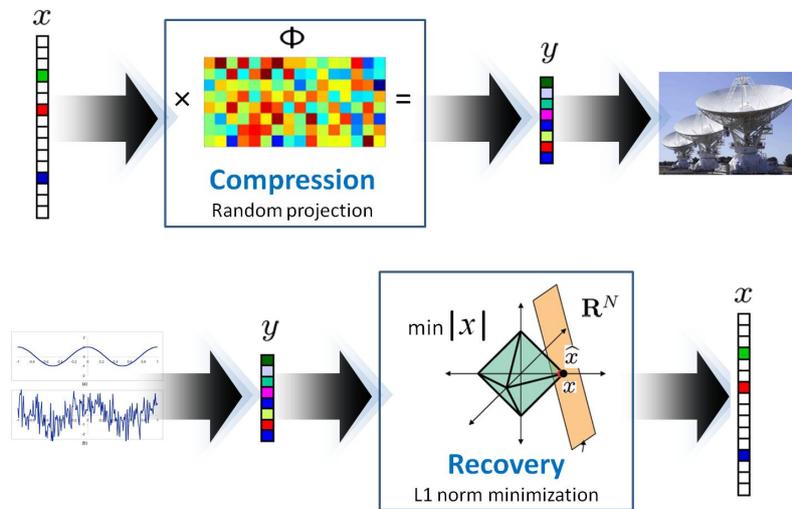


FIGURE 6.3: Schéma global de l'échantillonnage compressé.

Maintenant, entrons plus en détails sur les notions essentielles à la théorie de l'échantillonnage compressé avec tout d'abord, la propriété d'isométrie restreinte (RIP). Notons δ_S la constante d'isométrie d'ordre S correspondant à la plus petite valeur telle que la base Ψ utilisée pour la reconstruction d'un signal f S -parcimonieux vérifie :

$$(1 - \delta_S) \|f\|_{l_2}^2 \leq \|\Psi f\|_{l_2}^2 \leq (1 + \delta_S) \|f\|_{l_2}^2.$$

On dit alors, que Ψ vérifie la condition de RIP d'ordre S si δ_S est suffisamment différent de la valeur 1. Dans ce cas, chaque sous-ensembles de S colonnes de Ψ sont considérés comme quasiment orthogonaux, Ψ préserve ainsi quasiment la norme de tout signal S -parcimonieux. Pour l'échantillonnage compressé, si la constante δ_S est suffisamment inférieure à 1 cela assure une solution robuste au problème 6.3.

Le choix des bases Ψ et Φ doit être réalisé minutieusement selon les applications et doit entrer dans le cadre défini précédemment. De nombreux travaux explorent cette problématique. Il a été prouvé que les bases d'acquisition Φ suivantes dont les com-

posantes sont tirées aléatoirement peuvent être utilisées avec différentes bases Ψ fixées (ondelettes, DCT, ...) :

- Base construite à partir d'une matrice Gaussienne.
- Base de Fourier.
- Base orthonormale.

Nous pouvons à présent revenir sur le problème 6.3 : en pratique, la plupart des signaux observés ne sont pas rigoureusement parcimonieux dans chaque base transformée Ψ et de plus il est nécessaire de considérer l'addition d'un bruit dû à la précision des outils d'acquisition. L'addition d'une telle incertitude dans les mesures est modélisée de la façon suivante :

$$y = \Phi f + e,$$

où e est l'erreur stochastique ou déterministe de telle manière que son énergie soit bornée $\|e\|_{l_2} \leq \varepsilon$.

Une variante classiquement utilisée du problème 6.3 consiste alors à relâcher la contrainte d'attache aux données :

$$\hat{f} = \arg \min_f \|\Psi f\|_{l_1}, \text{ tel que } \|\Phi f - y\|_{l_2} \leq \varepsilon. \quad (6.5)$$

Finalement, ce problème restant complexe à résoudre comme cela a été mentionné précédemment, c'est le problème d'optimisation sans contrainte suivant qui est utilisé :

$$\hat{f} = \arg \min_f \left\{ \|\Psi f\|_{l_1} + \lambda \|y - \Phi f\|_{l_2} \right\}, \quad (6.6)$$

où λ est un coefficient qui contrôle le poids du terme d'attache aux mesures.

Au cours des dernières années, beaucoup d'approches différentes ont été proposées pour résoudre le problème de reconstruction 6.6. Chacune de ces méthodes de reconstruction jongle de façon classique avec un compromis entre qualité de la reconstruction et complexité de calcul. Citons deux algorithmes utilisés dans nos travaux :

ADM De l'anglais *alternating direction method*, cette méthode [Yang & Zhang 09] est considérée comme rapide et efficace. Le problème 6.6 est équivalent à :

$$\min_{x,z} \{g(x) + h(z), \text{ tel que : } Ax + Bz = b\}, \quad (6.7)$$

où nous avons défini :

$$x = \begin{pmatrix} \phi f \\ \psi f \end{pmatrix} \in \mathbb{R}^m \times \mathbb{R}^p, B = \begin{bmatrix} \phi \\ \psi \end{bmatrix}, A = -Id, b = \begin{pmatrix} y \\ 0 \end{pmatrix} \quad (6.8)$$

Le Lagrangien augmenté de ce problème est alors défini par :

$$\mathcal{L}_A(x, z, \lambda) = g(x) + h(z) - \lambda^\top (Ax + Bz - b) + \frac{\beta}{2} \|Ax + Bz - b\|^2, \quad (6.9)$$

où $\lambda \in \mathbb{R}^l$ est le multiplicateur de Lagrange et $\beta > 0$ est un paramètre de pénalité.

Plutôt que de résoudre le double problème de minimisation en (x, z) , l'algorithme ADM résout alternativement chaque minimisation, comme suit :

$$\begin{cases} x^{k+1} \leftarrow \operatorname{argmin}_x \mathcal{L}_A(x, z^k, \lambda^k) \\ z^{k+1} \leftarrow \operatorname{argmin}_z \mathcal{L}_A(x^{k+1}, z, \lambda^k) \\ \lambda^{k+1} \leftarrow \lambda^k - \gamma\beta(Ax^{k+1} + Bz^{k+1} - b) \end{cases} \quad (6.10)$$

PL En anglais *projected Landweber*, cette méthode introduite dans [Haupt & Nowak 06] fait partie de la classe des techniques basées sur des projections. Elle est caractérisée par une complexité de calcul réduite et la possibilité d'incorporer facilement de nouveaux critères d'optimisation. Cette méthode consiste en une succession de projection et de seuillage.

En partant d'une solution initiale \hat{f}^0 , l'approximation du signal reconstruit à l'itération $i + 1$ est obtenue par :

$$\hat{f}^i = \hat{f}^i + \frac{1}{\lambda} \Psi \Phi^T (y - \Phi \Psi^{-1} \hat{f}^i), \quad (6.11)$$

$$\hat{f}^{i+1} = \begin{cases} \hat{f}^i, & \left| \hat{f}^i \right| \geq \tau^i \\ 0 & \text{sinon.} \end{cases} \quad (6.12)$$

où λ est un paramètre fixé par exemple à la valeur propre de $\Phi^T \Phi$ la plus élevée, et τ^i est un seuil mis à jour à chaque itération.

Pour clore cette section, il est important de noter que le fort enthousiasme qui entoure l'échantillonnage compressé réside à la fois dans ses fondements multidisciplinaires qui regroupe plusieurs communautés et dans ses nombreuses applications. Celle qui nous intéresse et sera développée dans la section 2 est bien-sûr la compression de données mais on trouve aussi des applications dans le traitement du signal en général (détection, classification, estimation...), en imagerie médicale pour la résolution de problèmes inverses (imagerie par résonance magnétique), en acquisition de données où plusieurs systèmes d'acquisition dédiés ont été développés, ou encore pour la correction d'erreur ou en estimation statistique.

Dans la section suivante nous présentons l'échantillonnage compressé appliqué à un signal de type vidéo et son utilisation pour la compression de tels signaux.

6.2.2 Echantillonnage compressé et vidéo

L'application de l'échantillonnage compressé à un signal de type vidéo nécessite des modifications du procédé initial afin d'exploiter les caractéristiques des vidéos. Nous présentons dans cette section les principales contributions sur le sujet puis nous détaillons les approches intégrant un tel système dans un codeur vidéo moderne.

Précisons ici que notre contribution que nous introduirons dans la section 8.2.1 s'inscrit dans le cadre des travaux présentés dans la première sous-section (6.2.2.1) tandis que la seconde mise en œuvre que nous présenterons a été réalisée dans le codeur vidéo HEVC (8.2.2) et s'inscrit dans le cadre des travaux de la seconde sous-section ci-dessous (6.2.2.2).

6.2.2.1 Echantillonnage compressé appliqué à un signal vidéo

La première idée intuitive pour l'application de l'échantillonnage compressé à un signal de type vidéo, proposée dans [Wakin *et al.* 06], consiste à appliquer à chaque image d'une séquence vidéo les méthodes développées pour le traitement d'images fixes. Dans cet article, chaque image est reconstruite indépendamment en utilisant l'opérateur ondelette 2D. Cette solution demeure néanmoins limitée du fait de la non-exploitation de la dimension temporelle du signal. Dans la même contribution une seconde approche est proposée en considérant le signal comme étant la séquence vidéo complète et en lui appliquant un opérateur d'acquisition 3D. Bien que cette solution permette d'explorer la dimension temporelle, la séquence vidéo est reconstruite en une passe ce qui rend les calculs très coûteux. Pour résoudre ce problème, les auteurs de [Park & Wakin 09] proposent d'exploiter les redondances temporelles en réalisant un échantillonnage image par image combiné à des étapes d'estimation de mouvement et de compensation de mouvement.

Avant de présenter les contributions suivantes, insistons sur le fait que pour un signal de type vidéo, encore plus que pour les images fixes, le temps de traitement lié à la résolution du problème 6.3 appliqué au signal dans sa totalité est très élevé et engendre des difficultés de stockage en mémoire. Un tel schéma est donc irréaliste dans le cadre d'applications réelles. Une solution proposée par Gan [Gan 07] est par conséquent de diviser le problème en sous-problèmes appliqués bloc par bloc. La matrice d'échantillonnage Φ est ainsi définie par :

$$\Phi = \begin{bmatrix} \Phi_B & 0 & \dots & 0 \\ 0 & \Phi_B & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \Phi_B \end{bmatrix}$$

L'acquisition peut alors être réalisée bloc par bloc : $y_i = \Phi_B x_i$, avec x_i le $i^{\text{ème}}$ bloc de l'image courante. Ce procédé est utilisé dans les contributions détaillées ci-dessous.

Les auteurs de [Mun & Fowler 11] s'appuient sur un tel schéma combiné à un algorithme de reconstruction PL où une étape de lissage à l'aide d'un filtre de Wiener a été ajoutée (en anglais, *smooth projected Landweber*, noté SPL). Ils proposent de traiter chaque bloc de façon itérative en utilisant une compensation de mouvement pour affiner la reconstruction par échantillonnage compressé du bloc résiduel obtenu. Ledit résiduel reconstruit est ensuite ajouté au bloc obtenu par compensation de mouvement qui est

utilisé pour l'itération suivante. Ainsi, cette méthode propose une solution ingénieuse pour exploiter les redondances temporelles. Elle améliore très nettement les résultats par rapport à l'approche 3D et permet des gains significatifs par rapport à [Wakin *et al.* 06]. Un second article [Tramel & Fowler 11] propose d'explorer un autre outil propre aux codeurs vidéo modernes : des prédictions temporelles de l'image courante basées sur des hypothèses multiples. Chaque bloc de l'image courante est échantillonné de manière classique mais c'est le résiduel qui est reconstruit au moment du décodage avant d'être ajouté au bloc prédicteur. Cette prédiction est réalisée à l'aide d'une combinaison linéaire optimale des prédictions déterminée à l'aide d'un opérateur de régularisation de Tikhonov.

En addition de ces techniques cherchant la meilleure mise en œuvre de l'échantillonnage compressé appliqué à un signal de type vidéo, des travaux proposent d'ajouter la notion d'adaptativité. Dans [Stankovic *et al.* 08], les blocs catégorisés comme étant parcimonieux sont traités à l'aide de l'échantillonnage compressé tandis que les autres sont échantillonnés de façon classique. Les auteurs des travaux [Liu *et al.* 10, Liu *et al.* 11] vont plus loin et proposent de réaliser un échantillonnage adaptatif reposant sur une étude précise des redondances temporelles. Chaque bloc est ainsi catégorisé suivant son activité temporelle par rapport au bloc colocalisé de l'image précédente, puis la stratégie d'acquisition des mesures est adaptée selon la caractérisation de chaque bloc de même que la reconstruction de chaque bloc. Ainsi pour les blocs statiques (très corrélés au bloc colocalisé), un nombre restreint de mesures est effectué et le bloc est reconstruit selon deux méthodes, les blocs avec peu de changement utilisent un nombre un peu plus élevé de mesures et c'est la différence avec le bloc colocalisé qui est reconstruite avant addition avec ce dernier, enfin les blocs trop différents nécessitent un nombre élevé de mesures et sont reconstruits de façon classique, indépendamment du bloc colocalisé. Une amélioration significative de 1.5 à 7.8dB est rapportée en comparaison aux travaux précédents en vidéo. Ces travaux proposent donc une nouvelle voie prometteuse en introduisant la notion d'adaptativité. L'utilisation systématique du bloc colocalisé pour réaliser la caractérisation semble cependant limiter la méthode pour les nombreuses situations où ce bloc n'est pas corrélé au bloc courant.

Finalement, insistons sur le fait qu'au-delà de l'application initiale de l'échantillonnage compressé, utilisé au moment de l'acquisition d'un signal brut, et des avantages détaillés dans 6.2.1, les contributions présentées ci-dessus peuvent être regardées comme un nouveau moyen de définir un codeur vidéo. Le signal vidéo est ainsi représenté par un nombre de mesures limité puis il est reconstruit par un décodeur connaissant l'opérateur d'acquisition Φ utilisé. L'idée est attractive, elle peut devenir très intéressante lorsqu'elle est combinée à une exploitation efficace des propriétés du signal vidéo (redondances spatiales et temporelles) comme dans certaines des contributions précédentes. Bien sûr, de tels schémas ne sont pour l'instant pas compétitifs avec un codeur vidéo type H.264/AVC très efficace grâce à la mise en compétition d'un grand nombre de modes de codage qui exploitent les redondances spatiales et temporelles. Les possibilités offertes par l'échan-

tillonnage compressé et le dynamisme de la communauté scientifique sur le sujet laissent cependant entrevoir une alternative sérieuse aux codeurs vidéo modernes. Pour ce faire, la notion d'adaptativité à la fois de l'échantillonnage et de la reconstruction, combiné à la transmission des paramètres associés, semble être un point clé du succès d'un codeur basé échantillonnage compressé.

6.2.2.2 Echantillonnage compressé intégré dans un codeur vidéo

Plutôt que de redéfinir un codeur vidéo basé sur l'échantillonnage compressé, quelques travaux ont fait le choix d'intégrer directement l'échantillonnage compressé dans un codeur vidéo existant. Nous introduisons ces travaux dans cette section.

Une première application est proposée dans [Zhang *et al.* 08]. La réduction du nombre de mesures nécessaires pour reconstruire correctement le signal est exploitée afin d'améliorer l'efficacité de codage. Ainsi, le résiduel r transformé et quantifié après prédiction est échantillonné par un opérateur Φ . A partir des mesures y_r obtenues puis déquantifiées (\tilde{y}_r), le résiduel est reconstruit dans le domaine Ψ par résolution de l'équation suivante :

$$\hat{r} = \underset{r}{\operatorname{arg\,min}} \{ \|\Psi r\|_1 + \lambda \|\tilde{y}_r - \Phi r\|_2 \} \quad (6.13)$$

Le résiduel obtenu est alors ajouté au signal prédictif pour reconstruire le bloc. Une transformée DCT 1D est proposée par les auteurs à la place de la DCT 2D classique afin de s'aligner avec l'équation initiale de l'échantillonnage compressé. La méthode est évaluée en codage Intra sous des conditions simplifiées avec une taille de partition fixée à 8×8 pixels et un seul mode de prédiction (DC) activé. La méthode est ainsi en compétition avec le mode de codage classique (DCT 2D) et signalée au décodeur par un élément de syntaxe si le bloc contient au moins un coefficient AC non nul. Les gains présentés sont encourageants et démontrent l'intérêt d'une telle méthode, particulièrement sur des contenus ayant de grandes régions homogènes bien délimitées comme dans la figure 6.4.



FIGURE 6.4: Exemple de contenu adapté pour la méthode [Zhang *et al.* 08].

Dans [Do *et al.* 10], la méthode précédente est améliorée : afin de mieux s'adapter

au cadre de la compression vidéo, le signal est directement reconstruit dans le domaine pixel et non plus dans le domaine résiduel. L'information du signal prédictif y_p est ajoutée dans le problème d'optimisation 6.13 et le premier terme est remplacé par la minimisation de la variation totale :

$$\hat{f} = \arg \min_f \left\{ TV(f) + \lambda \left\| (\tilde{y}_r + y_p) - \Phi f \right\|_2 \right\}, \quad (6.14)$$

où la variation totale d'un signal f à deux dimensions est définie par :

$$TV(f) = \sum_{i,j} \sqrt{(f_{i+1,j} - f_{i,j})^2 + (f_{i,j+1} - f_{i,j})^2}.$$

L'opérateur d'acquisition utilisé conserve les 40 premiers coefficients d'un bloc transformé par une DCT parcouru en zig-zag. Les auteurs montrent que dans le cadre de leur étude cet opérateur est plus efficace qu'un opérateur aléatoire classiquement utilisé en échantillonnage compressé. Les gains obtenus sont intéressants, notamment sur des séquences spécialement sélectionnées. La méthode semble cependant ne pas tirer profit au maximum des possibilités offertes par l'échantillonnage compressé. Ainsi, le nombre de coefficients pour l'échantillonnage est fixé à 40 mais ce chiffre n'est évidemment pas optimal pour toutes les séquences et toutes les partitions. Comme expliqué dans la conclusion de l'article, le choix de la base DCT pour l'acquisition est adapté pour les blocs de texture mais une autre base serait mieux adaptée pour d'autres types de contenus.

En addition de ces travaux visant à intégrer un schéma d'échantillonnage compressé au sein d'un codeur vidéo, nous pouvons citer d'autres travaux explorant l'ajout d'approches itératives dans des codeurs modernes et qui exploitent notamment la notion de parcimonie du signal. En particulier, les travaux [Martin *et al.* 08, Thoreau *et al.* 10, Martin *et al.* 07] qui visent à améliorer l'efficacité de la prédiction par la recherche du meilleur dictionnaire. Ce type d'approche est mis en compétition avec une approche basée *Template Matching* dans [Turkan & Guillemot 10] pour le codage d'image Intra.

6.3 Bilan

Dans ce chapitre, nous avons tout d'abord présenté l'intérêt du transfert d'intelligence au niveau du décodeur dans le cadre du codage vidéo distribué mais surtout dans un cadre de codage classique. Ces approches sont très prometteuses en terme d'amélioration d'efficacité de codage bien qu'elles entraînent une augmentation de la complexité du décodeur. Une seconde section a ensuite permis d'introduire la théorie émergente de l'échantillonnage compressé qui présente un fort intérêt dans le cadre de la compression vidéo.

Nous nous appuyons sur cet état de l'art dans les chapitres suivants pour introduire

dans un premier temps un schéma de codage construit sur l'utilisation d'un décodeur actif. Nous appliquons ensuite ce schéma à l'échantillonnage compressé afin de confirmer le potentiel d'une telle méthode de codage et proposer une solution permettant d'en tirer profit au maximum dans un cadre de codage vidéo.

Chapitre 7

Approche décodeur intelligent avec application au codage Intra

Dans ce chapitre, nous proposons une solution pour anticiper les limites des schémas de codage vidéo conventionnels basés sur la compétition de multiples choix de codage qui exploitent efficacement les différentes corrélations du signal vidéo. Comme cela a été mis en lumière dans le chapitre 2, le coût de signalisation résultant de cette compétition est d'une part élevé et rend d'autre part délicat l'ajout de nouveaux modes de codage. Afin de contourner ces contraintes, l'approche en rupture que nous introduisons ici consiste à reporter cette compétition au niveau du décodeur afin d'économiser le coût de signalisation tout en se donnant un nouveau degré de liberté concernant l'ajout de nouveaux modes. On parle ainsi de décodeur intelligent.

Cette approche va plus loin que les méthodes évoquées dans l'état de l'art précédent dans la mesure où elle rend potentiellement reproductible l'ensemble des opérations réalisées au codeur dans un but d'amélioration de l'efficacité de codage. Cette méthode alternative passe cependant par une augmentation de la complexité du décodeur et donne lieu à différentes problématiques de mises en œuvre pour lesquelles nous proposons des solutions au cours de ce chapitre. Afin de concrétiser et mesurer son potentiel, une application de la méthode à la sélection des modes de prédiction Intra est proposée et évaluée dans le cadre du futur standard HEVC. Cette étude étant encore au stade préliminaire, nous concluerons finalement le chapitre avec différentes perspectives prometteuses.

7.1 Description de l'approche proposée

7.1.1 Motivation

Les chapitres 1 et 2 de ce manuscrit ainsi que la première partie du chapitre précédent nous ont permis d'évoquer une limite des schémas de codage conventionnels liées à la mise en compétition de multiples choix de codage. La motivation de ce chapitre est de

proposer une alternative à cette approche.

Les schémas de codage actuels sont basés sur l'exploitation des redondances spatiales et temporelles du signal à l'aide de multiples choix de codage en compétition. L'amélioration incrémentale de tels schémas implique alors, entre autres, une augmentation significative du nombre de paramètres en compétition à chaque génération de standards. Les choix nécessitent en effet d'être de plus en plus abondants afin d'améliorer l'efficacité de codage de la texture. C'est ainsi que le projet HEVC introduit notamment une compétition entre cinq vecteurs de mouvement prédictifs parmi de nombreux candidats alors que seul le médian spatial utilisant trois vecteurs voisins était disponible dans H.264/AVC et que MPEG-2 n'utilisait qu'un unique vecteur provenant du bloc de gauche. L'augmentation du nombre de vecteurs prédictifs a ici pour effet de réduire le coût de codage des vecteurs de mouvement au prix d'un surcoût de signalisation de la compétition qui demeure inférieur au gain engendré, l'effort consenti est donc récompensé. La même démarche est utilisée en Intra où l'amélioration de la prédiction est rendue possible par un nombre de prédictifs quasiment multiplié par quatre. Cette augmentation significative des choix de codage en compétition contraint cependant l'évolution futur de tels encodeurs dans la mesure où l'amélioration de la prédiction du signal nécessite d'être systématiquement supérieur à un coût de signalisation grandissant. Nous formalisons ci-dessous cette évolution des encodeurs vidéos afin d'analyser ce constat de façon théorique :

Soit (R_1, D_1) et (R_2, D_2) les couples débit-distorsion générés pour des paramètres de codage C_1 et C_2 appartenant respectivement à un ensemble de N_1 et N_2 paramètres en compétition qui représentent deux générations d'encodeurs vidéos successives telles que $N_1 < N_2$. Nous nous basons sur la représentation Lagrangienne du problème à partir du critère $J_k = D_k + \lambda R_k$, $k \in \{1, 2\}$, où λ est un multiplicateur de Lagrange. Pour que la seconde génération d'encodeur soit plus efficace que la première ($J_1 > J_2$), il est nécessaire que :

$$D_1 - D_2 > \lambda(R_2 - R_1), \quad (7.1)$$

soit, en séparant chaque composante de débit R_i , $i \in \{1, 2\}$ en une composante liée à la transmission des choix de codage R_{C_i} et une seconde composante liée aux résiduels de textures R_{res_i} :

$$D_1 - D_2 > \lambda((R_{C_2} - R_{C_1}) + (R_{res_2} - R_{res_1})). \quad (7.2)$$

Nous considérons maintenant que dans le futur le coût des résiduels sera borné du fait des nets progrès en prédiction de la texture, $R_{res_2} \rightarrow R_{res_1}$. Bien que cela n'est pas vérifié rigoureusement avec le passage d'H.264/AVC à HEVC, nous avons montré dans l'analyse du chapitre 2 que les écarts avaient significativement diminué. Sous cette hypothèse, on peut alors écrire :

$$R_{res_2} = R_{res_1} - \varepsilon. \quad (7.3)$$

En remplaçant le terme R_{res_2} dans (7.2), il vient ainsi :

$$D_1 - D_2 > \lambda((R_{C_2} - R_{C_1}) - \varepsilon). \quad (7.4)$$

En supposant finalement que les performances des codeurs entropiques demeurent stables, ce qui implique que les paramètres de codage C_2 soient plus coûteux que les paramètres C_1 , nous introduisons Δ le surcoût de signalisation nécessaire pour permettre davantage de compétition tel que $\Delta > 0$, $R_{C_2} = R_{C_1} + \Delta$. Par conséquent, nous obtenons :

$$D_1 - D_2 > \lambda\Delta - \lambda\varepsilon. \quad (7.5)$$

Ainsi, ce résultat indique que la différence de qualité de reconstruction entre deux générations d'encodeurs est bornée par le surcoût de signalisation résultant de l'augmentation des choix de codage en compétition. Lorsque ce nombre de choix tend vers une valeur très élevée, $\Delta \gg \varepsilon$, l'expression (7.5) devient en effet :

$$D_1 - D_2 > \lambda\Delta, \quad (7.6)$$

où le terme $\lambda\Delta$ est de plus en plus élevé avec les futures générations d'encodeurs du fait du surcoût de codage lié à l'ajout de choix. L'amélioration en qualité de reconstruction étant bornée par ce terme, elle nécessite donc d'être encore meilleure pour assurer une progression dans l'efficacité de codage. Cela n'est néanmoins pas garanti du fait des limites dans la qualité de la prédiction. C'est cette contradiction que nous souhaitons contourner en introduisant dans ce chapitre une alternative aux schémas d'encodage actuels. Ainsi, notre étude ne se dirige pas vers un nouveau mode de prédiction ou l'amélioration du codage des données mais plutôt un nouvel environnement de codage visant à répartir différemment les ressources entre l'encodeur et le décodeur. Ce dernier sera ainsi davantage exploité pour devenir acteur du schéma.

Avant de détailler l'approche proposée, citons de nouveau la solution de l'adaptativité évoquée dans le chapitre précédent et qui permet déjà de solliciter d'une manière plus active le décodeur. Cette méthode consiste à utiliser uniquement un ensemble réduit de paramètres en compétition générés de manière pertinente pour chaque partition à coder en fonction des caractéristiques du voisinage causal. Elle demeure néanmoins une solution délicate à envisager du fait des difficultés liées au décodage du flux qui limitent voire empêchent une mise en œuvre pratique. L'approche que nous proposons dans ce chapitre est différente, nous en introduisons le schéma général dans la sous-section suivante avant d'en détailler les différentes étapes par la suite.

7.1.2 Schéma global

Pour résumer le schéma proposé, nous souhaitons exploiter le décodeur pour déterminer certains paramètres de codage dont la transmission peut être alors économisée au

niveau de l'encodeur dans la mesure où il est capable de dériver les paramètres à l'identique. Ainsi, l'approche consiste plus précisément à hériter un ensemble de paramètres de codage à partir d'une partition causale judicieusement sélectionnée. De manière à conserver une efficacité de codage significative, les choix de codage ainsi dérivés doivent être suffisamment pertinent. A cette efficacité de codage conservée s'ajoute alors les bénéfices de la méthode dont nous discuterons après avoir décrit, ci-dessous, la méthode et introduit les notations qui seront utilisées dans la suite de ce chapitre.

Nous notons P_i la partition courante à coder et L_i la liste regroupant les paramètres de codage à déterminer pour la partition P_i . Nous définissons alors P'_i la partition causale que nous allons exploiter pour dériver les paramètres L_i . Cette partition qui doit être disponible lors du traitement de P_i est indiquée par une information de mouvement notée vm_i dont nous discuterons dans la section suivante. Notre approche est divisée en trois étapes successives :

1. Sélection de la partition causale P'_i .
2. Mise en compétition de N^* paramètres sur la partition causale P'_i .
3. Codage de la partition P_i à partir des paramètres hérités de l'étape précédente.

Ces étapes sont représentées sur la figure 7.1 en regard du schéma de l'état de l'art.

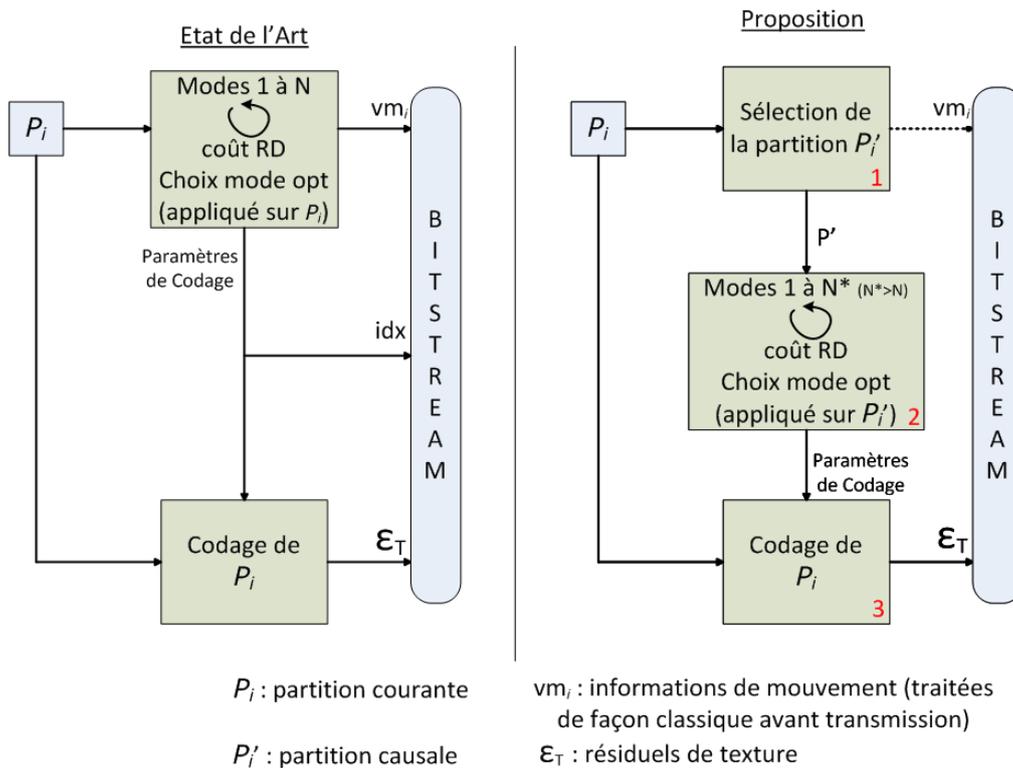


FIGURE 7.1: Schéma de codage d'une partition courante de façon classique (à gauche) ou selon notre proposition (à droite).

Dans ce dernier, on traite la partition originale P_i pour laquelle les paramètres de codage optimaux sont déterminés lors d'une compétition entre N modes de codage, une information de mouvement est alors transmise (dans un cas Inter) accompagnée de l'ensemble des paramètres de codage optimaux qui sont aussi utilisés pour le codage de la partition courante donnant lieu par ailleurs à des résiduels de textures ε_T .

Ainsi, les redondances spatiales et temporelles entre pixels ne sont plus ici exploitées pour prédire d'autres pixels mais pour prédire des moyens de codage qui peuvent alors eux-mêmes exploiter les redondances pixelliques en tant que telles. La figure 7.2 schématise l'approche proposée selon ce point de vue. Les paramètres de codage sont déterminés grâce à l'utilisation d'une partition causale suffisamment corrélée à la partition courante, P'_i jouant donc le rôle de la partition de référence en se substituant à la partition originale. La liste des paramètres L_i ainsi déterminée, ces derniers sont hérités pour coder la partition courante et ils ne nécessitent alors pas d'être transmis au décodeur dans la mesure où celui-ci est capable de les déterminer à son tour. La méthode est par conséquent décodable dès lors que codeur et décodeur utilisent la même partition causale P'_i et que le processus de sélection des paramètres optimaux est identique. Plus précisément, afin de répéter ce processus de décision, le décodeur effectue les opérations de simulation de codage des données ainsi que de reconstruction de la partition courante pour l'ensemble des paramètres en compétition. En ce sens, nous parlons de *décodeur intelligent*.

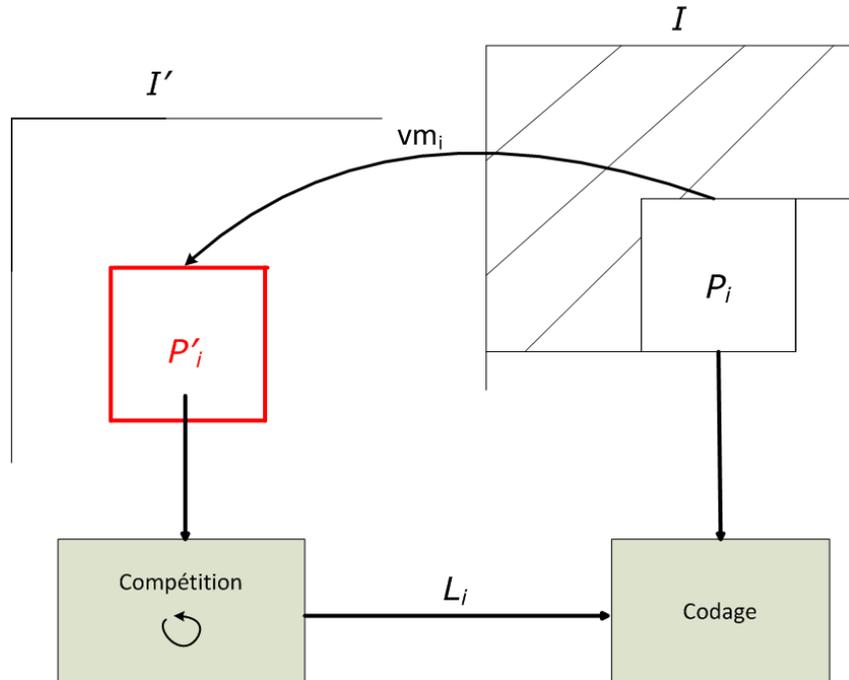


FIGURE 7.2: Représentation de l'approche *décodeur intelligent*.

Après cette description, dont les trois points principaux seront détaillés dans les sous-sections suivantes, discutons de l'apport de cette approche par rapport au problème des schémas de codage conventionnels énoncé dans 7.1.1. En considérant que les choix de codage dérivés sont suffisamment pertinents, deux voies d'améliorations sont alors ouvertes par la méthode :

1. Réduction des coûts de signalisation grâce à l'économie des informations de compétition.

Il s'agit de l'apport immédiat de l'approche qui en permettant de dériver à l'identique les informations de codage au niveau du décodeur réduit leur proportion dans le flux dans la mesure où elles ne sont plus transmises. La proportion du flux ainsi réduite peut alors être en partie reportée sur les résiduels de textures pour améliorer la qualité de reconstruction du signal.

2. Affranchissement de la contrainte liée au nombre de paramètres en compétition.

Dans les codeurs conventionnels, le nombre N de choix disponibles pour un paramètre donné est limité. Il est classiquement défini de façon à être suffisamment pertinent du point de vue du compromis débit-distorsion. Dans le cadre de notre approche, cette quantité de choix en compétition N^* n'est plus limitée et peut donc être augmentée sans se soucier du coût de signalisation : $N^* > N$. N^* est alors uniquement contraint par la complexité visée qui augmente en conséquence et fait apparaître une nouvelle façon d'aborder la question de la scalabilité en complexité dont nous discuterons à la fin de ce chapitre.

Pour aller plus loin dans cette analyse, développons l'impact de la méthode proposée à partir du critère débit-distorsion qui est modifié. Ce critère à minimiser est classiquement noté J_i , il est utilisé pour la sélection des choix de codage de la partition P_i et formulé comme suit :

$$J_i = D_i + \lambda R_i, \quad (7.7)$$

où :

- D_i est la distorsion entre la partition originale et la partition reconstruite pour un ensemble de choix de codage donné C_i .
- R_i est le débit généré par le codage de l'ensemble des informations relatives à un ensemble de choix de codage donné C_i .
- λ est le multiplicateur de Lagrange permettant de balancer entre débit et distorsion en tenant compte du paramètre de quantification courant.

Dans ce cadre, la liste L_i des paramètres traités à l'aide de notre approche constitue un sous-ensemble de C_i ($L_i \subset C_i$) et on note alors \tilde{R}_i le débit résiduel généré comme suit :

$$\tilde{R}_i = R_i - \varsigma_L(L_i), \quad (7.8)$$

où $\varsigma_L(L_i)$ représente le coût de codage associé à la transmission des paramètres de la liste L_i dans un schéma classique sans notre méthode.

Le critère débit-distorsion (7.7) s'écrit alors sous la forme suivante :

$$J_i = D_i + \lambda(\tilde{R}_i + \varsigma_L(L_i)). \quad (7.9)$$

Notre proposition dans le cadre de l'approche décodeur intelligent consiste à substituer au coût de codage des paramètres L_i , le coût de codage $\varsigma_{vm}(vm_i)$ de l'information de mouvement vm_i permettant de définir la partition causale P'_i , soit :

$$J'_i = D'_i + \lambda(\tilde{R}'_i + \varsigma_{vm}(vm_i)), \quad (7.10)$$

où $\tilde{R}_i \neq \tilde{R}'_i$ et $D_i \neq D'_i$ dans la mesure où notre méthode impacte sur la reconstruction de P_i .

Ainsi, la méthode devient avantageuse lorsque :

1. l'information de mouvement est moins coûteuse que les paramètres de codage traités, $\varsigma_{vm}(vm_i) < \varsigma_L(L_i)$.
2. les couples (D_i, R_i) et (D'_i, \tilde{R}'_i) sont suffisamment proches, c'est à dire que la prédiction est restée suffisamment efficace.

Il est important de noter que nous intervenons sur ces deux points de façon étroitement liée. Le coût que nous allouons à la signalisation de la partition causale influence en effet la pertinence des paramètres hérités à l'aide de notre méthode, qui agissent eux-mêmes sur la qualité de la reconstruction dont dépend finalement le débit résiduel \tilde{R}'_i au travers des résiduels de texture à coder. L'étape de sélection de la partition causale et sa signalisation sont par conséquent les étapes clés de notre étude.

Dans la sous-section suivante, nous débutons justement la description des étapes par la sélection de P'_i .

7.1.3 Sélection de la partition causale

Comme expliqué dans la première section de ce chapitre, la sélection de la partition causale est une étape clé de notre méthode dans la mesure où une partition suffisamment corrélée à la partition courante permettra de déterminer les meilleurs paramètres. A l'inverse, une « mauvaise » partition P'_i engendrera des paramètres sous-optimaux pouvant donner lieu à une mauvaise prédiction et finalement un surcoût de codage des résiduels de texture résultant. La sélection de la partition P'_i est réalisée parmi l'ensemble des partitions causales précédemment traitées au cours du codage. Notons donc que dans le cadre d'un codage Intra de l'image I , la partition P'_i appartient nécessairement à la zone causale de l'image courante tandis que pour une image Inter, elle peut en plus appartenir à l'image de référence ou mieux dans un cadre bi-prédicatif à l'une des images de référence.

Nous introduisons ci-dessous quatre solutions envisagées réparties en deux types de méthodes selon qu'un vecteur de mouvement soit explicitement transmis ou pas.

Solutions avec signalisation du vecteur de mouvement

Pour déterminer la partition P'_i nous proposons d'utiliser dans un premier temps un algorithme classique de mise en correspondance de blocs (en anglais Block-Matching, noté BMA). Le BMA explore les corrélations entre pixels dans une zone de recherche définie au sein de la zone causale. Le critère de sélection est un critère débit-distorsion tenant compte du coût de codage de l'information de mouvement associée vm_i , la distorsion étant pour sa part estimée par la somme des différences absolues (SAD) entre la partition courante P_i et chaque partition causale candidate $P'_{i,k}$. Cette première solution a l'avantage d'être celle classiquement utilisée dans les codeurs vidéo utilisant une prédiction par compensation de mouvement. De nombreux algorithmes très performants sont disponibles et l'intégration dans un codeur vidéo conventionnel est ainsi facilitée.

En alternative du BMA classique, nous proposons d'utiliser un BMA modifié réalisé dans un domaine approprié aux paramètres à déterminer. Notons J_{vm} le critère débit-distorsion utilisé dans l'estimation de mouvement du logiciel de référence de la norme H.264/AVC où le débit R est égal au débit relatif au codage du mouvement R_{vm} [Girod 94] :

$$J_{vm} = D + \lambda_{vm} R_{vm}. \quad (7.11)$$

Le critère modifié, qui demeure convexe, devient alors :

$$J'_{vm} = \hat{D} + \lambda_{vm} R_{vm}, \quad (7.12)$$

où \hat{D} représente un critère de ressemblance entre deux partitions dans un domaine prédéfini, différent du domaine pixel du macrobloc.

De telles estimations de mouvement seront introduites dans la section perspective ainsi que dans le chapitre suivant où l'approche est appliquée dans le cadre de l'échantillonnage compressé et \hat{D} représente alors une mesure de la distorsion dans le domaine transformée.

Concernant la signalisation de P'_i , les informations de mouvement vm_i peuvent être transmises au décodeur dans le cadre d'une mise en œuvre intuitive. Les procédés classiques de réduction du coût de codage des vecteurs de mouvement sont alors directement appliqués et permettent de signaler P'_i à moindre coût. Notons que dans le cadre d'un codage Intra-image, le domaine de recherche de la partition causale comprend uniquement la zone au-dessus et à gauche de la partition courante et que cela devra être pris en compte pour le codage du vecteur de mouvement.

Solutions sans signalisation du vecteur de mouvement

La signalisation par transmission du vecteur étant cependant non-négligeable, nous envisageons des solutions moins coûteuses (voir gratuites) dans le but d'introduire une telle méthode dans un codeur vidéo moderne. La partition causale est alors simplement

définie comme l'une des partitions voisines spatialement ou temporellement sans avoir recours à une estimation de mouvement. Les corrélations avec ces partitions sont en effet souvent suffisamment élevées pour assurer le succès de la méthode sans augmenter le coût de codage. Nous envisageons ainsi un large ensemble de partitions $P'_{i,j}$ introduites ci-dessous qui constituent la liste des partitions causales candidates $L_{P'_i}$:

- partitions voisines spatialement, notées respectivement A , B , C et D , comme représentées sur la figure 7.3 :

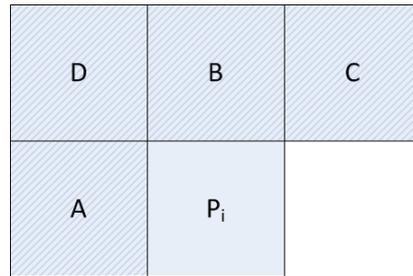


FIGURE 7.3: Partitions causales voisines spatialement (en hachuré) de la partition courante.

- partition dans l'image précédente pointée par le vecteur de mouvement prédicteur de la partition courante (par exemple, le vecteur médian des vecteurs de mouvement des trois voisins spatiaux comme dans H.264/AVC), comme représentée sur la figure 7.4 :

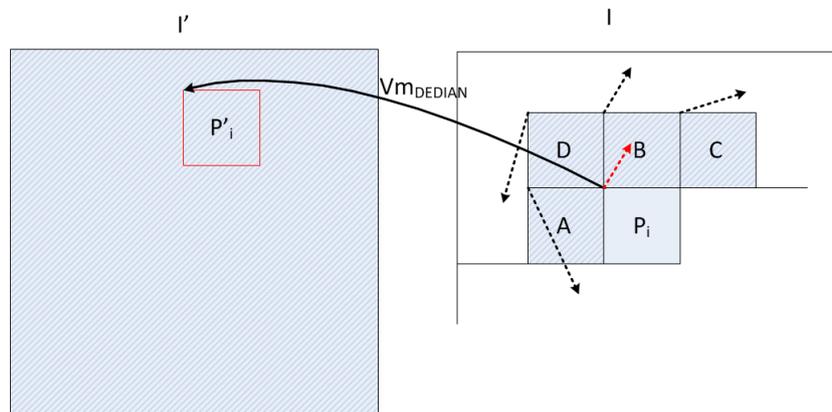


FIGURE 7.4: Partition causale pointée par le vecteur médian.

La partition causale P'_i à utiliser est alors, soit fixée par défaut à l'une de ces partitions soit sélectionnée parmi l'ensemble de ces partitions à l'aide d'un critère de similarité avec la partition courante P_i . Dans ce cas, un indice de signalisation est associé à chaque partition possible et on transmet cet indice au décodeur.

Enfin, pour économiser le coût de cet indice, une dernière solution consiste à utiliser les paramètres optimaux correspondants au meilleur compromis débit-distorsion indépendamment de la partition causale P'_i . Ainsi, nous considérons que chacune des partitions causales est une partition suffisamment corrélée à la partition courante pour la détermination de la liste des paramètres optimaux L_i^* . Le problème (7.10) est alors modifié comme nous le verrons dans la sous-section suivante pour réaliser l'optimisation à la fois sur la liste L_i et les partitions causales $P'_{i,j}$.

Dans la sous-section suivante, nous décrivons comment la sélection des paramètres optimaux est réalisée dans ce dernier cas de figure après avoir introduit le cas plus général.

7.1.4 Sélection des paramètres optimaux

L'étape de sélection des paramètres optimaux est calquée sur ce qui se fait actuellement dans un encodeur conventionnel, à ceci près que la décision doit nécessairement être reproductible au niveau du décodeur. Dans la mesure où la partition utilisée en lieu et place de la partition originale est effectivement disponible lorsque cela est nécessaire, le point restant à contrôler concerne le critère de décision. Ainsi, contrairement aux critères de sélection utilisés dans les codeurs actuels tels que le critère débit-distorsion, le critère utilisé dans notre solution devient normalisable puisqu'il est utilisé lors du décodage. En ce sens, il ne constitue plus un choix d'implémentation laissé libre aux concepteurs d'encodeur qui adaptent le processus de décision en fonction du type d'application et des contraintes de complexité. Un critère fixé reproductible de manière exacte assure dans notre approche que le décodeur hérite les mêmes paramètres que l'encodeur et décode parfaitement la séquence.

En pratique, dans le but d'utiliser un critère débit-distorsion optimal, cela ne pose pas de problème pour le calcul de la composante de distorsion D mais complexifie notre approche par rapport à la composante de débit R qu'il est alors nécessaire d'estimer au cours du décodage en simulant le codage des données à l'aide du codeur entropique. Cela devient notamment complexe dans le cas du codeur entropique CABAC qui nécessite l'utilisation de contextes dans un état identique à l'encodage et au décodage. On peut cependant envisager d'utiliser un coût approché à l'aide d'un codeur entropique CAVLC ou bien d'un codeur CABAC pour lequel les contextes seraient remis à jour uniquement au niveau de chaque image par exemple.

Afin de réaliser une première évaluation de notre approche, une solution intuitive mais néanmoins pertinente sera mise en œuvre à la fin de ce chapitre. Une mesure de l'entropie H des données à coder sera ainsi exploitée pour obtenir une approximation du débit [Cover & Thomas 91]. En limitant ces données aux résiduels de pixels transformés et quantifiés ε_{P_i} , le débit sera alors calculé comme suit :

$$R_i = nb_{\varepsilon_{P_i}} H(\varepsilon_{P_i}), \quad (7.13)$$

où $nb_{\varepsilon_{P_i}}$ représente le nombre de résiduels de pixels.

Détaillons maintenant la décision à partir de ce critère de sélection calculé à l'aide du couple débit-distorsion (R_i, D_i) . Deux solutions sont considérées selon que la méthode est mise en œuvre avec une unique partition causale P'_i ou bien plusieurs partitions causales $P'_{i,j}$, $j \in \{1, \dots, N_{P'}\}$ comme nous l'avons introduit dans la précédente sous-section.

Dans le premier cas, la partition causale P'_i est sélectionnée indépendamment des paramètres et il s'agit alors d'utiliser la partition P'_i à la place de la partition originale P_i lors de la génération du critère de sélection des paramètres de codage. En reprenant les notations de la première section 7.1.2 avec L_i la liste des paramètres traités à l'aide de notre approche et $L_i^c = C_i - L_i$ les paramètres codés classiquement, nous utilisons donc le critère suivant :

$$J'_i = D'_i + \lambda R'_i. \quad (7.14)$$

Soit, si la détermination de P'_i nécessite la transmission d'information de mouvement vm_i :

$$J'_i = D'_i + \lambda(\tilde{R}'_i + \varsigma_{vm}(vm_i)), \quad (7.15)$$

ou bien, dans le cas où cette signalisation est économisée en définissant, par exemple, une partition P'_i par défaut :

$$J'_i = D'_i + \lambda \tilde{R}'_i. \quad (7.16)$$

Dans le second cas impliquant de multiples partitions candidates, la problématique de détermination des paramètres optimaux est conjuguée à celle de la détermination de la partition causale optimale dès lors que l'on s'interdit la signalisation. Le critère $J'_{i,j}$ associé à chaque partition candidate $P'_{i,j} \in L_{P'_i}$ devient ainsi :

$$J'_{i,j} = D'_{i,j} + \lambda \tilde{R}'_{i,j}, \quad (7.17)$$

où $D'_{i,j}$ représente la distorsion engendrée par une prédiction donnée associée à un ensemble de choix de codage candidat et la partition $P'_{i,j}$, le débit $\tilde{R}'_{i,j}$ étant calculé relativement aux résiduels résultants par rapport à la partition $P'_{i,j}$.

La particularité vient donc ici de la comparaison de critères calculés avec des références différentes. Ce choix permet de combiner les deux étapes de sélection en faisant l'hypothèse qu'en héritant les paramètres minimisant le critère (7.17), il s'agisse des plus pertinents parmi l'ensemble des paramètres optimaux $L^*_{i,j}$ associés à chaque $P'_{i,j}$. Ainsi, nous évaluerons cette solution face à une plus intuitive consistant à retenir les paramètres $L^*_{i,j}$ les plus sélectionnés pour chaque partition $P'_{i,j}$.

Dans la sous-section suivante, nous clôturons la description théorique de notre approche à travers la dernière étape d'application des choix de codage ainsi dérivés.

7.1.5 Traitement de la partition courante

Une fois les paramètres optimaux déterminés à l'aide d'une ou plusieurs partitions causales, le codage de la partition courante est divisé en deux parties. D'une part, le codage classique de la partition et d'autre part la signalisation de la partition causale. Pour le codage de la partition, la liste de paramètres hérités de la précédente étape est utilisée de façon aveugle par le codeur qui prédit alors P_i à l'aide des choix optimaux déterminés sur P'_i avant de transmettre uniquement les résiduels de pixels résultants. La signalisation de P'_i qui a déjà été abordée précédemment, consiste soit en la transmission d'informations de mouvement (vecteur résiduel, image de référence, prédicteur) soit la transmission d'un indice indiquant la partition optimale parmi un ensemble prédéfini. Cette signalisation peut enfin être économisée dans le cas d'un critère exactement reproductible au décodeur comme nous l'expérimenterons dans la section suivante.

Ainsi, comme nous l'avons déjà mentionné, les redondances spatiales et temporelles entre pixels ne sont plus ici exploitées pour prédire d'autres pixels mais pour prédire des moyens de codage qui peuvent alors eux-mêmes exploiter les redondances pixelliques en tant que telles. Notons d'ailleurs que pour exploiter l'approche différemment et étendre la dualité avec la prédiction des pixels, il est possible d'envisager la transmission de résiduels de paramètres de codage à la manière du *most probable mode* utilisé dans H.264/AVC. Nous développerons ce point dans la section perspective 7.3.

Attardons-nous finalement sur le décodeur qui devient acteur de notre système. Celui-ci procède comme suit lorsqu'une partition courante à décoder P_i utilise notre approche :

1. Détermination de la partition causale P'_i après lecture éventuelle d'informations de mouvements.
2. Détermination des paramètres de décodage par simulation du codage de la partition P'_i .
3. Décodage de la partition P_i à l'aide des paramètres hérités de l'étape précédente.

Ainsi, l'appellation *décodeur intelligent* prend tout son sens ici dans la mesure où le schéma de décodage est symétrique de l'encodage. En particulier, dans le cas où la partition P'_i n'est pas signalée explicitement, le décodeur est alors exactement identique à l'encodeur. C'est notamment le cas lors de l'utilisation d'une partition causale prédéfinie ou de multiples partitions en compétition.

Dans le cadre de notre étude, que ce soit pour l'application aux modes Intra présentée dans la section suivante ou bien à l'échantillonnage compressé au chapitre suivant, nous n'avons pas considéré la problématique de la complexité engendrée au décodeur. Indiquons cependant qu'une telle approche bouleverse bien évidemment la conception et l'utilisation de celui-ci du fait des opérations qui lui sont demandées. Un couple encodeur et décodeur configurable en fonction de l'application est néanmoins envisageable du fait des progrès considérables dans les performances matérielles et logicielles des équipements multimédia. L'approche permet notamment d'envisager le sujet de la scalabilité

en complexité de façon conjointe à l'encodeur et au décodeur comme nous l'évoquerons dans les perspectives.

Dans la section suivante, nous mettons en pratique l'approche proposée afin d'évaluer son potentiel.

7.2 Application pour l'héritage de modes Intra

7.2.1 Description de la méthode

Afin de mettre en application la théorie développée dans ce chapitre, nous avons choisi d'appliquer notre solution au codage des prédicteurs Intra, l'ensemble des autres informations étant traitées classiquement. Les raisons de ce choix tiennent en trois observations soulignées notamment par l'analyse du chapitre 2 :

- Les prédicteurs Intra représentent 10% du flux total en proportion dans le standard H.264/AVC et 12% dans HEVC. L'impact potentiel de notre méthode est donc significatif.
- Afin d'améliorer l'efficacité de codage de la texture, le nombre de prédicteurs Intra en compétition augmente de façon significative dans les dernières générations d'encodeurs standards : 1 seul prédicteur dans MPEG 2 ; de 4 à 9 prédicteurs dans H.264/AVC et jusqu'à 34 prédicteurs dans le standard HEVC. Cette information de codage est donc particulièrement pertinente relativement à la motivation exprimée au début de ce chapitre.
- La sélection des prédicteurs optimaux repose sur la direction de la texture et nous sommes confiants pour que l'utilisation d'un bloc causal suffisamment corrélé au bloc courant permette de déterminer le prédicteur optimal du bloc courant de façon efficace.

L'application de l'approche décodeur intelligent à la sélection des prédicteurs Intra apparaît donc comme un choix pertinent afin de mesurer le potentiel de la méthode. La liste L_i des paramètres à déterminer est ainsi limitée au seul prédicteur Intra dont la place dans le flux total est suffisamment importante pour démontrer un effet significatif de la méthode proposée.

En pratique, nous avons intégré notre schéma dans le codeur HEVC en compétition avec les modes Intra et Inter classiques. Notre solution est appliquée de deux façons selon les partition causales disponibles parmi celles décrites dans 7.1.3 :

- pour les images Intra prédites uniquement spatialement, nous utilisons les blocs voisins à gauche et au-dessus de la partition courante.
- pour les images Inter prédites temporellement ainsi que spatialement, nous utilisons les mêmes blocs voisins ainsi que les blocs causaux de l'image précédente pointés par les vecteurs de mouvement des quatre blocs voisins.

Nous évaluerons les différentes solutions liées à la détermination et la signalisation de cette partition optimale. La signalisation de notre méthode est quant à elle réalisée

comme le décrit la figure 7.5 : pour chaque unité de prédiction (PU) le mode de prédiction Intra, Inter ou Skip est signalé ; puis dans le cas d'une PU Intra, un indice additionnel indique si la PU est traitée de façon classique ou bien à l'aide de notre solution notée *Intra DI* (pour Intra Décodeur Intelligent). S'il s'agit d'une PU Intra classique, selon le partitionnement $2N \times 2N$ ou $N \times N$, un ou quatre prédicteurs Intra sont respectivement transmis alors que cette transmission est économisée dans le cas où le mode Intra DI est utilisé.

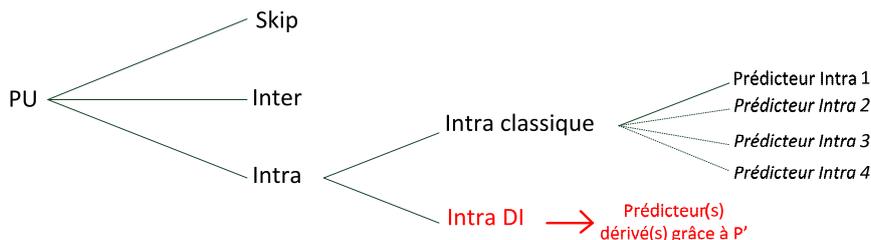


FIGURE 7.5: Modification de la signalisation HEVC intégrant le mode Intra DI proposé.

La détermination du prédicteur Intra optimal par le mode Intra DI consiste à substituer l'une des partitions causales candidates à la partition originale habituellement utilisée dans le critère débit-distorsion. L'ensemble du processus de sélection demeure alors équivalent à l'état de l'art : calcul du résiduel entre la partition prédictrice générée par chaque mode Intra et la partition causale candidate, suivi de l'étape de transformation et de quantification du résiduel avant de finalement calculer la distorsion entre la partition ainsi reconstruite et la partition causale. Dans la mesure où l'intégration d'un estimateur du coût Cabac au niveau du décodeur est relativement complexe, la composante de débit est quant à elle simulée à partir de l'entropie du résiduel généré. L'ensemble de ce traitement est ainsi exactement reproductible au décodeur.

Nous présentons et analysons les résultats obtenus pour cette mise en œuvre élémentaire mais néanmoins pertinente dans la sous-section suivante.

7.2.2 Résultats expérimentaux

7.2.2.1 Configuration de tests

Nous avons évalué la méthode dans l'encodeur HEVC HM 3.0 avec l'ensemble des outils activés afin de réellement mesurer le potentiel de notre méthode, excepté pour la sous-section 7.2.2.2 où une modification est apportée au nombre de prédicteurs Intra en compétition. Il est ainsi important de noter que cette expérimentation est mise en œuvre dans un encodeur très mature et à l'efficacité de codage nettement supérieure de celle de H.264/AVC (environ 40% de réduction du débit à qualité équivalente). Le profil utilisé est un des meilleurs profils actuellement considérés et il utilise une prédiction unidirectionnelle. Deux ensembles de QPs ont été évalués pour les moyens et bas débits

respectivement sur deux types de contenus : des séquences classiques issues du projet HEVC (7.2.2.2) puis des séquences de cartes de profondeur (7.2.2.3) qui présentent un intérêt particulier pour la méthode du fait des fortes corrélations entre partitions. Cette configuration est résumée ci-dessous :

Codeur HEVC HM 3.0.

Profil Low Delay High Efficiency (Prédiction uni-directionnelle).

QPs moyens débits : $QP = \{22, 27, 32, 37\}$; bas débits : $QP = \{27, 32, 37, 42\}$.

Séquences 8 séquences 2D du projet HEVC des formats qwvga à 1080p (7.2.2.2) et 5 séquences de cartes de profondeur (7.2.2.3).

Evaluation Pourcentage de réduction de débit moyen calculé à l'aide de la métrique de Bjontegaard.

7.2.2.2 Evaluation sur des séquences 2D

La première évaluation réalisée sur des séquences classiques est divisée en deux expérimentations permettant d'analyser respectivement les deux axes d'améliorations abordés par l'approche décodeur intelligent :

- réduction des coûts de signalisation des choix de codage.
- augmentation du nombre de choix en compétition.

Evaluation de l'efficacité de codage

Dans cette première sous-section nous évaluons la réduction des coûts de signalisation apportée par notre méthode au travers de l'amélioration de l'efficacité de codage dans une configuration complète du codeur HEVC.

Afin de confirmer le potentiel de la méthode et décider de la mise en œuvre la plus prometteuse par rapport à la problématique de signalisation de la partition causale à utiliser, nous mesurons tout d'abord les gains maximaux envisageables. Pour ce faire, nous proposons quatre tests dont il est important de préciser qu'ils sont non-décodables dans la mesure où la partition P'_i n'y est pas explicitement signalée :

- Test ND1 : mise en compétition du codage Intra classique de HEVC avec la méthode proposée où la partition causale utilisée est la partition originale ($P'_i = P_i$).
- Test ND2 : test précédent où la partition causale fait partie de l'ensemble $L_{P'}$ défini dans la section 7.2.1 et où le mode optimal hérité à l'aide de la méthode est déterminé pour la partition originale P_i parmi l'ensemble réduit des modes candidats sélectionnés sur les partitions causales de l'ensemble $L_{P'}$. Ce test permet d'évaluer le gain dans le cas où le coût de signalisation de la partition causale optimale est considéré comme nul.
- Test ND3 : test précédent où l'ensemble des partitions candidates est réduit aux deux partitions retournant les modes optimaux au sens du critère débit-distorsion.

Ce test permet d'évaluer une solution où un indice binaire serait suffisant pour indiquer la partition P'_i .

- Test ND4 : dans ce test, P'_i correspond à la partition de l'estimation de mouvement classique dont le vecteur de mouvement n'est pas transmis.

Les résultats de ces tests sont regroupés dans le tableau 7.1 pour les huit séquences HEVC testées à moyen débit. Les résultats du Test ND1 sont évidemment les plus favorables de façon significative et font état d'un gain potentiel maximum d'environ 10%. Ce test est loin d'être réaliste puisque la partition utilisée pour la détermination des prédicteurs Intra optimaux est la partition originale mais il délimite l'impact potentiel d'une telle approche dans un cas idéal en tenant compte de la signalisation additionnelle du nouveau mode basé sur notre méthode. Les tests suivants sont davantage réalistes puisqu'ils utilisent des prédicteurs hérités en utilisant une partition causale, c'est uniquement la signalisation de cette dernière qui n'est pas décodable. On observe dès le Test ND2 une diminution très significative des gains potentiels qui sont en moyenne de 1.1% pour ce second test où l'on utilise le meilleur prédicteur parmi l'ensemble des prédicteurs optimaux déterminés sur les partitions causales candidates. Il est intéressant de noter que les gains sont globalement équivalents pour l'ensemble des séquences testées avec une plage de valeurs allant de 0.9% à 1.3%. Le Test ND3 qui se calque sur le même principe mais en choisissant le meilleur prédicteur sur uniquement deux partitions causales valide notre façon de choisir les partitions optimales à l'aide d'un critère débit-distorsion global puisque les gains restent proches du test précédent avec 0.9% en moyenne. Enfin, le Test ND4 est peu concluant puisque sans transmettre les vecteurs de mouvement, l'utilisation de partitions obtenues par l'estimation de mouvement classique apporte seulement 0.4% de gains. Cette solution n'a par conséquent pas été étudiée davantage dans notre travail mais demanderait certainement une étude plus approfondie. En particulier, une estimation de mouvement plus adaptée, réalisée dans un autre domaine que le domaine pixel pourrait fournir des partitions causales davantage pertinentes.

	Test ND1	Test ND2	Test ND3	Test ND4
BasketBallPass qwvga	7.2	0.9	0.9	0.4
BlowingBubbles qwvga	9.2	0.9	0.7	0.3
RaceHorses qwvga	9.2	1.2	0.9	0.5
RaceHorses wvga	9.1	1.2	0.8	0.4
Vidyo 1 720p	13.5	0.9	0.8	0.5
Parkscene 1080p	8.7	1.3	1.2	0.5
Cactus 1080p	9.9	1.2	0.8	0.3
BasketBallDrive 1080p	12.7	1.3	1.0	0.4
Moyenne	9.9	1.1	0.9	0.4

TABLE 7.1: Pourcentages de réduction de débit maximums de notre méthode Intra DI pour quatre configurations non-décodables à moyen débit.

A titre de conclusion intermédiaire, l'approche, bien que très prometteuse au vue de la place conséquente des données traitées dans le flux total, se trouve notamment limitée du fait de la délicate étape de choix de la partition causale P'_i .

Des gains significatifs demeurent cependant envisageables à partir d'une solution utilisant un ensemble de partitions causales prédéfini comme dans le Test ND2 et sans transmission de signalisation additionnelle de la partition causale optimale.

A partir de cette conclusion, nous évaluons maintenant la méthode dans un cadre décodable où deux configurations ont été testées :

- Test 1 : Test ND2 avec sélection de la partition P'_i optimale à l'aide d'un critère débit-distorsion.
- Test 2 : Test ND2 avec sélection de la partition P'_i optimale en fonction du mode optimal le plus fréquent parmi les partitions candidates et en cas d'égalité, sélection selon le critère débit-distorsion précédent.

Ces résultats sont rapportés dans le tableau 7.2 à la fois pour les moyens débits à mettre en regard des résultats précédents mais aussi pour les bas débits qui donnent une indication intéressante des performances de la méthode pour un type de débit ou les modes Intra sont classiquement peu utilisés.

	Moyen Débit		Bas Débit	
	Test 1	Test 2	Test 1	Test 2
BasketBallPass qwvga	0.1	0.4	0.3	1.0
BlowingBubbles qwvga	0.3	0.3	0.6	0.8
RaceHorses qwvga	0.4	0.3	0.8	0.6
RaceHorses wvga	0.2	0.3	0.4	0.8
Vidyo 4 720p	0.5	0.3	0.4	0.9
Parkscene 1080p	0.4	0.5	0.4	0.6
Cactus 1080p	0.2	0.3	0.4	0.5
BasketBallDrive 1080p	0.2	0.4	0.3	0.6
Moyenne	0.3	0.4	0.5	0.7

TABLE 7.2: Pourcentages de réduction de débit moyen de notre méthode d'héritage des modes Intra basée décodeur intelligent pour deux configurations différentes à moyen et bas débit.

Tout d'abord, nous observons pour les moyens débits, un gain moyen de 0.3% et 0.4% respectivement pour le Test 1 et le Test 2.

La méthode apporte ainsi une amélioration systématique et non-négligeable pour le codage des séquences qui atteint notamment 0.5% pour la séquence Parkscene 1080p. Les gains sont plus élevés à bas débit où ils atteignent en moyenne 0.5% et 0.7% respectivement pour les deux expérimentations.

Une amélioration particulièrement intéressante pour les séquences *BasketBallPass* qwvga et *Vidyo 4* 720p est observée pour le second test. Ces gains supérieurs à bas débits s'expliquent par l'augmentation de l'utilisation de modes de prédiction Intra pour ce type de débit qui en contient sensiblement moins classiquement du fait du fort coût de signalisation des prédicteurs. Cette augmentation se répercute alors directement dans la qualité de la reconstruction des séquences. Pour les séquences aux plus forts gains, nous observons que les corrélations entre partitions sont très adaptées à notre choix réalisé sur des partitions causales prédéfinies. Par rapport aux deux tests évalués, le Test 2 semble être ici un peu plus performant, notamment à bas débit. Le prédicteur le plus sélectionné parmi l'ensemble des partitions causales candidates constitue en effet un bon pronostic du prédicteur réel et confirme la pertinence de notre liste de partitions candidates.

Afin d'analyser plus en détails ces résultats, nous proposons d'observer les pourcentages de sélection de notre mode de codage et ceux des modes Intra et Inter classique de HEVC pour le Test 2 à moyen débit. Ces statistiques sont rapportées dans la figure 7.6 avec le codeur de référence (a) puis le codeur modifié avec notre nouveau mode respectivement en moyenne sur l'ensemble des séquences (b), pour la séquence *BasketBallDrive* 1080p (c) et enfin la séquence *BasketBallPass* qwvga (d). Notons tout d'abord que pour l'ensemble des figures, la sélection des modes Intra (Intra seul pour (a) et l'union du mode Intra classique et le mode Intra introduit pour (b), (c) et (d)) augmente avec le pas de quantification. Précisons ici que ces pourcentages de sélection sont calculés en nombre d'unités de prédiction (notées PU, voir le chapitre 1 pour les détails de notations) codées avec un des trois types de modes et non en termes de pourcentage de la séquence globale. Ainsi, une PU de taille 4×4 a par exemple autant d'importance qu'une PU 64×64 dans ces statistiques afin de bien évaluer l'impact de notre méthode qui agit sur le codage de l'indice d'une PU qui s'avère avoir le même poids dans le flux pour ces deux tailles de PU. Cela explique l'augmentation de la proportion des modes Intra avec le QP alors que classiquement les modes Intra sont davantage utilisés à haut débit : les partitionnements Inter sont moins précis pour les forts QP alors que ceux des PU Intra demeurent relativement petits et c'est donc bien la proportion de PU Intra qui augmente avec le QP. Celle-ci atteint 18% pour le codeur de référence au QP 37 (a) confirmant le potentiel de gain de notre méthode et on observe qu'en considérant les deux modes Intra pour le codeur modifié, cette proportion est augmentée de 10% en moyenne pour l'ensemble des séquences (b). Le mode introduit à l'aide de notre approche est alors sélectionné à 18% pour ce même QP, soit le double du mode Intra classique. Il est intéressant de noter que sur l'ensemble des QPs, notre mode convertit non seulement des partitions codées classiquement en Intra mais aussi en Inter. La proportion de ces

dernières diminue ainsi de quelques pourcents à haut débit jusqu'à 10% à bas débit.

Finalement, les deux figures détaillant les statistiques pour deux séquences particulières (c, d) sont enrichissantes si on les met en regard des résultats précédents. Pour la séquence *BasketBallDrive* 1080p (c) où la prédiction Intra joue un rôle très important, nous observons alors une sélection très élevée de notre mode allant de 10% à 30% et des gains associés qui ne sont cependant pas les meilleurs avec 0.4% pour ces moyens débits. A l'inverse, une sélection plus limitée est observée pour la séquence *BasketBallPass* qwvga (d) qui fournit les mêmes résultats à moyen débit mais atteint 1.0% de gain à bas débits. L'étape de détermination du prédicteur à hériter semble donc être perfectible mais les pourcentages de sélection élevés associés à des gains prometteurs permettent globalement de valider l'intérêt de notre approche.

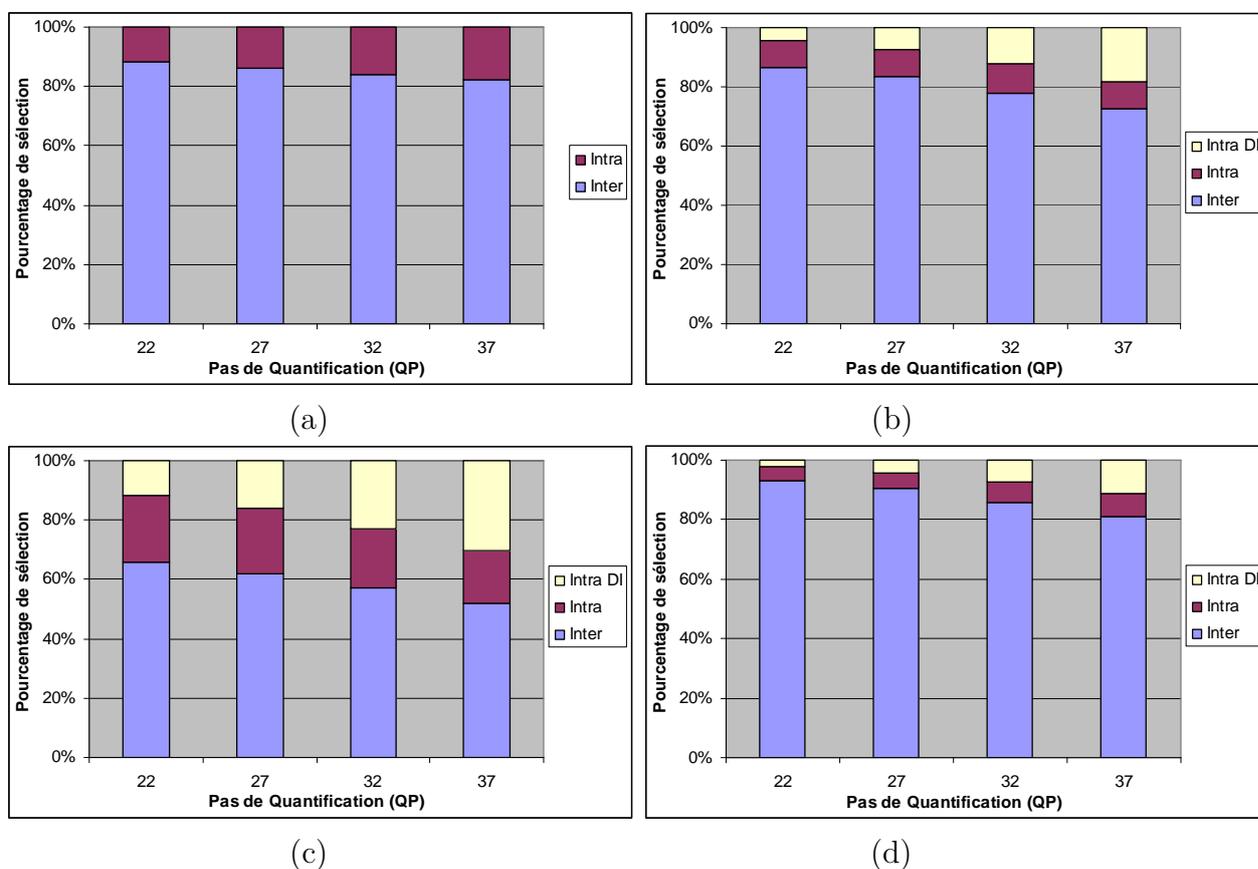


FIGURE 7.6: Pourcentage de sélection des différents modes de codage selon le QP calculés en moyenne pour l'ensemble des images Inter des séquences testées. Statistiques obtenues avec le codeur de référence (a), et le codeur modifié avec notre nouveau mode en moyenne sur l'ensemble des séquences (b) ; pour la séquence *BasketBallDrive* 1080p (c) ; et la séquence *BasketBallPass* qwvga (d).

Enfin, nous évaluons avec la figure 7.7 la répartition de la sélection des modes en fonction du type d'image et du partitionnement des unités de prédiction. Ces statistiques sont calculées en moyenne pour les quatre QP de 22 à 37. Par rapport aux types d'images, on observe qu'un plus grand nombre de PU sont codées à l'aide du mode proposé dans les images Intra, atteignant 28% pour les partitionnements $2N \times 2N$ contre 12% dans les images Inter. Ce résultat confirme l'intérêt de l'approche puisque, évaluée uniquement avec des modes Intra similaires mais nécessitant la transmission des indices des prédicteurs, notre solution est très souvent sélectionnée malgré l'utilisation de partitions causales potentiellement moins corrélées que pour les images Inter (les PU de gauche et du dessus). La sélection de notre mode parmi les images Inter demeure néanmoins significative et il est très intéressant d'observer que parmi les PU $2N \times 2N$, notre mode y est davantage sélectionné que le mode Intra classique (3% de sélection supplémentaire). Ceci s'explique car notre méthode est plus efficace pour ce type de partition comme on a aussi pu l'observer pour les images Intra du fait de l'unique prédicteur à déterminer. À l'inverse, les PU $N \times N$ nécessitent de déterminer quatre prédicteurs Intra et l'efficacité de notre méthode est limitée dans ce cas car l'héritage d'un mauvais prédicteur entraîne une mauvaise reconstruction de la partition associée ce qui peut avoir un effet très négatif sur la composante de distorsion. Le mode proposé est alors peu sélectionné malgré le gain associé en débit du fait de la non-transmission des quatre indices de prédicteurs. Une sélection non-négligeable de 13% et 5% est néanmoins observée pour ce type de PU respectivement en Intra et en Inter.

Ces forts pourcentages de sélection prouvent globalement l'intérêt de notre approche pour laquelle l'utilisation de partitions causales davantage corrélées demeure l'enjeu majeur.

Evaluation de l'influence du nombre de prédicteurs en compétition

Nous souhaitons maintenant évaluer la seconde voie d'amélioration ouverte par notre approche qui concerne l'augmentation des paramètres en compétition. Pour ce faire, nous restreignons le nombre de prédicteurs Intra en compétition dans le mode classique à 4 prédicteurs (Horizontal, Vertical, DC, Planar) tout en laissant un nombre élevé de prédicteurs mis en compétition dans le mode défini à l'aide de notre approche décodeur intelligent. Le tableau 7.3 présente les résultats obtenus pour ce test réalisé à moyen débit avec respectivement 4, 17 et 34 prédicteurs utilisés dans la méthode décodeur intelligent. Ces trois configurations sont comparées au schéma de référence avec les 4 prédicteurs Intra uniques définis ci-dessus. Notons que bien que cela n'ait pas été évalué au cours de cette thèse, il serait envisageable d'ajouter encore d'autres prédicteurs en compétition pour par exemple doubler le nombre de directions qui serait porté à 68 comme cela a été évoqué au début du développement d'HEVC.

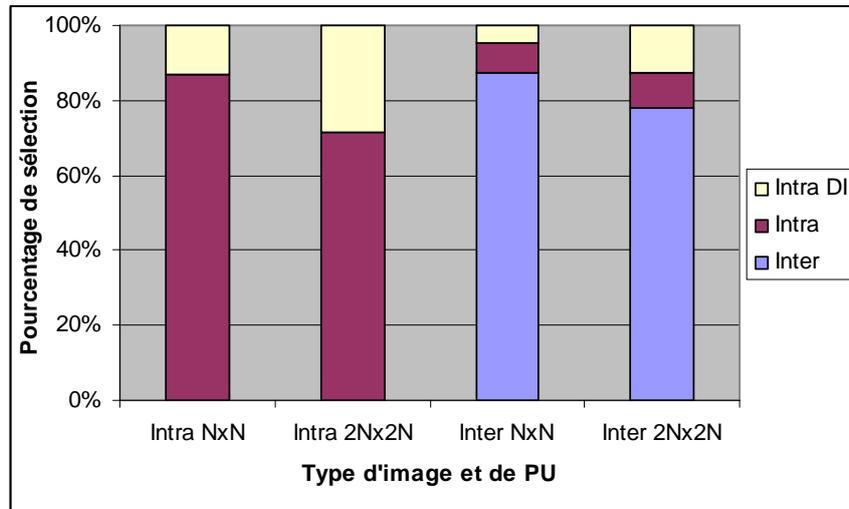


FIGURE 7.7: Pourcentages de sélection des différents modes de codage selon le type d'image (Intra ou Inter) et de partition ($N \times N$ ou $2N \times 2N$). Moyenne calculée à partir des résultats des quatre QP.

	4 pred	17 pred	34 pred
BasketBallPass qwvga	0.1	0.2	0.1
BlowingBubbles qwvga	0.4	0.5	0.3
RaceHorses qwvga	0.2	0.6	0.6
RaceHorses wvga	0.1	0.6	0.4
Vidyo 1 720p	-0.1	0.9	0.7
Parkscene 1080p	0.1	0.3	0.4
Cactus 1080p	0.3	0.5	0.7
BasketBallDrive 1080p	0.1	0.7	0.7
Moyenne	0.2	0.5	0.5

TABLE 7.3: Pourcentages de réduction de débit moyen de notre méthode décodeur intelligent appliquée au mode Intra pour différents nombres de prédicteurs en compétition (4, 17 et 34). Test réalisé pour un ensemble de QP à moyen débit.

Tout d'abord, notons que les gains relativement faibles de cette étude forcent à prendre des précautions pour tirer des conclusions et précisons de plus que le nombre de prédicteurs dans le mode classique étant de 4, ces derniers sont codés sur au maximum 2 bits tandis que la signalisation entre ce mode classique et notre mode est réalisée pour sa part sur 1 bit. La marge est par conséquent beaucoup plus limitée par rapport aux tests précédents (7.2.2.2) où le mode classique pouvait utiliser jusqu'à 6 bits. Néanmoins,

nous observons de façon suffisamment remarquable l'apport de l'augmentation de 4 à 17 prédicteurs en compétition dans notre mode basé décodeur intelligent. Cet accroissement des choix qui ne nécessite pas de signalisation additionnelle permet en effet une amélioration du gain sur le débit de 0.3% en moyenne pour un gain total de 0.5% avec 17 prédicteurs. Cette amélioration est systématique pour l'ensemble des séquences et atteint 1.0% pour la séquence Vidyo 1 720p qui tire particulièrement profit de l'apport des nouvelles directions introduites à l'aide de cette seconde configuration. La dernière configuration où les 34 directions de HEVC sont mises en compétition n'apporte globalement pas d'amélioration supplémentaire puisque le gain moyen demeure de 0.5%. Cela semble venir de la qualité limitée des partitions causales utilisées pour ces tests qui ne permettent pas d'exploiter suffisamment l'ajout d'un nombre très élevé de prédicteurs. Ce gain est cependant légèrement supérieur au gain obtenu dans la section précédente où les résultats de référence étaient générés avec un codeur ayant lui aussi 34 prédicteurs en compétition (tableau 7.2).

Ces observations tendent donc à confirmer l'intérêt de notre approche qui permet ici d'augmenter le nombre de prédicteurs en compétition pour améliorer la qualité de la reconstruction sans impacter sur le coût de signalisation de ces prédicteurs.

Afin d'apprécier l'impact en termes de sélection des différents modes, nous avons représenté leurs pourcentages de sélection sur la figure 7.8 pour les trois configurations testées. Ces pourcentages sont conformes aux résultats du tableau 7.3 avec une augmentation de la sélection du mode Intra DI de 1% pour le passage de 4 à 17 prédicteurs accompagnée d'une très légère augmentation de la sélection du mode Intra classique qui tire certainement profit d'un meilleur codage de ses prédicteurs. Enfin, la répartition de la sélection des modes reste quasiment identique avec la dernière configuration à 34 prédicteurs.

7.2.2.3 Evaluation sur des séquences de cartes de profondeur

Dans cette dernière évaluation de notre méthode, nous avons souhaité appliquer notre solution au codage de cartes de profondeur (en anglais, *Depth-Map*) utilisées en codage multi-vues. Ces cartes de profondeur [Shade *et al.* 98, Yoon *et al.* 07] associées à une vidéo de texture permettent de représenter une scène 3D que ce soit pour des applications de vidéos en relief ou de sélection de point de vue. La vidéo 2D classique apporte ainsi les informations de texture, d'intensité des couleurs ainsi que de structure de la scène tandis que la vidéo de profondeur définit la distance pour chaque pixel entre la caméra et les différents points 3D de la scène. Typiquement, les cartes de profondeur sont des images en niveau de gris où la valeur 255 est associée aux points les plus proches de la caméra et la valeur 0 aux points les plus éloignés (pour une représentation classique avec 8 bits). Leur principal avantage par rapport au format multi-vues classique est de permettre la

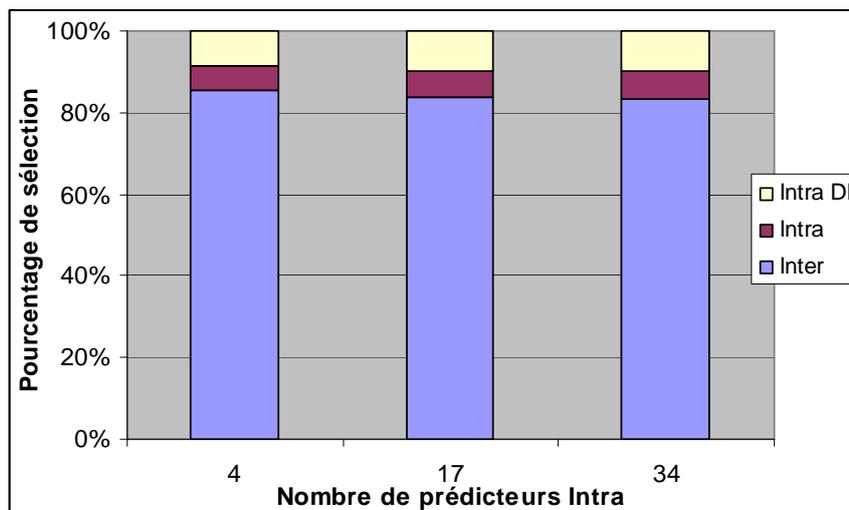


FIGURE 7.8: Pourcentages de sélection des différents modes de codage en fonction du nombre de prédicteurs en compétition dans la méthode proposée. Moyenne calculée à partir des résultats des quatre QP.

synthèse d'un nombre important de vues à partir d'un nombre limité de vues en entrée et de découpler les étapes de cartes de profondeur véritable enjeu pour le développement de cette solution. Le standard MVC (*Multi-view Video Coding*) [Merkle *et al.* 07] permet actuellement le codage des cartes de profondeur mais n'exploite pas l'ensemble de leurs particularités. Ainsi, un appel à proposition pour un nouveau standard de compression de vidéos 3D a récemment eu lieu (résultats en décembre 2011) et ce sont des séquences issues de cet appel que nous avons utilisées pour l'évaluation de notre méthode sur ce contenu particulier.

Les cartes de profondeur sont particulièrement adaptées à la solution décodeur intelligent proposée dans cette section pour deux raisons :

- La proportion des prédicteurs Intra dans le flux est plus élevée que sur des contenus classiques de type texture (20% en moyenne pour les cartes de profondeur contre 10% pour la texture). En effet, dans la mesure où les prédictions Intra sont efficaces plus souvent en générant des résiduels de pixels très faibles voir nuls, il est alors souvent moins coûteux de signaler un prédicteur Intra plutôt qu'un vecteur de mouvement.
- Les corrélations sont très fortes entre les blocs appartenant à un même niveau de profondeur. Les blocs causaux utilisés pour notre approche sont alors plus pertinents que sur des séquences classiques.

Afin de réellement percevoir les gains apportés pour le codage des cartes de profondeur, nous avons évalué en parallèle le codage des vidéos de textures associées. L'ensemble de ces résultats sont rapportés dans le tableau 7.4.

	Carte de profondeur	Texture
GT_Fly 1920×1088	0.6	-0.1
PantoMime 1280×960	1.4	0.1
Kendo 1024×768	1.0	-0.1
Balloons 1024×768	0.0	0.4
LoveBird1 1024×768	0.8	0.2
Moyenne	0.8	0.1

TABLE 7.4: Pourcentages de réduction du débit moyen de notre méthode appliquée aux séquences de cartes de profondeur et textures associées.

L'observation principale est que des gains supérieurs sont obtenus avec 0.8% de réduction du débit en moyenne pour les cartes de profondeur contre seulement 0.1% pour les séquences de texture associées.

Cette différence est notamment très forte pour la séquence *PantoMime* où elle atteint 1.3% tandis qu'à l'inverse, la méthode ne donne pas de gain sur la carte de profondeur de la séquence *Balloons* alors qu'un gain de 0.4% est rapporté en texture. Pour expliquer ces résultats, précisons de plus que les cartes de profondeur sont généralement moins coûteuses que les séquences classiques et les méthodes adaptées comme celle introduite dans ce chapitre ont alors davantage d'impact sur l'efficacité de codage globale.

A titre d'exemple, la figure 7.9 présente une image de la séquence *Kendo* compressée avec la méthode à QP 32. On y observe clairement le second point propre aux cartes de profondeur décrit précédemment : les zones appartenant à une même profondeur sont très corrélées, ce qui assure de bonnes partitions causales P'_i pour notre méthode.



FIGURE 7.9: Image compressée avec la méthode proposée au QP=32 tirée de la séquence de carte de profondeur Kendo 1024×768.

7.3 Perspectives

L'approche présentée dans ce chapitre est de type prospective et l'application aux modes Intra bien que prometteuse demeure le résultat d'une étude préliminaire. Avant d'approfondir le potentiel de la méthode dans une autre mise en pratique au chapitre suivant, nous souhaitons présenter plusieurs perspectives intéressantes d'ordre générale mais aussi liées à l'application Intra.

7.3.1 Perspectives générales

Au-delà des améliorations des étapes de sélection de la partition causale et de la mise en compétition des choix de codage, une perspective ambitieuse concernant l'intégration de l'approche au sein du schéma de codage est à envisager. Comme nous l'avons évoqué dans 7.1.5, notre méthode permet de réaliser une prédiction des moyens de codage et il est par conséquent possible de poursuivre la démarche de recherche d'une dualité avec la prédiction de la texture en transmettant un résiduel de prédiction de modes. Cette idée est déjà exploitée avec l'utilisation du « most probable mode » en Intra qui permet de coder à coût réduit le mode le plus probable déterminé à l'aide du voisinage. L'expérimentation réalisée dans 7.2 nous a montré que les prédicteurs déterminés par la méthode ne sont pas systématiquement optimaux et il pourrait donc être intéressant de les corriger.

Une seconde voie d'étude ouverte par notre méthode concerne la complexité. En effet, l'approche proposée reporte uniquement sur la complexité les questions liées à l'ajout de nouveaux modes en compétition. Ainsi, encodeur et décodeur fonctionnant alors « ensemble », un nouvel angle d'approche est notamment ouvert pour le sujet de la scalabilité en complexité. Actuellement, la complexité des encodeurs est gérée par la définition de profils dont les plus simples sont constitués d'un ensemble réduit des outils disponibles ce qui limite par conséquent l'efficacité de codage pour ces profils. En alternative, certains travaux récents [Nam *et al.* 10, Lee *et al.* 09] s'intéressent à ce sujet afin de proposer de nouvelles solutions offrant une complexité variable en fonction du matériel utilisé tout en garantissant une bonne qualité d'encodage. Différentes approches peuvent être envisagées mais dans le cadre du schéma proposé dans ce chapitre, il serait alors aisé de modifier l'ensemble de modes en compétition sans impacter négativement sur le coût de signalisation associée dans la mesure où celle-ci est économisée. A l'inverse, s'il est possible actuellement de faire le choix de restreindre les modes testés à l'encodeur, le nombre de bits alloué à leur signalisation est pour sa part incompressible même si un tel choix est pris en compte par le codeur entropique. L'encodeur signale alors des choix sans prendre en compte qu'une partie des modes n'est pas accessible ce qui est sous-optimal en termes d'efficacité de codage.

Afin d'étudier ce sujet qui représente un fort enjeu au moment où la complexité des encodeurs devient de plus en plus élevée, une thèse sur le sujet va débutée au sein de

l'équipe codage vidéo avancé d'Orange Labs. L'étude de la scalabilité en complexité dans le cadre de l'approche décodeur intelligent constitue ainsi une des voies de recherche envisagée.

7.3.2 Perspectives dédiées à l'application aux modes Intra

Concernant plus précisément l'application aux modes Intra, deux voies d'améliorations ambitieuses principales pourraient être étudiées. Elles concernent la sélection de P'_i d'une part et l'ajout de prédicteurs en compétition d'autre part.

Une solution particulièrement prometteuse quant à la sélection de la partition causale consisterait ainsi à réaliser l'estimation de mouvement dans le domaine des prédicteurs Intra. Comme représenté de façon schématique sur la figure 7.10, la partition P'_i définie est la partition dont les modes de codage sont les plus proches des modes optimaux de la partition courante. La zone causale où est réalisée la mise en correspondance est alors représentée en terme de modes de prédiction et non plus de pixels et après avoir déterminé les prédicteurs optimaux pour la partition courante, ces derniers sont recherchés au sein de la zone causale. En pratique cela peut être fait en associant un indice à chaque mode de façon judicieuse afin d'utiliser directement un algorithme d'estimation de mouvement classique, les directions de prédiction voisines seront ainsi associées à des valeurs proches. Cette solution augmente cependant de façon significative la complexité qui pourrait être par ailleurs régulée par la taille des partitions sur lesquelles la recherche est effectuée.

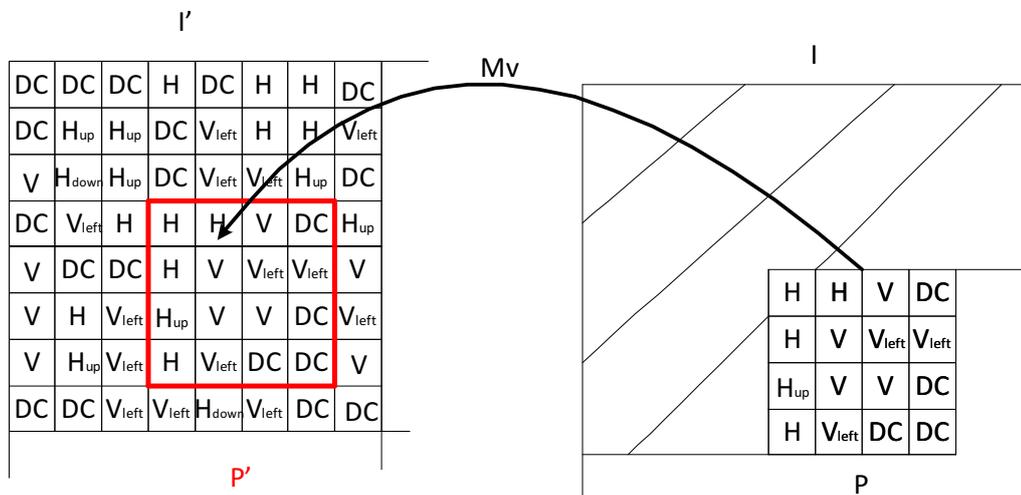


FIGURE 7.10: Exemple d'estimation de mouvement dans le domaine des modes Intra.

La seconde perspective que nous envisageons est l'ajout de nouveaux modes de prédiction afin d'exploiter davantage la souplesse offerte par l'approche décodeur intelligent. Nous avons en effet mis en évidence dans la seconde partie des résultats 7.2.2.2 que l'augmentation de choix en compétition améliore l'efficacité de la méthode en passant de

quatre à dix-sept directions de prédictions. Le passage à trente-quatre prédicteurs ne permet cependant pas d'augmenter à nouveau les gains et nous pensons par conséquent que l'ajout de nouveaux prédicteurs dans le cadre de cette approche doit passer par des nouveaux modes plutôt que d'autres directions. Il est en effet délicat de départager des prédicteurs trop proche à l'aide de la méthode du fait de la qualité limitée de la partition P'_i mais des prédicteurs davantage diversifiés seraient particulièrement pertinents. De nombreux travaux proposant des nouveaux prédicteurs Intra ont justement été réalisées ces dernières années, citons notamment les approches parcimonieuses [Martin *et al.* 07], l'utilisation de résolutions multiples [Park *et al.* 07], la création de prédictions pondérées générées par un gradient ou de prédicteurs bi-dimensionnels [Matsuo *et al.* 09]. La plupart de ces types de travaux ont en commun leur mise en œuvre par substitution à un autre mode afin de ne pas engendrer de surcoût de signalisation. Chaque méthode a ainsi prouvé son efficacité pour des contenus spécifiques mais elles demeurent limitées pour être utilisées dans un standard du fait du coût de signalisation associé. Par conséquent, leurs ajout dans le cadre de notre approche prendrait tous son sens, chaque méthode pouvant alors apporter un gain en qualité de prédiction non-compensé par la transmission du choix optimal qui serait dérivé lors du décodage. Enfin, afin de contraindre l'augmentation de la complexité ajoutée par un nombre très élevé de prédicteurs, l'ensemble des modes en compétition pourrait être adapté au cours du traitement en fonction d'un classement réalisé identiquement par l'encodeur et le décodeur.

7.4 Bilan

En rupture avec les schémas de codage conventionnels, nous avons introduit dans ce chapitre une approche intitulée *décodeur intelligent* visant à exploiter davantage le décodeur. À l'aide d'une partition causale judicieusement sélectionnée, différents paramètres de codage sont hérités pour la partition courante qui bénéficie alors d'une réduction significative des coûts de signalisation. Nous entrevoyons de ce fait un schéma permettant non-plus de prédire les pixels mais bel et bien les modes de codage dont la systématique augmentation peut devenir une limite aux futurs codeurs. Cette approche nécessite cependant de nombreuses modifications des schémas de compression actuels, en particulier au niveau du décodeur qui devient acteur du système. Plusieurs de ces questions ont été abordées au cours de ce chapitre comme la sélection de la partition causale à utiliser et sa signalisation.

Dans le cadre d'une évaluation préliminaire du potentiel de la méthode, une application aux prédicteurs Intra dans le codeur HEVC a ensuite été réalisée. Cette étude, bien qu'élémentaire, a permis de révéler des gains prometteurs au sein d'un encodeur aux performances déjà extrêmement élevées. La méthode apporte notamment des gains significatifs à hauts débits et fait état d'un fort pourcentage de sélection atteignant 18% en moyenne pour ce type de débits. Ces promesses ont été finalement mises en perspectives par différentes voies d'améliorations dont certaines sont mises en œuvre dans le

chapitre suivant où nous proposons une application plus ambitieuse de cette approche à la théorie de l'échantillonnage compressé. Les méthodes d'échantillonnage compressé s'inscrivent en effet idéalement dans le cadre de l'approche décodeur intelligent car elles nécessitent un paramétrage abondant et minutieux, en particulier lors d'une utilisation dans un but de compression.

Chapitre 8

Approche décodeur intelligent appliquée à l'échantillonnage compressé pour le codage vidéo

Nous avons introduit l'échantillonnage compressé au chapitre 6, il s'agit d'une théorie émergente du traitement du signal qui affirme : sous certaines conditions, uniquement un faible nombre de mesures d'un signal projeté dans une base donnée sont nécessaires pour obtenir une bonne reconstruction du signal d'origine. De nombreuses applications en traitement du signal ont vu le jour dont quelques travaux en compression vidéo comme nous l'avons présenté précédemment.

Dans ce chapitre, nous étudions l'intégration d'une telle méthode dans un procédé de codage vidéo classique. Afin de proposer une solution aux problèmes rencontrés par une telle mise en œuvre, la méthode est étudiée dans le cadre de l'approche décodeur intelligent décrite dans le chapitre précédent. Cette approche est en effet particulièrement adaptée aux méthodes itératives comme l'échantillonnage compressé du fait des nombreux paramètres nécessaires pour assurer l'efficacité de la méthode. Après la description de la méthode, nous proposons une première expérimentation théorique définissant un schéma de codage uniquement basé sur notre solution que nous évaluons finalement au sein du futur standard HEVC.

8.1 Description de la méthode proposée

8.1.1 Schéma global

L'état de l'art sur l'échantillonnage compressé et son application à un signal de type vidéo (6.2.1) a tout d'abord montré que cette nouvelle théorie possède de nombreux avantages. Nous avons ensuite vu que l'élaboration d'un schéma de codage vidéo utilisant l'échantillonnage compressé nécessite une habile sélection des nombreux paramètres

et des opérateurs afin d'optimiser l'efficacité de la méthode. Enfin, l'exploitation des redondances spatiales et surtout temporelles nous a semblé être un autre point clé d'un tel schéma.

Cependant, nous souhaitons mettre en avant deux points particuliers relatifs aux méthodes existantes :

1. la notion d'adaptativité n'est pas ou peu considérée : selon le contenu du signal, des stratégies différentes doivent être mises en œuvre pour optimiser le rapport entre nombre de mesures collectées et qualité de la reconstruction.
2. la transmission du masque utilisé par l'opérateur d'acquisition Φ est une limite majeure de l'échantillonnage compressé qui est principalement résolue par l'utilisation de matrices aléatoires ou bien par la définition d'un masque « par défaut ». Ces solutions ne sont cependant pas optimales en termes de nombre de mesures collectées. A l'inverse, un masque optimal entraîne un surcoût de transmission rendant difficile la mise en œuvre d'un schéma de compression compétitif.

En considérant ces observations et en nous appuyant sur l'approche décodeur intelligent introduite précédemment, nous proposons dans ce chapitre un schéma de codage vidéo basé sur l'échantillonnage compressé. Dans ce contexte, notre approche apparaît en effet particulièrement adaptée au vu des limites dues au délicat paramétrage de la méthode d'une part et d'autre part dans le but de proposer une solution innovante exploitant les redondances spatiales et temporelles. Nous verrons au cours de cette étude comment notre solution répond notamment aux deux points introduits ci-dessus.

Calqué sur le plan du chapitre précédent nous introduisons tout d'abord le schéma proposé à partir des notations et de l'équation de reconstruction de l'échantillonnage compressé que nous adaptons à l'environnement de l'approche décodeur intelligent. Nous détaillons ensuite les paramètres considérés avant de présenter à nouveau les trois étapes de l'approche décodeur intelligent adaptées à notre nouvelle application.

Nous proposons d'utiliser le traitement par échantillonnage compressé dans le cadre introduit par Gan [Gan 07] : chaque image I est divisée en partitions P_i traitées séquentiellement, typiquement des blocs. Pour chaque partition, nous réalisons l'acquisition d'un nombre de mesures inférieur au nombre de pixels de la partition, puis la reconstruction est réalisée par résolution de l'équation 6.6. Ce traitement est guidé par un 4-uplet de paramètres, à savoir :

$$\forall P_i, L_i = (\Phi_i, \Psi_i, S_i, \Lambda_i) \in \mathcal{F}_\Phi \times \mathcal{F}_\Psi \times [0, 100]_{\mathbb{N}} \times \mathbb{R}^k \quad (8.1)$$

où :

- Φ_i représente l'opérateur d'acquisition du signal.
- Ψ_i la base transformée utilisée pour la reconstruction du signal.
- S_i le pas de sous-échantillonnage définissant le nombre de mesures conservées.
- Λ_i l'ensemble des paramètres propres à la méthode de reconstruction par résolution de l'équation 6.6.

L'ensemble de ces paramètres et leurs interactions sont détaillés dans la section 8.1.2.

Pour chacune des partitions courante P_i , une stratégie différente pour la mise en œuvre du schéma d'échantillonnage compressé est utilisée. Nous cherchons ainsi à déterminer la liste des paramètres optimaux \hat{L}_i telle que la partition P_i soit la mieux reconstruite possible avec à la fois un minimum de mesures y_m et un coût de signalisation minimal de ces paramètres \hat{L}_i . Cette liste optimale est déterminée par la résolution du problème (\mathcal{P}_1) suivant :

$$(\mathcal{P}_1) \begin{cases} \text{Min} [\|\Psi_i P_i\|_1 + \lambda \|\sum_m y_m - \Phi_i P_i\|_2 + \nu(\varsigma_y(y_m) + \varsigma_L(L_i))] \\ L_i \in \mathcal{F}_\Phi \times \mathcal{F}_\Psi \times [0, 100]_{\mathbb{N}} \times \mathbb{R}^k \end{cases}, \quad (8.2)$$

où $\varsigma_y(y_m)$ représente le coût associé aux mesures y_m , $\varsigma_L(L_i)$ représente le coût associé à la transmission des paramètres de la liste L_i et ν est un paramètre fixé permettant de pondérer le poids de la somme des coûts des données transmises.

L'efficacité d'une telle méthode dépend de la pertinence des paramètres en compétition ainsi que du nombre de choix disponibles pour chacun de ces paramètres. Ce second point a un impact direct sur l'efficacité dans la mesure où un nombre plus élevé de choix fait augmenter le coût de signalisation et limite par conséquent les gains apportés par la méthode. Evidemment, ce coût, même avec peu de paramètres en compétition, est déjà pénalisant pour la méthode. Notre solution consiste donc à se placer dans le cadre de l'approche décodeur intelligent. Nous utilisons une partition causale P'_i judicieusement sélectionnée pour déterminer la liste L_i^* des paramètres, dits *quasi-optimaux*, à utiliser pour le traitement de la partition courante P_i . Finalement, le problème d'optimisation considéré (\mathcal{P}_2) devient :

$$(\mathcal{P}_2) \begin{cases} \text{Min} [\|\Psi_i P'_i\|_1 + \lambda \|\sum_m y_m - \Phi_i P'_i\|_2 + \nu(\varsigma_y(y_m) + \varsigma_{Vm}(Vm_i))] \\ L_i \in \mathcal{F}_\Phi \times \mathcal{F}_\Psi \times [0, 100]_{\mathbb{N}} \times \mathbb{R}^k \end{cases}, \quad (8.3)$$

où Vm_i représente le vecteur mouvement pointant vers la partition P'_i et $\varsigma_{Vm}(Vm_i)$ le coût de signalisation associé. Ce coût se substitue au coût des paramètres $\varsigma_L(L_i)$ présent dans l'équation 8.2 : au lieu de transmettre la liste L_i^* , nous signalons dans le flux le vecteur mouvement Vm_i permettant de retrouver de manière exacte les paramètres L_i^* lors du décodage. Comme évoqué dans le chapitre 7, rappelons que de manière plus générale $\varsigma_{Vm}(Vm_i)$ représente le coût de l'information de mouvement (prédicteur de mouvement, vecteur résiduel, image de référence...) permettant de déterminer la partition P'_i , ce coût pouvant éventuellement être nul comme cela est détaillé dans la section 8.1.3.1.

Sous une hypothèse de forte corrélation entre les partitions P_i et P'_i , la liste de paramètres L_i^* constitue une bonne approximation de \hat{L}_i . La partition courante P_i est alors efficacement reconstruite avec un minimum de mesures et un surcoût de signalisation limité à l'indication de la partition causale P'_i à l'aide de l'information de mouvement

Vm_i . De nouveau, nous nous retrouvons avec le problème de la signalisation du mouvement qui est un sujet largement traité par la communauté du codage vidéo comme nous avons pu le voir dans le chapitre 3. Nous pouvons ainsi nous concentrer sur la compétition permettant de déterminer les paramètres de codage optimaux pour lesquels il est donc important d'insister sur l'absence de limite liée au coût pour le nombre de paramètre en compétition. La limite demeure uniquement liée à la complexité. En particulier, le pas de sous-échantillonnage S_i^* peut être défini de façon très précise pour acquérir uniquement le nombre nécessaire de mesures sans se soucier du coût de codage tandis que si celui-ci il était transmis directement dans le flux, on le limiterait à un sous-ensemble de valeurs fixes pour des problèmes de signalisation.

En regard de la figure 7.2 du chapitre précédent, la méthode est résumée sur la figure 8.1.

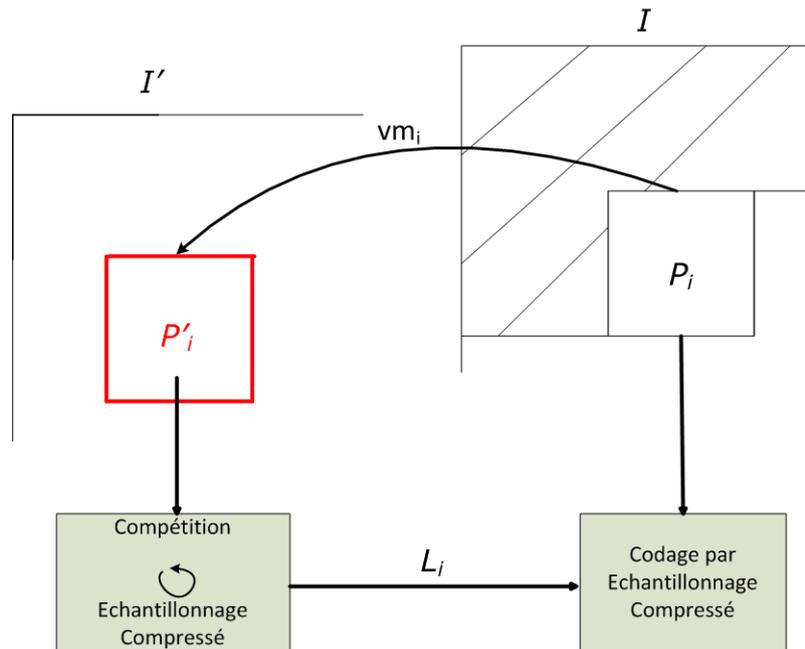


FIGURE 8.1: Schéma simplifié de la méthode d'échantillonnage compressé combinée avec l'approche décodeur intelligent.

Pour le codage de la partition P_i appartenant à l'image I , une partition P'_i appartenant à la zone causale (ici, l'image précédente I') est déterminée puis signalée au décodeur à l'aide du vecteur Vm_i . Le même procédé est alors appliqué au niveau de l'encodeur comme du décodeur : une compétition sur l'ensemble des paramètres L_i est réalisée puis la liste des paramètres optimaux L_i^* ainsi déterminée est utilisée pour le traitement par échantillonnage compressé de la partition P_i . Comme représenté ici pour l'encodage, ce traitement consiste en l'acquisition puis la reconstruction selon les paramètres L_i^* où l'étape de reconstruction permet de plus de rendre la partition disponible pour la suite

du codage. De façon inverse, lors du décodage la reconstruction est obtenue à partir des mesures lues dans le flux et des mêmes paramètres optimaux L_i^* . Ainsi, on économise comme attendu le coût de transmission des paramètres L_i^* au prix de la signalisation de la partition causale P_i' .

Avant de détailler les différentes parties de la méthode proposée, nous présentons les paramètres mis en compétition dans la sous-section suivante.

8.1.2 Constitution de la liste des paramètres en compétition

A l'aide de la méthode proposée, plusieurs paramètres sont mis en compétition sur une partition causale P_i' préalablement déterminée afin de coder par la suite la partition courante P_i . Comme introduit dans la section 8.1.1, la liste L_i de ces paramètres est définie comme suit pour la partition courante P_i :

$$\forall P_i, L_i = (\Phi_i, \Psi_i, S_i, \Lambda_i) \in \mathcal{F}_\Phi \times \mathcal{F}_\Psi \times [0, 100]_{\mathbb{N}} \times \mathbb{R}^k. \quad (8.4)$$

Les paramètres les plus influents de cette méthode sont :

- Φ_i , l'opérateur d'acquisition qui permet de s'adapter au contenu de chaque partition.
- S_i , le pas de sous-échantillonnage qui régule le débit de la méthode.

Nous définissons l'opérateur Φ_i^S comme la combinaison de l'opérateur d'échantillonnage Φ_i et du masque d'échantillonnage M_{S_i} généré pour un pas de sous-échantillonnage S_i :

$$\begin{aligned} \Phi_i^S : \mathbb{R}^{N \times N} &\longrightarrow \mathbb{R}^{N \times N} \\ f &\longmapsto (M_{S_i} \circ \Phi_i)(f), \end{aligned} \quad (8.5)$$

où le masque M_{S_i} conserve un pourcentage S_i d'éléments non nuls du signal f de taille $N \times N$, M_{S_i} est défini par :

$$\begin{aligned} M_{S_i} : \mathbb{R}^{N \times N} &\longrightarrow \mathbb{R}^{N \times N} \\ f &\longmapsto M_{S_i}(f) \text{ tel que : } \text{Card}(M_{S_i}(f) \mid M_{S_i}(f) \neq 0) \leq S_i \times \text{Card}(f). \end{aligned} \quad (8.6)$$

Il existe un nombre élevé d'opérateurs vérifiant la définition ci-dessus pour un pas d'échantillonnage donné S_i . Dans le cadre de notre étude, toutes les possibilités de masque pour chaque pas de sous-échantillonnage ne sont pas testées afin de limiter la complexité de la méthode. Nous avons ainsi défini un masque adapté pour chaque opérateur d'acquisition que nous introduisons ci-après en relation avec les deux opérateurs Φ_i considérés au cours de notre étude :

DCT le masque utilisé conserve un pourcentage S_i du nombre d'éléments total pris parmi les premières mesures de la partition parcourue à l'aide d'un ordre de parcours en zig-zag débutant en haut à gauche du bloc. Les coefficients situés à ces

positions correspondent aux basses fréquences et sont en effets les plus significatifs comme nous l'avons exposé dans le chapitre 1 avec la figure 1.4 notamment.

Ondelette nous proposons deux masques, le premier est directement hérité de la partition causale et correspond au masque conservant un pourcentage S_i des éléments les plus significatifs :

$$M_{S_i} : \mathbb{R}^{N \times N} \longrightarrow \mathbb{R}^{N \times N} \quad (8.7)$$

$$f \longmapsto M_{S_i}(f) \text{ tel que : } \text{Card}(M_{S_i}(f) \mid |M_{S_i}(f)| > \tau) \leq S_i \times \text{Card}(f),$$

où τ est choisi tel qu'il y ait une proportion S_i d'éléments supérieurs à τ en valeur absolue.

Le second masque proposé ajoute la contrainte de conserver la totalité des coefficients de la dernière sous-bande où sont en effet regroupés les valeurs les plus significatives dites coefficients d'approximation.

Les exemples des masques correspondants sont introduits dans la figure ci-dessous pour un sous-échantillonnage fixé à 50%.

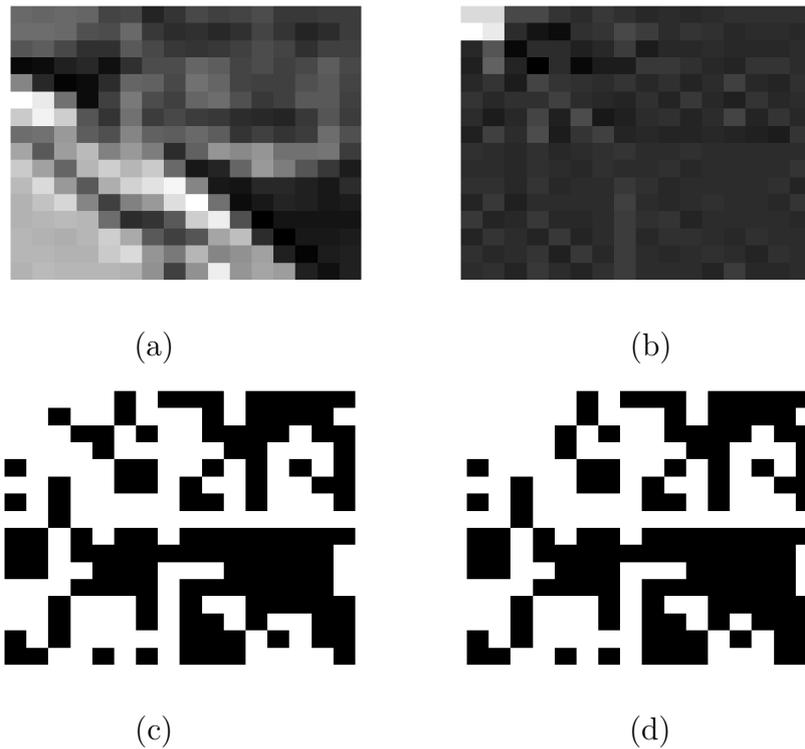


FIGURE 8.2: Exemples, pour un bloc donné (a) transformée par une ondelette avec 3 niveaux de décomposition (b), du premier et second masques (c et d respectivement) obtenus pour l'opérateur ondelette avec un échantillonnage de 50% (la zone blanche correspond aux valeurs conservées).

Ce choix de masque pour l'ondelette entraîne cependant une difficulté dont nous discuterons dans la section résultats (8.2.1). En effet, si les partitions causales et courantes ne sont pas suffisamment corrélées dans le domaine transformé par ondelette, le masque hérité de la partition causale peut être significativement différent du masque optimal pour la partition courante.

La compétition entre ces deux opérateurs d'acquisition permet d'adapter l'échantillonnage en fonction du contenu des partitions. La DCT est en effet efficace pour les contenus texturés tandis que l'ondelette retranscrit très fidèlement les contours et sera donc particulièrement efficace pour les partitions plus difficiles à coder.

De la même manière, grâce à la méthode proposée et modulo l'utilisation d'une bonne partition causale P'_i , le pourcentage de sous-échantillonnage S_i s'adapte habilement à la difficulté de codage de chaque partition. Les partitions faciles à coder héritent ainsi d'un faible pas d'échantillonnage S_i qui limite le nombre de mesures au minimum nécessaire pour obtenir une bonne qualité de reconstruction. Inversement, les zones plus complexes bénéficient automatiquement d'un nombre plus élevé de mesures.

Par rapport aux opérateurs de reconstruction Ψ_i , nous introduisons de la même manière une compétition entre deux opérateurs. Afin d'envisager des opérateurs supplémentaires, nous encourageons le lecteur à se reporter aux travaux [Mun & Fowler 09] où différentes transformées sont comparées. Les opérateurs que nous avons testés dans ce travail sont les suivants :

Variation Totale la variation totale (TV) est couramment utilisée ([Candes *et al.* 06]) comme opérateur de reconstruction dans le cadre de l'échantillonnage compressé appliqué aux signaux 2D tels que les images. La variation totale est en effet efficace dans la mesure où de nombreuses partitions d'une vidéo ont un gradient parcimonieux. Pour une partition f d'une image, elle s'exprime par :

$$TV(f) = \sum_{i,j} \sqrt{(f_{i+1,j} - f_{i,j})^2 + (f_{i,j+1} - f_{i,j})^2},$$

où $f_{i,j}$ est le pixel à la position (i, j) dans le bloc f .

DCT la DCT 2D est utilisée pour sa simplicité d'intégration dans le schéma proposé et ses performances satisfaisantes.

Finalement, les paramètres propres à l'algorithme de reconstruction, Λ_i , peuvent être inclus dans la compétition. Dans les méthodes de reconstruction introduites dans l'état de l'art (6.2.1), il s'agit respectivement des paramètres γ et β pour la méthode ADM [Yang & Zhang 09] et du paramètre λ pour la méthode PL [Haupt & Nowak 06]. Selon le type d'algorithme de reconstruction, un nombre plus élevé de paramètres peut intervenir. Dans la plupart des cas d'utilisation, ils sont classiquement fixés de façon empirique mais conservent une influence non-négligeable sur l'efficacité de reconstruction. Néanmoins, il faut jouer avec finesse avec ces paramètres car ils peuvent aussi avoir un effet sur le temps de convergence des algorithmes.

Nous finissons ainsi la liste non-exhaustive des paramètres intervenant dans les méthodes d'échantillonnage compressé pour les quatre types de paramètres considérés. Nous nous sommes limités à ces paramètres au cours de notre étude mais d'autres choix pertinents peuvent être mis en compétition, notamment pour les opérateurs d'acquisition et de reconstruction.

Dans la section suivante, nous détaillons le fonctionnement de la méthode au travers de la sélection de la partition causale, la détermination des paramètres optimaux et leur utilisation pour le codage de la partition courante.

8.1.3 Détails de la méthode

8.1.3.1 Sélection de la partition causale

La sélection de la partition causale constitue toujours l'étape clé de notre méthode : une partition causale suffisamment corrélée à la partition courante permettra de déterminer des paramètres pertinents tandis qu'une « mauvaise » partition P'_i engendrera un échantillonnage trop coûteux ou bien une mauvaise reconstruction, voir les deux dans le pire des cas. Les méthodes utilisées sont les mêmes que celles introduites dans le chapitre précédent (7.1.3), notamment pour les solutions sans transmission de vecteur de mouvement dont nous expliciterons la mise en œuvre dans la section suivante. En addition de ces solutions, nous détaillons une alternative à l'algorithme de mise en correspondance de blocs (BMA) qui est adaptée pour l'application à l'échantillonnage compressé.

Notre proposition consiste à utiliser un BMA réalisé dans le domaine de l'opérateur d'acquisition Φ . Ainsi, nous cherchons à déterminer la partition P'_i la plus corrélée à la partition courante P_i dans le domaine qui sera utilisé pour l'acquisition des mesures y_m . En reprenant l'équation 7.12 où nous prenons en compte le coût de l'information de mouvement générée pour réaliser l'estimation de mouvement, la méthode recherche alors la partition minimisant le critère :

$$J'_{vm} = \hat{D} + \lambda_{vm} R_{vm}, \quad (8.8)$$

où \hat{D} est ici une mesure de la distorsion prise dans le domaine de l'opérateur d'acquisition Φ . Soit, en utilisant la SAD comme métrique :

$$\hat{D} = \sum_{x,y} |\Phi(P'_i)(x, y) - \Phi(P_i)(x, y)|. \quad (8.9)$$

où (x, y) représentent les coordonnées des partitions transformées. Notons que cette métrique s'apparente à la SATD (*Sum of Absolute Transformed Differences*) qui est habituellement calculée avec la transformée d'Hadamard.

Par rapport à un unique BMA réalisé dans le domaine pixel, a solution que nous proposons ici implique une augmentation de la complexité de l'algorithme dans la mesure où il faut déterminer autant de vecteurs mouvement Vm_i^j que d'opérateurs d'acquisition Φ_j

en compétition. Notons de plus que l'opérateur Φ_j n'est pas nécessairement l'opérateur optimal pour la partition causale $P'_{i,j}$ associée à l'opérateur Φ_j : une partition causale déterminée par BMA réalisé dans le domaine transformé par DCT, peut avoir une ondelette pour opérateur d'acquisition optimal si l'ondelette permet de bien reconstruire ladite partition avec moins de mesures que le nécessiterait l'opérateur DCT. En effet, en se plaçant au niveau du décodeur l'information de mouvement lue permet de pointer vers une partition causale sans savoir si elle a été déterminée lors de l'encodage par un BMA réalisé dans l'un ou l'autre des domaines transformés, il est alors nécessaire de réaliser la compétition pour l'ensemble des opérateurs. Un exemple de cette solution est présentée dans la figure 8.3 où une partition causale $P'_{i,1}$ associée au domaine DCT et une partition $P'_{i,2}$ associée au domaine ondelette sont sélectionnées pour être utilisées par la méthode.

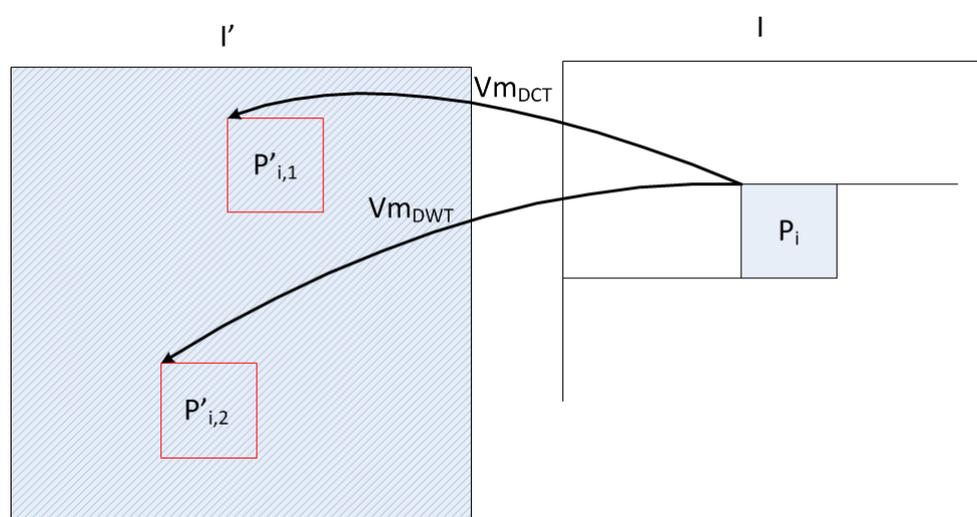


FIGURE 8.3: Exemple d'estimation dans les domaines transformés par DCT et ondelettes.

Pour aller plus loin et être notamment en accord avec l'opérateur d'acquisition Φ_i^S basé sur l'ondelette introduit dans la section précédente, nous proposons de déterminer autant de partitions causales que de pas de sous-échantillonnage S_i en compétition. Le masque pour cet opérateur étant très dépendant du contenu, il est particulièrement important que la partition causale et la partition courante soient corrélées au sens de Φ_i^S . A l'inverse, cela peut entraîner l'héritage d'un masque optimal pour la partition causale mais qui ne le serait pas pour la partition courante. L'acquisition de la partition courante serait alors affectée dans le sens où les mesures les plus significatives ne seront pas conservées tandis que la sélection des paramètres sur la partition causale n'aura pas tenu compte de cette inexactitude.

Dans la sous-section suivante, nous détaillons comment la sélection des paramètres optimaux est réalisée dans ce dernier cas de figure ainsi que dans le cas général.

8.1.3.2 Sélection des paramètres optimaux

La sélection des paramètres optimaux est réalisée sur la partition causale P'_i et doit nécessairement être reproductible au niveau du décodeur afin de retrouver exactement la même liste de paramètres optimaux L_i que lors du codage. Pour l'ensemble des paramètres mis en compétition il s'agit donc de résoudre le problème d'optimisation (\mathcal{P}_2) (8.3) :

$$(\mathcal{P}_2) \begin{cases} \text{Min} [\|\Psi_i P'_i\|_1 + \lambda \|\sum_m y_m - \Phi_i P'_i\|_2 + \nu(\varsigma_y(y_m) + \varsigma_{V_m}(V m_i))] \\ L_i \in \mathcal{F}_\Phi \times \mathcal{F}_\Psi \times [0, 100]_{\mathbb{N}} \times \mathbb{R}^k \end{cases} .$$

Nous cherchons ainsi l'ensemble des paramètres optimaux L_i^* qui minimise le critère pour une valeur ν donnée. Notons que c'est ce multiplicateur ν qui régule le poids aloué au coût de transmission des mesures et de l'information de mouvement éventuel par rapport à la qualité de reconstruction de la partition. Ainsi, une valeur élevée de ν produira un encodage à débit faible associée à une reconstruction de qualité limitée tandis qu'une très bonne reconstruction combinée à un débit plus élevé sera obtenue avec un ν plus faible.

Comme nous le verrons en détails dans la sous-section suivante, la méthode proposée entraîne une discussion par rapport à la complexité. Afin de limiter cette complexité nous avons cependant étudié l'impact du nombre d'itérations réalisées lors des différentes reconstructions par échantillonnage compressé évaluées au crous de l'étape de compétition.

Notons nb'_{iter} et nb_{iter} le nombre d'itérations de l'algorithme de reconstruction respectivement lors de la compétition sur P'_i et lors de l'application à la partition courante P_i . Nous avons empiriquement vérifié qu'une solution hybride avec un nombre d'itérations $nb'_{iter} < nb_{iter}$ peut être utilisée efficacement dans la mesure où le même nombre d'itérations nb'_{iter} est utilisé pour l'ensemble de la compétition. La figure 8.4 présente la courbe d'évolution des gains pour un nombre d'itérations $nb'_{iter} = nb_{iter}$ ainsi que des solutions hybrides avec $nb_{iter} = 100$ et $nb'_{iter} = \{10, 50, 75\}$. Nous avons retenu la solution $nb'_{iter} = 50$, $nb_{iter} = 100$ dans la mesure où elle représente un bon compromis entre complexité et performance de la méthode. Ces résultats ont été obtenus avec l'implémentation HEVC dans un cadre non-décodable (Test 1, sous-section 8.2.2) pour la séquence RaceHorses qwvga et sont confirmés pour d'autres séquences.

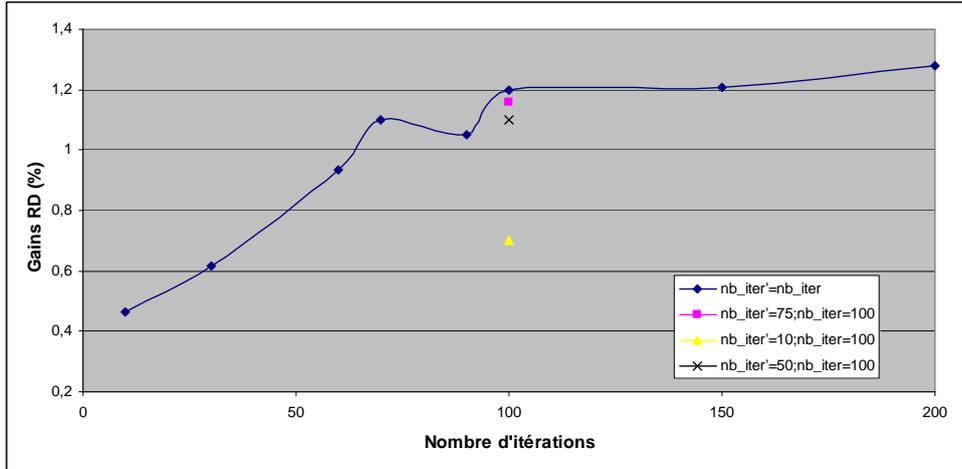


FIGURE 8.4: Evolution du gain débit-distorsion (%) de la méthode comparée à HEVC selon le nombre d'itérations pour un nombre d'itérations fixé (courbe bleue) et trois solutions hybrides. Tests avec l'implémentation HEVC pour la séquence RaceHorses qwvga.

Enfin, traitons le cas d'une mise en œuvre de la méthode où il y a plusieurs partitions causales $P'_{i,j}$, le problème à résoudre est défini comme suit :

$$(\mathcal{P}_3) \begin{cases} \text{Min} \left[\left\| \Psi_i P'_{i,j} \right\|_1 + \lambda \left\| \sum_m y_m - \Phi_i P'_{i,j} \right\|_2 + \nu (\varsigma_y(y_m) + \varsigma_{Vm}(Vm_{i,j})) \right] \\ L_i \in \mathcal{F}_\Phi \times \mathcal{F}_\Psi \times [0, 100]_{\mathbb{N}} \times \mathbb{R}^k \\ P'_{i,j} \in L_{P'_i} \end{cases}, \quad (8.10)$$

où $L_{P'_i}$ est la liste des M partitions qui correspondent aux :

- M partitions causales obtenues par estimation de mouvement dans les domaines transformés correspondants aux opérateurs Φ_i en compétition.
- M partitions causales prédéterminées comme cela est défini à la fin de la sous-section précédente (8.1.3.1).

Un couple $(L_i^*, P'_{i,j}^*)$ optimal est ainsi obtenu et il est nécessaire de transmettre l'information de mouvement correspondante $Vm_{i,j}$ dans le premier cas tandis que le second cas est exactement reproductible au décodeur sans signalisation additionnelle.

Dans les deux cas, cette solution avec partitions causales multiples augmente la complexité de la méthode mais permet d'hériter les meilleurs paramètres pour la partition courante.

Pour conclure avec la description de la méthode, nous détaillons, dans la sous-section suivante, l'application des paramètres optimaux sur la partition courante et discutons quelques points relatifs au schéma proposé.

8.1.3.3 Codage et reconstruction de la partition courante

Pour cette dernière sous-section, nous présentons le schéma complet de la méthode dans la figure 8.5. Pour le codage de la partition courante P_i nous avons précédemment détaillé les étapes de sélection de la partition causale P'_i puis des paramètres optimaux L_i^* . La partition courante P_i est alors codée avec les paramètres optimaux ainsi déterminés. L'acquisition des mesures de la partition P_i est réalisée à l'aide de l'opérateur optimal $\Phi_i^{S_i^*}$ défini à partir de Φ_i^* et du pas de sous-échantillonnage optimal S_i^* . La partition peut alors être reconstruite en utilisant de plus l'opérateur de reconstruction optimal Ψ_i^* et les paramètres Λ_i^* propres à l'algorithme utilisé pour résoudre :

$$\hat{P}_i = \underset{P_i}{\operatorname{arg\,min}} \left\{ \|\Psi_i^* P_i\|_{l_1} + \lambda \|y - \Phi_i^S P_i\|_{l_2} \right\}. \quad (8.11)$$

Finalement, les partitions ainsi reconstruites sont disponibles pour la répétition du processus pour les partitions suivantes.

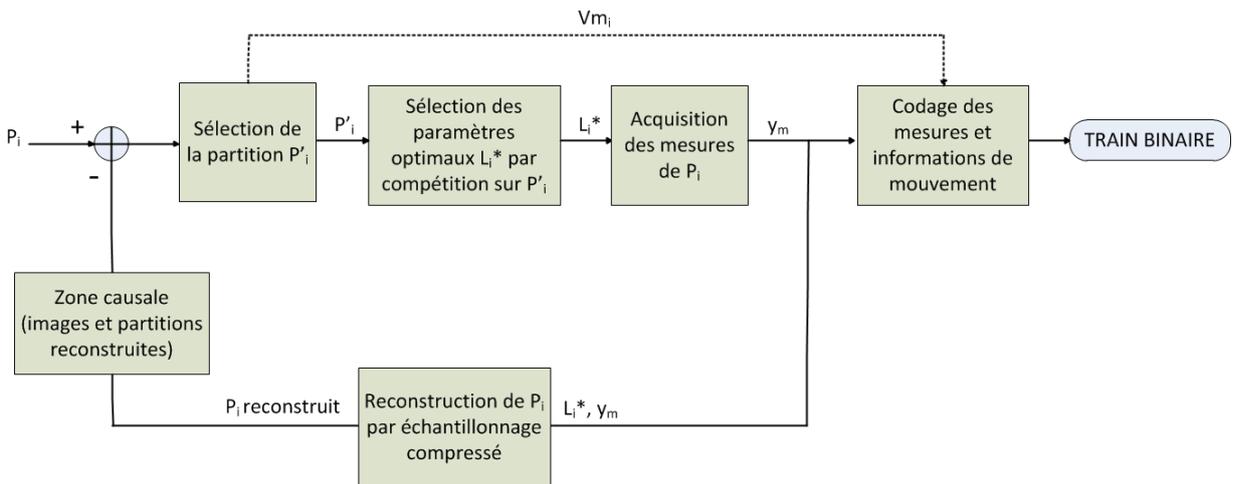


FIGURE 8.5: Schéma détaillé de la méthode d'échantillonnage compressé combinée avec l'approche décodeur intelligent.

Insistons ici sur le fait que la sélection des paramètres optimaux sur une partition causale P'_i entraîne une maîtrise limitée de la reconstruction et du coût effectif résultant de l'application des paramètres L_i^* sur la partition courante. Une corrélation trop faible entre P_i et P'_i provoque ainsi une mauvaise reconstruction de la partition courante ou un sous-échantillonnage pas assez fort de la partition originale (un nombre de mesures conservées plus élevé que nécessaire).

Par rapport à la transmission des informations de codage : les mesures y_m et les informations de mouvement éventuelles Vm_i sont écrites classiquement dans le train binaire à l'aide d'un codeur entropique. Afin de rendre un tel schéma compétitif avec les schémas de codage classique, il est nécessaire d'ajouter une étape de quantification. La quantification des mesures issues de l'échantillonnage compressé est un sujet délicat que nous n'avons pas directement traité dans nos travaux mais il est possible de se reporter notamment à l'étude réalisée dans [Dai *et al.* 09b] pour plus d'informations. Pour le besoin de nos expérimentations, nous avons utilisé une quantification simple au cours de l'application théorique réalisée en Matlab (8.2.1), elle est définie par :

$$Q_T(x) = \begin{cases} 0 & \text{si } |x| < T \\ \text{signe}(x) \times (\lfloor \frac{x}{T} \rfloor + 0.5) \times T & \text{sinon} \end{cases} \quad (8.12)$$

La quantification originale du codeur HEVC est quant à elle utilisée dans l'expérimentation présentée dans 8.2.2.

Pour conclure cette description du schéma de codage par échantillonnage compressé couplé à l'approche décodeur intelligent, revenons sur la notion de complexité. De la même manière que pour l'application de l'approche décodeur intelligent aux modes Intra, la notion de scalabilité en complexité vient s'ajouter aux avantages de la méthode. Ainsi, l'augmentation du nombre de paramètres en compétition ne peut plus avoir d'influence négative sur le débit dans le sens où un plus grand choix de paramètres n'entraîne pas de surcoût de signalisation. Au contraire, ce plus grand nombre de paramètres en compétition améliore la reconstruction du signal et permet ici de réduire le coût des mesures transmises. En revanche, l'influence de cette compétition est directement repercutée sur la complexité du codage et du décodage. Il est donc crucial de limiter les paramètres en compétition aux plus influents dans le cadre d'une application à complexité limitée. A cette notion de scalabilité en complexité liée à la compétition s'ajoute une seconde scalabilité en complexité liée à la méthode d'échantillonnage compressé elle-même. L'algorithme de reconstruction étant itératif, la complexité de la méthode est régulée par le nombre d'itérations. Nous avons fait intervenir ce point dans la section 8.1.3.2 où l'efficacité de la méthode est conservée lorsque le nombre d'itérations au cours de la compétition est réduit par rapport au nombre d'itérations utilisé lors de l'application de la méthode sur la partition courante P_i . Sur la figure 8.4 nous constatons aussi l'impact du nombre d'itérations sur l'efficacité de la méthode. Celle-ci se stabilise pour un nombre d'itérations $nb_{iter} > 100$. En fonction du contenu de la séquence traitée et du cadre de l'application de la méthode, nous proposons donc de signaler ce nombre d'itérations au décodeur afin d'optimiser la qualité de la reconstruction pour une cible de complexité donnée.

Dans la section suivante nous présentons les résultats obtenus avec la méthode : tout d'abord dans un cadre théorique où nous évaluons un schéma de codage vidéo entiè-

rement basé sur la méthode proposée puis dans le cadre du codeur vidéo HEVC où la méthode est introduite comme un nouveau mode de codage en compétition avec les modes de référence.

8.2 Résultats expérimentaux

8.2.1 Expérimentation théorique

Le but de cette expérimentation théorique réalisée avec Matlab est d'évaluer les différentes solutions introduites dans la précédente section pour finalement proposer une solution complète de codage basée sur l'échantillonnage compressé. Cette solution est comparée d'une part à d'autres solutions de l'état de l'art sur l'échantillonnage compressé appliqué à un signal vidéo et d'autre part à un schéma de codage vidéo prédictif classique.

8.2.1.1 Détails de l'implémentation

Nous avons implémenté en Matlab la méthode présentée dans 8.1. Chaque séquence est donc traitée image par image, les images étant regroupées en GOP (groupement d'images consécutives) de taille fixe où chaque image de début de GOP, dite Intra, est codée sans référence à une autre image. Chaque image est partitionnée en blocs de taille 16×16 pixels traités séquentiellement.

Par défaut, les images Intra utilisent un paramétrage prédéterminé de la méthode d'échantillonnage compressé avec la DCT utilisée pour l'opérateur d'acquisition Φ_i , un pas de sous-échantillonnage $S_i = 0.7$, l'utilisation de la minimisation de la variation totale pour la reconstruction Ψ_i et le paramètre de l'algorithme de reconstruction $\nabla = 3$.

Pour les images suivantes, une estimation de mouvement en Full Search est réalisée avec une fenêtre de recherche de ± 15 pixels. Par défaut, cette estimation de mouvement est réalisée dans le domaine pixel mais nous avons aussi évalué l'impact d'une estimation de mouvement réalisée dans le domaine transformé des opérateurs d'acquisition Φ_i en compétition. Dans ce cas, l'étape de calcul du coût dans l'algorithme d'estimation de mouvement est modifiée par le calcul de la SAD entre la partition originale et la partition testée transformées dans le domaine Φ_i . La partition causale P'_i utilisée par notre méthode correspond alors à la partition dans le domaine pixel pointée par le vecteur mouvement issue de l'estimation de mouvement.

La reconstruction de chaque bloc est réalisée à l'aide de l'algorithme de reconstruction PL¹ introduit dans l'état de l'art (6.2.1) pour lequel le nombre d'itérations maximales est fixé à 400.

1. Implémentation Matlab disponible à l'adresse <http://www.ece.msstate.edu/~fowler/BCSSPL/> pour laquelle nous avons supprimé l'étape de lissage par un filtre de Wiener car chaque partition est reconstruite indépendamment dans notre méthode.

Enfin, l'évaluation de la méthode dans un cas de codage des mesures utilise un module de quantification comme défini par la formule 8.12 et un codeur arithmétique classique² qui retourne le nombre de bits utilisés pour le codage des mesures de chaque partition.

Notre solution, notée DI-CS (pour *décodeur intelligent compressed sensing*), est comparée dans cette section aux méthodes de l'état de l'art suivantes :

BCS-PL pour *block-based compressed sensing with projected Landweber*, il s'agit d'une méthode classique basée sur l'échantillonnage compressé : chaque image est traitée indépendamment avec un opérateur d'acquisition Φ fixé qui conserve les $S\%$ des coefficients placés en premiers selon le parcours zig-zag dans chaque bloc transformé par DCT, la variation totale est utilisée pour l'opérateur de reconstruction Ψ . Notons que cette méthode correspond à l'environnement de base dans lequel notre méthode DI-CS a été bâtie, on peut voir cette méthode comme notre solution avec une liste de paramètres optimaux L_i^* figée pour l'ensemble de la séquence.

MC-BCS-SPL pour *motion-compensated block-based compressed sensing with smooth projected Landweber* [Mun & Fowler 11], il s'agit d'une méthode basée sur l'échantillonnage compressé qui exploite les redondances temporelles en raffinant la reconstruction de chaque bloc par l'application successive d'une méthode d'échantillonnage sur le résiduel obtenu par compensation de mouvement avec le bloc reconstruit à l'itération précédente. L'opérateur d'acquisition Φ utilisé est une matrice aléatoire orthonormée avec un pourcentage d'échantillonnage S fixé à 0.1 et la DCT comme opérateur de reconstruction Ψ .

MC-BPC pour *motion-compensated block-based predictive coding*, il s'agit d'un schéma de codage vidéo classique sans échantillonnage compressé : après estimation de mouvement, la partition obtenue par compensation de mouvement est utilisée pour prédire la partition courante, le résiduel obtenu est transformé à l'aide d'une DCT, quantifié puis codé à l'aide du codeur arithmétique. Pour les images codées en Intra, une prédiction DC est utilisée pour générer le résiduel.

8.2.1.2 Résultats

8.2.1.2.1 Evaluation des différents paramètres en compétition

Nous évaluons tout d'abord en détail la compétition du pas de sous-échantillonnage S_i puis des opérateurs d'acquisition Φ_i . Les performances de la méthode complète seront ensuite étudiées face aux différentes méthodes de l'état de l'art introduites ci-dessus.

Pour le paramètre de sous-échantillonnage S_i , nous le faisons varier de 0.1 à 0.7 avec un pas de 0.05 pour l'opérateur d'acquisition Φ_i fixé à la DCT. Les valeurs extrêmes testées sont définies ainsi d'une part car un sous-échantillonnage plus faible que 10% conserve trop peu de mesures et d'autre part car un sous-échantillonnage plus élevé que 70% n'est

2. Fonctions Matlab correspondantes disponibles à l'adresse
<http://www.ceremade.dauphine.fr/~peyre/teaching/wavelets/matlab/toolbox>.

pas significatif lorsque l'on considère l'ajout d'un bruit de quantification comme nous le verrons à la fin de cette sous-section. Trois valeurs $\{1, 2, 5\}$ du paramètre ν intervenant dans l'équation 8.3 ont été évalués ici, elles permettent de viser respectivement des haut, moyen et bas débits. Nous comparons ces tests à la courbe générée par la méthode BCS-PL avec les sous-échantillonnages fixes pour chaque expérimentation. Ces résultats sont reportés sur la figure 8.6.

La méthode proposée permet systématiquement d'obtenir une qualité de reconstruction équivalente pour un nombre de mesures significativement inférieur.

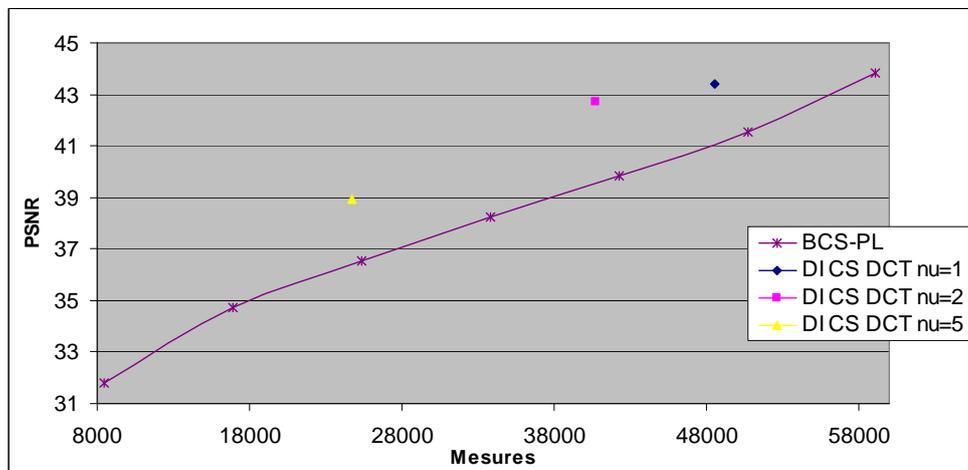


FIGURE 8.6: Variation du paramètre de sous-échantillonnage.

Nous évaluons maintenant la mise en compétition des opérateurs d'acquisition Φ_i . Pour un pas de sous-échantillonnage suffisamment représentatif fixé à $S_i = 0.5$ (des résultats similaires sont obtenus pour d'autres échantillonnages), les opérateurs DCT et ondelette sont mis en compétition pour une valeur du paramètre intervenant dans le problème 8.3, $\nu = 5$. Nous présentons les résultats dans le tableau 8.1 en évaluant les différentes solutions proposées pour l'opérateur ondelette. Ainsi, pour les tests avec la compétition DCT et ondelette, nous évaluons les deux masques proposés dans 8.1.2 puis la réalisation de l'estimation de mouvement dans les domaines transformés correspondant (8.1.3.1).

Nous observons que les trois tests ajoutant l'opérateur ondelette améliorent l'efficacité de la méthode par rapport au test avec la DCT uniquement : des faibles diminutions du nombre de mesures (jusqu'à 1% de mesures en moins) sont associées à une amélioration significative de la qualité (gains de 0.4dB à 0.7dB). La combinaison des deux opérateurs permet en effet d'adapter l'acquisition selon le contenu comme nous le verrons avec l'étude de la figure 8.8. La comparaison des deux masques tend à confirmer que le second masque, qui conserve systématiquement les coefficients de la dernière sous-bande,

est le plus efficace pour améliorer la qualité de la reconstruction. Ceci est d'autant plus vrai que le masque hérité de la partition causale est trop différent du masque optimal pour la partition courante. Le test présenté dans la dernière colonne apporte une solution à ce problème en réalisant l'estimation de mouvement dans le domaine d'acquisition. L'héritage du masque, notamment pour les partitions codées avec l'ondelette, est alors plus efficace dans la mesure où les partitions courantes et causales sont davantage corrélées dans le domaine d'acquisition.

	DCT	DCT/DWT masque 1	DCT/DWT masque 2	DCT/DWT masque 2 + EM*
PSNR	38.93	39.28	39.59	39.60
Nombre de mesures	24742	24563	24640	24512

TABLE 8.1: Evaluation de la mise en compétition des opérateurs d'acquisition DCT et Ondelette pour nos deux propositions de masque ainsi que l'estimation de mouvement réalisée dans le domaine transformé (EM*). Tests réalisés avec $\nu = 5$.

Des résultats similaires sont obtenus pour d'autres valeurs de ν ainsi que pour la compétition pour la base de reconstruction Ψ_i et les paramètres de la méthode de reconstruction Λ_i qui améliorent légèrement l'efficacité de la méthode.

Suite à cette étude préliminaire, nous analysons les performances de la méthode dans deux évaluations présentées dans les paragraphes suivants. La configuration utilisée pour ces tests est détaillée ci-dessous :

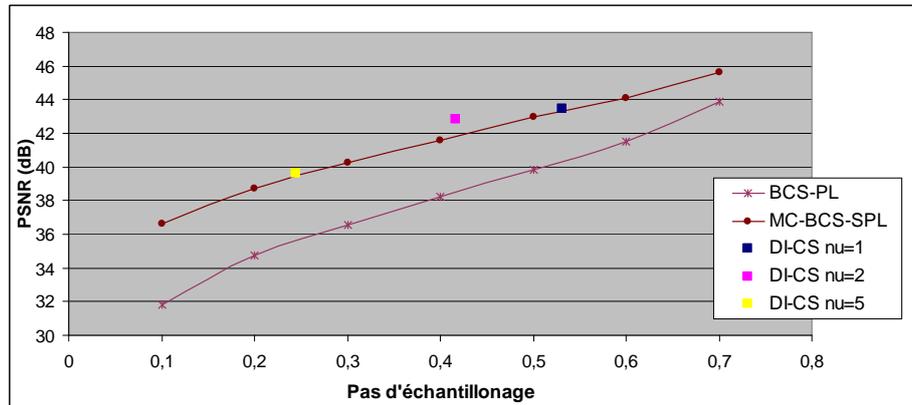
Codeur Codeur vidéo simulé à l'aide de Matlab.
Profil IPPP.
Partition Taille 16×16 .
BMA <i>Full Search</i> avec une fenêtre de recherche de taille 32 pixels.
Echantillonnage Compressé Algorithme de reconstruction PL [Haupt & Nowak 06].
DI-CS $\Phi_i = \{DCT, DWT\}$, $S_i = \{0.1, 0.15, \dots, 0.7\}$, $\Psi_i = \{TV, DCT\}$, $\Lambda_i = \{1, 5, 10\}$.
Séquences Trois séquences Cif avec la composante de Luma uniquement.
Evaluation PSNR en fonction du pas d'échantillonnage et PSNR en fonction du débit.

8.2.1.2.2 Evaluation du PSNR en fonction du pas d'échantillonnage

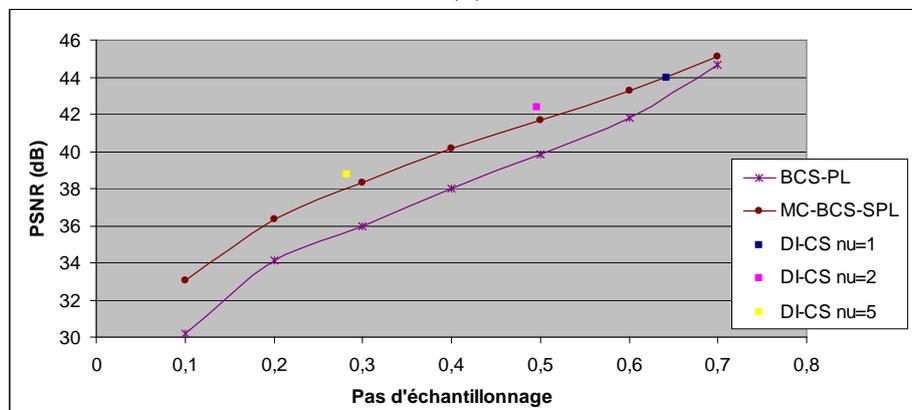
Pour la première évaluation (figure 8.7) nous restons dans le strict cadre de l'échantillonnage compressé et comparons, en termes de qualité selon le pas d'échantillonnage globale de la séquence S , différentes versions de notre méthode aux schémas de l'état de l'art BCS-PL et MC-BCS-SPL. L'étape de quantification n'est donc pas prise en

compte ici et c'est pourquoi nous ne nous comparons pas à un schéma de codage vidéo classique dans la mesure où ce schéma est sans perte lorsque l'on supprime l'étape de quantification. Cette expérimentation sans quantification conserve néanmoins une réelle signification car la quantification des mesures demeure une question très ouverte de l'échantillonnage compressé ce qui explique que la communauté préfère communément évaluer les méthodes en terme de qualité par rapport au nombre de mesures plutôt que par rapport au débit. Nous présenterons cependant de tels résultats après cette première évaluation dans le paragraphe 8.2.1.2.3. Enfin, notons que pour cette évaluation, chaque composante des vecteurs mouvement indiquant les partitions causales dans notre méthode DI-CS est comptabilisée comme une mesure supplémentaire pour chaque bloc et nous déduisons alors le pas d'échantillonnage moyen S de la séquence à partir du nombre total de mesures par rapport au nombre de pixels de la séquence.

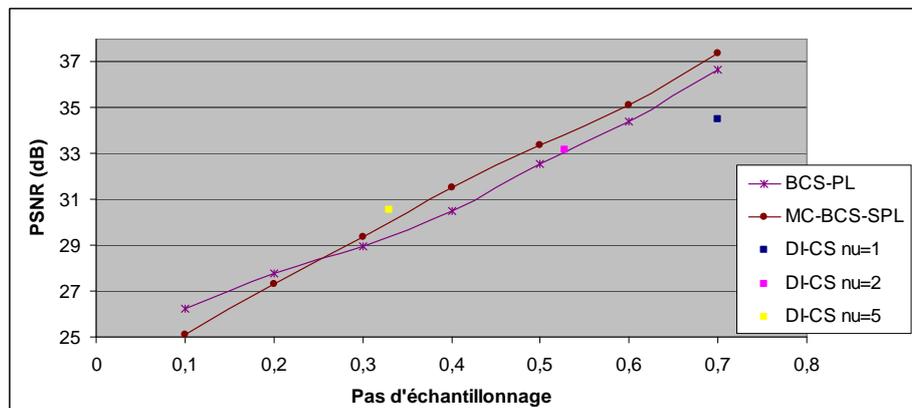
Trois versions de notre méthode DI-CS sont évaluées pour différentes valeurs du paramètre $\nu = \{1, 2, 5\}$ qui permettent de viser respectivement des haut, moyen et bas débits. Notre méthode permet d'obtenir une reconstruction de bonne qualité avec un nombre de mesures limité pour les séquences *Susie* (a) et *Foreman* (b), où les PSNRs sont supérieurs à 42dB pour les configurations à haut et moyen débits et entre 38dB et 40dB pour les points à bas débit des mêmes séquences. La séquence *Soccer* (c) étant plus complexe, la qualité de la reconstruction est moindre et varie de 30dB à 35dB. Par rapport aux deux méthodes de l'état de l'art, notons tout d'abord que la méthode MC-BCS-SPL qui exploite les redondances temporelles est plus efficace que la méthode plus classique BCS-PL pour l'ensemble des tests réalisés excepté pour la séquence *Soccer* (c) à très bas débit ($S \leq 0.2$) où l'estimation de mouvement réalisée sur le bloc reconstruit dans MC-BCS-SPL est rendue délicate du fait de la reconstruction de qualité très limitée. Ceci confirme l'intérêt de l'exploitation de la dimension temporelle des vidéos. De façon similaire, notre solution DI-CS est significativement meilleure que la méthode classique BCS-PL pour l'ensemble des tests réalisés excepté pour la séquence *Soccer* (c) à haut débit ($\nu = 1$) où la qualité de la reconstruction pour un même échantillonnage est inférieure de 2dB. Ceci s'explique par le fait que les composantes liées à la distorsion et au nombre de mesures ont le même poids dans l'équation 8.3 de notre solution DI-CS, le pas de sous-échantillonnage maximal ($S_i = 0.7$) est alors sélectionné de façon quasi systématique et il faut y ajouter les mesures des vecteurs mouvement ; combiné à l'utilisation très fréquente de l'opérateur ondelette dont l'héritage du masque n'est pas optimal, la reconstruction obtenue est limitée pour un pas d'échantillonnage relativement élevé. Ce comportement à haut débit se retrouve dans une autre mesure pour les deux premières séquences où l'amélioration des performance est moindre que pour les moyen ($\nu = 2$) et bas ($\nu = 5$) débits. Pour ces débits cibles, les gains sont particulièrement importants pour la séquence *Susie* (a) dont le pas d'échantillonnage est divisé par deux pour une même qualité de reconstruction à bas débit et réduit de 0.2 à moyen débit. Cela s'explique par la forte corrélation temporelle entre les images successives de cette séquence qui permet d'habilement hériter les paramètres de codage de chaque bloc.



(a)



(b)



(c)

FIGURE 8.7: Résultat complet de la méthode comparée aux méthodes de l'état de l'art en échantillonnage compressé. Evolution du PSNR en fonction du pas d'échantillonnage pour trois séquences Cif testées : *Susie* (a), *Foreman* (b) et *Soccer* (c).

Enfin, la comparaison avec la méthode MC-BCS-SPL est instructive pour notre étude :

les deux méthodes, qui possèdent une complexité voisine et exploitent chacune les redondances temporelles de façon différente, ont globalement une efficacité comparable et notre solution est parfois meilleure.

Ainsi, pour les deux premières séquences aux fortes corrélations temporelles, la méthode DI-CS permet un gain significatif à moyen débit s'élevant jusqu'à 1dB pour un même pas d'échantillonnage dans le cas de la séquence *Susie* (a) tandis qu'une même amélioration est aussi obtenue pour la séquence *Foreman* (b) à bas débit. Ces résultats sont très encourageants et confirment l'intérêt de notre approche qui, sans utiliser directement les pixels pour la reconstruction, permet de judicieusement réaliser l'acquisition et la reconstruction de chaque bloc à l'aide des paramètres déterminés sur les bloc caux. En revanche, les performances pour une séquence plus complexe telle que *Soccer* (c) sont moindres à haut et moyen débits alors qu'elles demeurent bonnes à bas débit. Nous avons précédemment expliqué les causes de ces résultats pour la méthode DI-CS qui viennent d'un héritage difficile des paramètres et n'affectent bien sûr par la méthode de l'état de l'art MC-BCS-SPL dont une reconstruction de qualité minimale correcte est assurée dans ce cas de figure.

Pour clore cette première évaluation, nous étudions en détail la répartition de la sélection des différents paramètres dans le cas de la séquence *Susie*. La figure 8.8 présente la version originale et le résultat du codage avec notre méthode de la seconde image de cette séquence ainsi que la répartition des choix optimaux des opérateurs d'acquisition et du pas de sous-échantillonnage.

Notons tout d'abord que l'image est bien reconstruite (b), ici avec uniquement 24 640 mesures ($S = 0.25$) pour un PSNR de 39,59dB et très peu de différences visibles. Par rapport au choix des opérateurs optimaux (c), les zones blanches utilisent l'ondelette tandis que les blocs noirs sont codés avec la DCT. L'ondelette est donc majoritairement choisie sur cet exemple, en particulier pour les zones incluant un contour. A l'inverse, les zones plus lisses comme le fond en haut à droite de l'image ou les zones du visage et du cou de la personne utilisent la DCT. Par rapport au pas de sous-échantillonnage S_i (d), les blocs noirs correspondent à un fort sous-échantillonnage tandis que les blocs blancs correspondent à un faible sous-échantillonnage (70% des mesures conservées).

On voit ainsi que l'échantillonnage déterminé sur l'image précédente s'adapte très bien au contenu puisqu'un faible sous-échantillonnage est utilisé sur les zones faciles à coder comme le fond ou certaines parties du visage alors que les autres zones plus difficiles nécessitent davantage de mesures, en particulier au niveau des cheveux.

De plus, il est intéressant de noter que les zones codées avec la DCT ont un pas de sous-échantillonnage souvent faible, en particulier au niveau du fond à droite ainsi

que du visage et du cou. Ceci s'explique par le fait que la DCT réussit à reconstruire efficacement ces zones avec un nombre de mesures très limité.

Finalement insistons sur le fait qu'en fonction de la valeur du paramètre ν , la répartition des choix évolue : pour une valeur élevée de ν il y aura plus de blocs codés avec la DCT et S_i sera plus faible dans la mesure où les composantes liées aux coût de codage dans le problème 8.3 ont plus d'importance, inversement pour une valeur faible de ν .

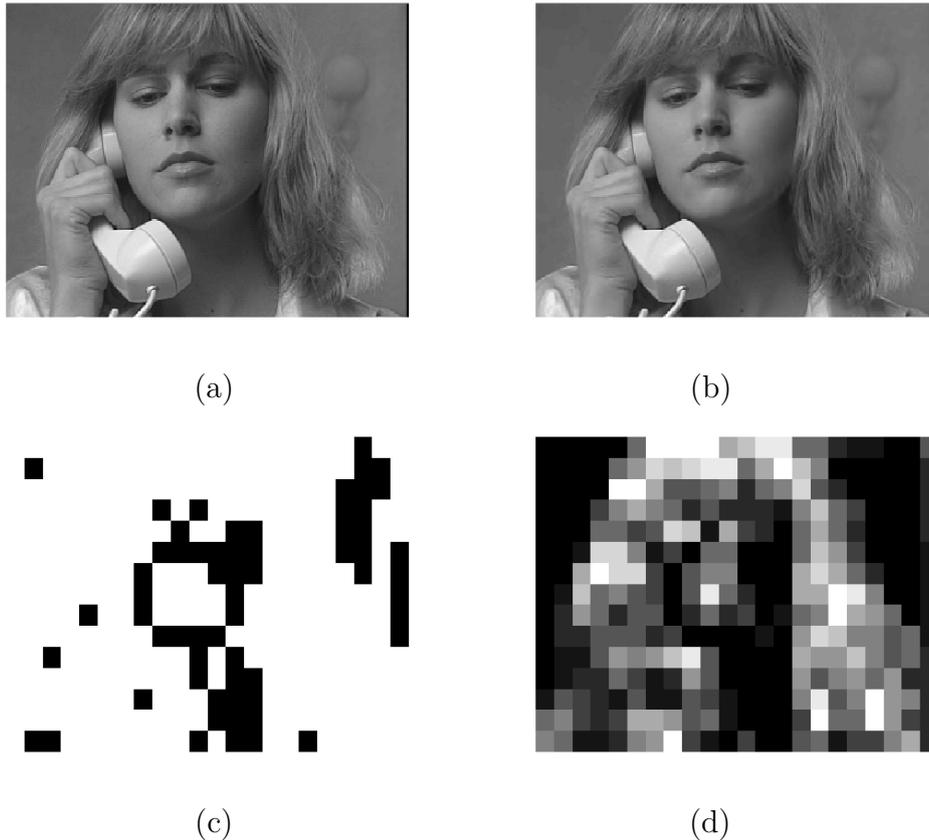


FIGURE 8.8: Seconde image de la séquence *Susie cif* originale (a) et encodée avec la méthode (b), répartition des opérateurs Φ_i optimaux DCT (en noir) et ondelette (en blanc) (c) et des échantillonnages optimaux S_i des plus forts échantillonnages (en noir) aux plus faibles (en blanc) (d).

8.2.1.2.3 Evaluation du PSNR en fonction du débit

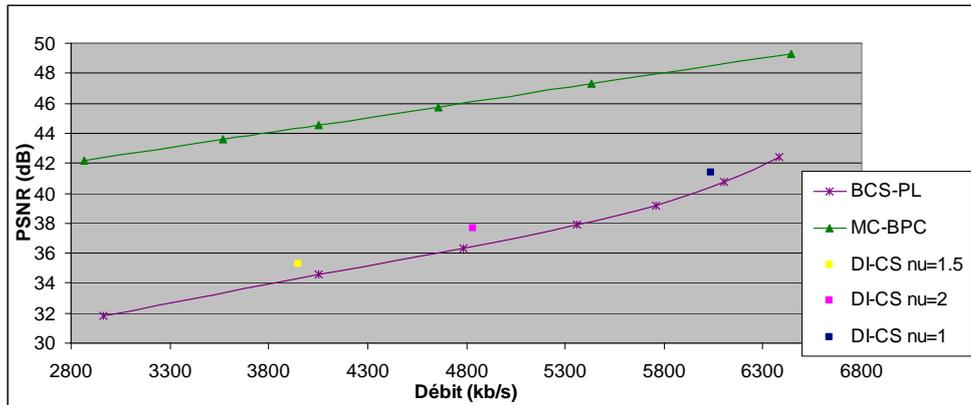
Dans cette seconde évaluation, notre but est de sonder le comportement de la méthode proposée (DI-CS) dans un cadre de codage vidéo traditionnel (qualité en fonction du débit) et de se comparer à un schéma de compression vidéo classique basé sur une prédiction temporelle (MC-BPC). L'évaluation inclut par conséquent une étape de quantification des mesures avant codage comme définie par la formule 8.12. Nous avons fixé le

paramètre de quantification $T = 3$, ce qui permet d'avoir une quantification significative mais qui conserve suffisamment de mesures nécessaires pour obtenir une reconstruction correcte. Comme précédemment, nous obtenons différents niveaux de débits en faisant varier le paramètre ν du problème 8.3 ($\nu = \{1, 1.5, 2\}$). La même quantification est appliquée aux mesures résultantes de la méthode d'échantillonnage compressé BCS-PL pour laquelle nous avons évalué les performances pour des pas d'échantillonnage variant de 0.1 à 0.7. Enfin, dans le cadre du schéma de codage vidéo classique MC-BPC, plusieurs niveaux de quantification variant de $T = 4$ (faible quantification) à $T = 16$ (forte quantification) sont appliqués aux résiduels de texture avant transmission.

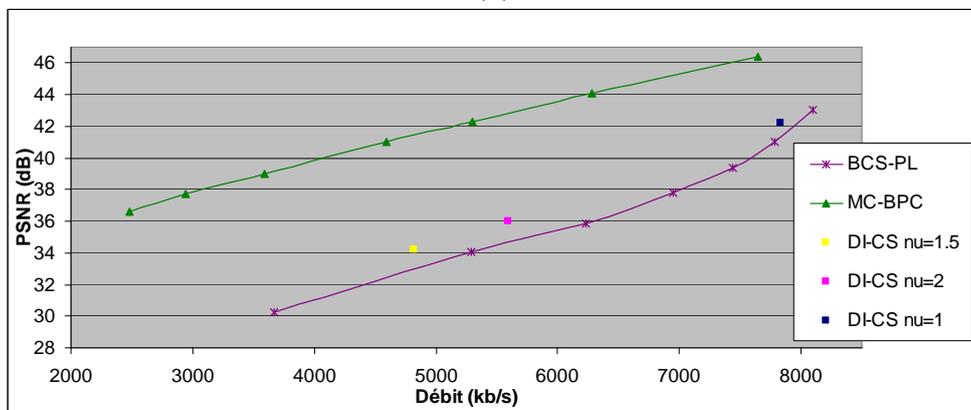
La figure 8.9 présente, pour les mêmes séquences que précédemment, les résultats obtenus par notre méthode en comparaison avec les méthodes de l'état de l'art. Elles permettent tout d'abord d'avoir un aperçu des performances de l'échantillonnage compressé face à un schéma de codage vidéo classique. La méthode MC-BPC, bien qu'étant elle-même éloignée des performances d'encodeurs réels (dans le sens où c'est une version simplifiée d'un codeur vidéo n'incluant pas de compétition de modes ni d'outils tels que la prédiction du mouvement, CABAC,...), est ainsi très significativement meilleure que le schéma BCS-PL avec une augmentation de la qualité allant de 5dB à 10dB à même débit. La structure prédictive de MC-BPC nécessite en effet beaucoup moins de données à transmettre pour assurer une reconstruction de bonne voir très bonne qualité. Notons de plus que c'est pour les hauts débits que les performances de l'échantillonnage compressé sont globalement le moins éloignées de celles de MC-BPC, en particulier pour les deux premières séquences. Ces résultats tendent à confirmer l'étude réalisée dans [Dubreil Vall & Stojanovic 11] qui compare un schéma d'échantillonnage compressé du type BCS-PL au standard H.264/AVC : un PSNR de 31dB y est obtenu pour un ratio de compression de 15% avec la méthode d'échantillonnage compressé, tandis que la même qualité de reconstruction est obtenue avec un ratio de seulement 0.45% pour H.264/AVC.

Par rapport à notre méthode DI-CS, nous observons tout d'abord qu'elle conserve un comportement similaire à celui de l'étude précédente par rapport au schéma d'échantillonnage compressé BCS-PL : la méthode est plus efficace pour les deux premières séquences (*Susie* (a) et *Foreman* (b)) ainsi qu'à bas et moyen débits pour la séquence *Soccer* (c) tandis qu'elle est de nouveau moins efficace pour cette même séquence à haut débit. Les gains par rapport à BCS-PL sont cependant moindres que ceux observés dans la figure 8.7 avec une augmentation moyenne de 1dB à moyen débit alors que DI-CS atteignait 4dB de plus pour ce même type de débit dans l'évaluation sans quantification. L'ajout de cette étape a pour effet d'harmoniser davantage les mesures, ce qui entraîne une perte des détails que notre méthode permet de judicieusement prendre en compte ordinairement. Enfin, nos résultats demeurent significativement éloignés des résultats obtenus avec la méthode MC-BPC.

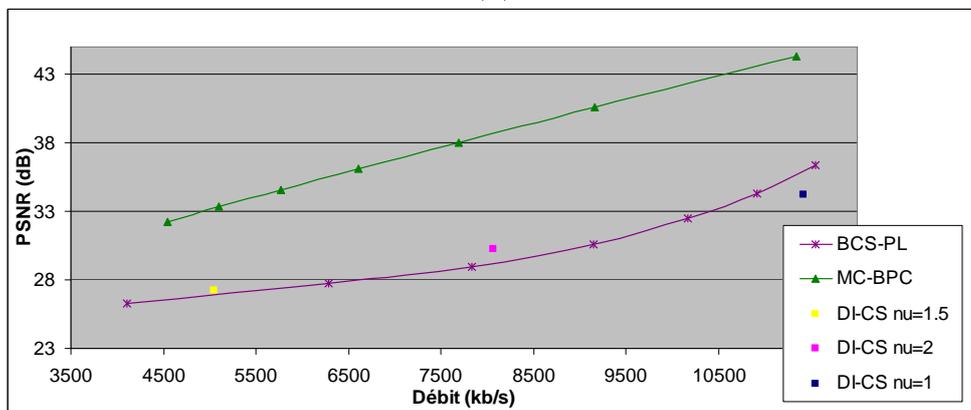
En effet, la question de la quantification des mesures ou plus largement de leur codage, constitue un point bloquant pour des méthodes par échantillonnage compressé n'utilisant aucune prédiction.



(a)



(b)



(c)

FIGURE 8.9: Résultat de la méthode dans un schéma incluant une quantification des mesures. Comparaison à une méthode de l'état de l'art en échantillonnage compressé (BCS-PL) et à une méthode de codage vidéo classique (MC-BPC). Evolution du PSNR en fonction du débit pour trois séquences Cif testées : *Susie* (a), *Foreman* (b) et *Soccer* (c).

Finalement, la figure 8.10 présente la répartition des choix des opérateurs d'acquisition et des paramètres de sous-échantillonnage dans le cas avec quantification. Ceci est à mettre en regard de la figure 8.8 sans quantification des mesures. Notons en particulier que pour l'opérateur d'acquisition Φ_i , la DCT (zones en noir) est beaucoup plus sélectionnée du fait de la quantification qui fait perdre trop de mesures pour l'opérateur ondelette. Ce dernier est uniquement sélectionné sur les zones très difficiles à coder comme les cheveux qui de façon générale utilisent un pas de sous-échantillonnage élevé permettant de conserver davantage de mesures.

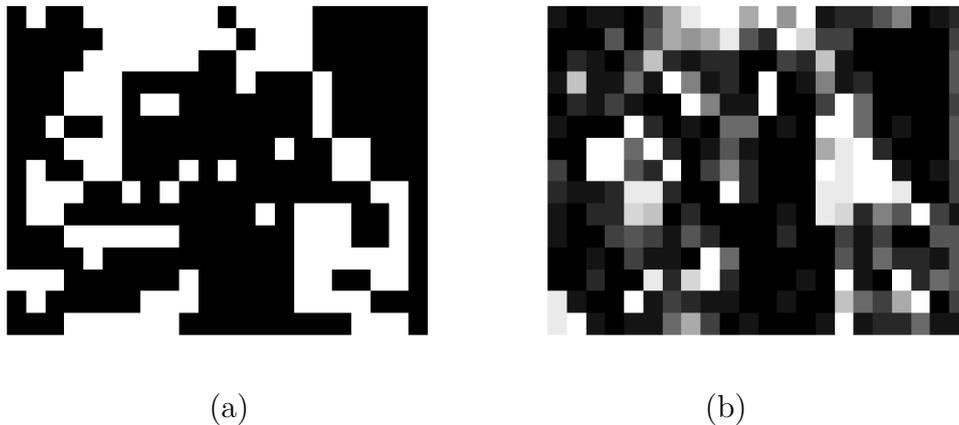


FIGURE 8.10: Répartition des opérateurs Φ_i optimaux avec la DCT en noir et l'ondelette en blanc (a) et des échantillonnages optimaux S_i avec les plus forts échantillonnages en noir et les plus faibles échantillonnages en blanc (b) pour la seconde image de la séquence *Susie cif*.

En conclusion de cette étude, nous avons évalué notre méthode dans un cadre théorique où nous avons comparé les performances à d'autres méthodes d'échantillonnage compressé. L'adaptativité et l'exploitation des redondances temporelles apportées par notre solution améliorent significativement une méthode classique du type BCS-PL et permettent des performances similaires voir meilleures qu'une autre solution exploitant les redondances temporelles (MCS-BCS-PL). Nous avons aussi pu vérifier que le choix des différents paramètres s'adapte aux contraintes de contenu des partitions ainsi qu'au débit ciblé.

Cependant, le schéma reste moins efficace comparé à une méthode de codage vidéo classique, c'est pourquoi nous proposons dans la sous-section suivante d'intégrer notre méthode en compétition avec d'autres modes de codage au sein d'un codeur vidéo de nouvelle génération afin de tirer profit des avantages de l'échantillonnage compressé uniquement pour les partitions au contenu le mieux adapté.

8.2.2 Evaluation dans un codeur vidéo moderne (HEVC)

L'intégration d'un schéma d'échantillonnage compressé dans un codeur vidéo moderne est très délicate comme l'ont montré les résultats limités des travaux [Zhang *et al.* 08, Do *et al.* 10]. En effet, il est difficile d'obtenir des mesures aussi peu coûteuses que le résiduel transformé et quantifié obtenu après une prédiction Intra ou Inter, même avec un échantillonnage optimal. Par conséquent, l'intégration efficace d'une méthode d'échantillonnage compressé dans un codeur moderne ne peut s'effectuer que par application à une donnée de type résiduel contrairement à l'étude précédente (8.2.1) réalisée dans un cadre plus théorique. Nous sommes alors exactement dans le cadre du problème posé dans ce chapitre 8.1.1 avec plusieurs paramètres en compétition nécessaire pour assurer l'efficacité de la méthode. Nous proposons dans cette sous-section d'appliquer la solution proposée ci-dessus au futur standard de compression vidéo HEVC.

Nous détaillons tout d'abord l'implémentation proposée puis nous présentons les résultats obtenus.

8.2.2.1 Détails de l'implémentation

Nous avons fait le choix de concentrer notre étude sur la prédiction Intra. En effet, deux raisons principales rendent délicates l'intégration de notre méthode en prédiction Inter :

1. Les modes Inter étant très efficace en terme de prédiction, une méthode basée sur l'échantillonnage compressé a très peu de sens. En effet, comme cela a été étudié dans le chapitre 2, une majorité des blocs Inter après prédiction, transformation et quantification possèdent un ensemble de coefficients nuls ou bien un nombre très faible de coefficients non-nuls. Réaliser une acquisition par échantillonnage compressé de ce type de bloc est alors inutile et le surcoût engendré par la signalisation de la méthode est trop pénalisant.
2. Dans le cadre décodeur intelligent, le choix d'une partition causale P'_i est délicat du fait de la redondance avec la partition déjà utilisée dans la prédiction Inter pour réaliser la compensation de mouvement. La solution envisagée serait d'utiliser une partition causale appartenant à l'image I'' correspondant à $t - 2$ mais les corrélations avec la partition courante deviennent alors plus faible.

Ainsi, nous introduisons un nouveau mode de codage Intra basé sur l'échantillonnage compressé d'une manière similaire aux travaux présentés dans [Zhang *et al.* 08, Do *et al.* 10]. Ce mode est mis en compétition avec les modes classiques, sélectionné à l'aide d'un critère débit-distorsion puis signalé à l'aide d'un indice transmis dans le flux. Le nouveau mode de codage est détaillé dans la figure 8.11 : après prédiction du bloc, le résiduel est transformé puis quantifié et nous ajoutons une étape de sous-échantillonnage des coefficients obtenus qui sont alors transmis au codeur entropique. La boucle de retour permet de reconstruire le résiduel en réalisant la quantification inverse des coefficients

puis la reconstruction à l'aide de l'échantillonnage compressé. Ce résiduel est finalement ajouté au signal prédicteur.

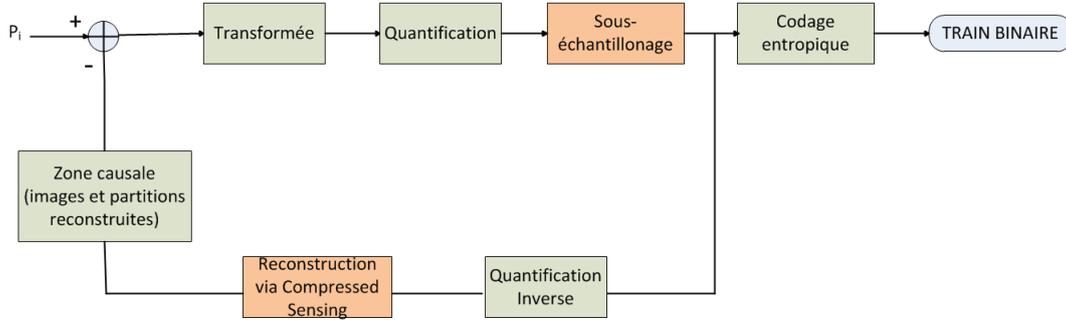


FIGURE 8.11: Mode de codage basé sur l'échantillonnage compressé.

Nous combinons cette mise en œuvre avec l'approche décodeur intelligent permettant de déterminer de façon adaptative les paramètres nécessaires à la méthode sans avoir à les transmettre au décodeur. Pour ce faire la partition causale P'_i est sélectionnée parmi les partitions voisines de la partition courante comme cela est décrit dans la section 8.1.3.1. Ces partitions sont en effet très corrélées à la partition courante. Au niveau du codeur comme du décodeur, l'estimation de la composante de débit associée à chaque simulation du codage de la partition P'_i est obtenue à partir d'une mesure de l'entropie des résiduels à transmettre, tandis que la distorsion est calculée entre la partition reconstruite et la partition causale.

Les paramètres mis en compétition sont le pas de sous-échantillonnage S_i et les paramètres Λ_i propres à la méthode de reconstruction :

$$L_i = \{S_i, \Lambda_i\} . \quad (8.13)$$

L'opérateur d'acquisition Φ_i est fixé à la DCT pour s'intégrer le plus simplement au schéma de codage HEVC et nous utilisons la variation totale pour la reconstruction Ψ_i . Enfin, c'est l'algorithme de reconstruction ADM³ introduit dans la section 6.2.1 qui est utilisé pour cette expérimentation avec un nombre d'itérations fixé à $nb'_{iter} = 50$ pendant la phase de compétition et $nb_{iter} = 100$ lors de l'application de la méthode avec les paramètres optimaux comme expliqué dans 8.1.3.2.

3. Algorithme développé en C++ à partir de la version Matlab disponible à l'adresse http://www.caam.rice.edu/~optimization/L1/2007/09/software_08.html

8.2.2.2 Résultats

La configuration utilisée pour l'ensemble des tests présentés dans cette section est la suivante :

<p>Codeur HEVC TMuC 0.7.</p> <p>Profil Intra High Efficiency.</p> <p>Modes Intra 16×16.</p> <p>Transformée DCT 16×16.</p> <p>CS algorithme de reconstruction ADM [Yang & Zhang 09].</p> <p>QPs 22, 27, 32, 37.</p> <p>Séquences 5 séquences cif sélectionnées et 5 séquences du projet HEVC au format qwvga et wvga.</p> <p>Evaluation Pourcentage de réduction de débit moyen calculé à l'aide de la métrique de Bjontegaard.</p>
--

8.2.2.2.1 Evaluation de la borne maximale

Afin de vérifier l'intérêt de la méthode dans le cadre HEVC, nous avons dans un premier temps évalué les gains potentiels de la méthode en se plaçant dans des configurations non-décodables :

- Test 1 : test avec $P'_i = P_i$.
- Test 2 : test avec $P'_{i,j} \in L_{P'_i} = \{A, B, C, D\}$ où chaque ensemble de paramètres optimaux déterminés sur chacune de ces partitions est ensuite testé directement sur P_i afin de retenir la meilleure configuration parmi les quatre. Les partitions $\{A, B, C, D\}$ étant celles représentées sur la figure 7.3.
- Test 3 : même test que le précédent avec $P'_{i,j} \in L_{P'_i} = \{A, B\}$.

Les gains obtenus sont présentés dans le tableau 8.2 sous la forme de pourcentage de réduction de débit. La première observation est que le Test 1 donne un gain moyen de 1.88% sur le débit avec l'efficacité la plus élevée obtenue sur les séquences Ice cif et BasketballDrill wvga qui atteignent respectivement 4.27% et 3.04% de gains. Ce test correspond au résultat maximal espéré dans la mesure où ce sont les paramètres optimaux déterminés sur la partition courante P_i elle-même qui sont utilisés. Ce premier test confirme l'intérêt d'une méthode basée sur l'échantillonnage compressé notamment pour les séquences disposant de nombreuses zones texturées comme la séquence Ice qui est composée à environ 50% de partitions correspondant à la glace d'une patinoire ou de la même manière la séquence BasketballDrill qui contient de grandes zones de parquet. A l'inverse, les séquences telles que Weather cif ou RaceHorses qwvga disposent aussi de ce type de zones mais en nombre beaucoup plus limité alors que le reste de ces séquences est un contenu difficile à coder.

La seconde observation des résultats vient des Tests 2 et 3 où nous avons évalué la perte potentielle liée à la sélection d'une partition causale P'_i dans le voisinage spatial de la partition courante P_i . Logiquement, en utilisant les paramètres optimaux déterminés sur l'une des quatre partitions voisines (Test 2), le gain moyen diminue fortement par rapport au premier test mais reste significatif avec 0.72% de gain sur le débit. Ceci est expliqué par la corrélation entre les partitions voisines ce qui est d'autant plus vrai au sein des grandes zones de textures où la méthode peut ainsi conserver son efficacité. On observe notamment cela pour la séquence BasketBallDrill wvga, tandis que la séquence Ice cif a un comportement différent avec un gain très faible de 0.11% que nous considérons comme un cas particulier où une suite de choix RD délicats a eu un effet global négatif sur le codage de la séquence (ceci est confirmé par les résultats du Test 3 où le gain de cette séquence redevient plus logique par rapports aux autres résultats). Enfin, le dernier test (Test 3), où uniquement la partition voisine de gauche et du dessus sont utilisées, donne des résultats qui restent consistants avec le Test 2 avec un gain moyen de 0.55%. De plus l'évolution des gains de chaque séquence est davantage en accord avec le Test 1 du fait de la proximité des deux partitions utilisées par rapport au second test.

	Test 1	Test 2	Test 3
Séquence	CIF		
Carphone	2.44	1.43	0.77
Ice	4.27	0.11	0.65
Modo	1.58	1.08	0.56
Sean	0.93	0.58	0.42
Weather	0.69	0.24	0.16
	QWVGA		
BasketBallPass	1.64	0.37	0.41
RaceHorses	1.10	0.60	0.42
	WVGA		
BasketBallDrill	3.04	1.08	0.89
BQMall	1.73	0.76	0.66
RaceHorses	1.39	0.95	0.60
Moyenne	1.88	0.72	0.55

TABLE 8.2: Pourcentages de réduction de débit maximums de notre méthode basée sur l'échantillonnage compressé et l'approche décodeur intelligent pour trois configurations non-décodables.

La conclusion de cette étude préliminaire est donc qu'une méthode basée échantillonnage compressé intégrée dans un codeur vidéo moderne comme HEVC peut apporter des gains intéressants dans la mesure où les paramètres de la méthodes sont déterminés de façon optimale (Test 1). La détermination desdits paramètres sur une partition causale dans le cadre de l'approche décodeur intelligent (Test 2 et 3) apporte une solution

qui permet de conserver une partie significative des gains de la méthode sous la condition d'être capable de retrouver la meilleure partition causale à utiliser. Dans ce but, la solution de limiter le nombre de partitions causales candidates à deux (Test 3) sera privilégiée dans le test décodable suivant dans la mesure où cela assure moins d'échecs lors du choix de la partition à utiliser. Si la partition causale optimale est signalée au décodeur cela permet alors de limiter le coût de signalisation par rapport à une solution avec un plus grand nombre de candidat.

8.2.2.2 Evaluation de la méthode

Nous présentons maintenant, dans le tableau 8.3, les résultats décodés de la méthode proposée. Nous avons comparé ces résultats aux méthodes de références introduites dans l'état de l'art sur l'échantillonnage compressé (section 6.2.2.2) et intégrées dans HEVC :

- CS1 correspond à la méthode introduite dans [Zhang *et al.* 08].
- CS2 correspond à la méthode introduite dans [Do *et al.* 10].

	DI-CS	CS1	CS2
Séquence	CIF		
Carphone	0.62	-0.27	-0.11
Ice	0.59	-0.11	-0.28
Modo	0.82	-0.24	0.19
Sean	0.47	0.13	0.18
Weather	0.41	-0.08	0.00
	QWVGA		
BasketBallPass	0.30	-0.28	-0.17
RaceHorses	0.25	-0.08	0.26
	WVGA		
BasketBallDrill	0.20	0.24	-0.05
BQMall	0.39	-0.11	-0.04
RaceHorses	0.49	-0.19	-0.20
Moyenne	0.45	-0.10	-0.02

TABLE 8.3: Pourcentages de réduction du débit de notre méthode (DI-CS) et des méthodes de l'état de l'art intégrées dans HEVC.

La première observation est que la méthode proposée apporte en moyenne un gain non-négligeable de 0.45% de réduction du débit et que la méthode améliore systématiquement l'efficacité de codage sur les séquences considérées.

Un gain minimum de 0.20% est rapporté pour la séquence BasketBallDrill wvga et un gain maximum de 0.82% est obtenu pour la séquence Modocif. A propos du gain

de cette séquence, notons qu'il est supérieur au gain espéré dans cette configuration de 0.56% selon le tableau 8.2 de la même manière que pour les séquences Sean et Weather Cif ce que l'on explique par des choix débit-distorsion locaux globalement meilleurs pour ces trois séquences. Dans l'ensemble, ces gains sont finalement plutôt limités par rapport au gain maximal envisagé (Test 1) puisqu'ils en représentent uniquement un quart mais ils demeurent consistants par rapport à la configuration du Test 3 du tableau 8.2.

Les deux autres colonnes donnent les résultats des méthodes proposées dans l'état de l'art que nous avons intégré dans HEVC. Ces deux méthodes ne sont pas efficaces sur les séquences testées avec une faible perte en moyenne pour la méthode CS1 tandis que la méthode CS2 est légèrement meilleure avec une efficacité de codage équivalente à la référence HEVC. En effet cette méthode constitue une amélioration de CS1 du fait de l'intégration de la prédiction lors de la reconstruction qui permet de reconstruire directement dans le domaine du pixel. Cependant, cette amélioration de la méthode permet tout juste de compenser le coût de signalisation. Ces résultats, différents de ceux présentés par les auteurs, s'expliquent par la configuration de test qui est moins restrictive et l'intégration de la méthode dans HEVC qui a une efficacité de codage plus élevée que H.264/AVC et est donc plus difficile à améliorer. En comparaison de ces méthodes, la méthode proposée avec combinaison de l'approche décodeur intelligent apporte une réelle amélioration.

Cela confirme que la fine paramétrisation de la méthode sans la signalisation associée est une solution prometteuse pour ce type de méthodes itératives qui sont très dépendantes des paramètres.

Les gains obtenus par notre méthode demeurent néanmoins limités relativement à la complexité ajoutée. Il est en effet délicat de faire cohabiter l'échantillonnage compressé avec un système de codage classique. Les résiduels issus de la prédiction du codeur HEVC étant déjà très faibles et avec un grand nombre de valeurs nulles, la méthode a moins de sens que lorsqu'elle est appliquée sur une partition originale. Ces résultats nous ont poussé à ne pas tester l'ajout d'un autre opérateur d'acquisition en compétition tel que cela est fait dans l'étude théorique 8.2.1 avec l'ondelette du fait des modifications du codeur que cela engendre.

Ces résultats concluent notre étude : l'approche décodeur intelligent permet en effet d'optimiser l'intégration d'une méthode basée échantillonnage compressé dans un codeur vidéo moderne mais son efficacité reste cependant limitée dans un tel cadre face aux méthodes de compression classique.

8.3 Bilan

Dans ce chapitre, nous avons proposé un schéma de codage vidéo basé sur l'échantillonnage compressé permettant d'exploiter efficacement les redondances spatiales et

temporelles grâce à son intégration dans le cadre de l'approche décodeur intelligent. Après sélection d'une partition causale adaptée, les paramètres de la méthode sont déterminés puis hérités pour le codage de la partition courante par échantillonnage compressé. Cela permet un gain significatif sur le coût de transmission des données ainsi que sur la qualité de la reconstruction.

Cette méthode a été évaluée d'une part dans un cadre théorique où sa pertinence est démontrée dans la mesure où elle permet une acquisition adaptative des données selon le contenu des partitions et le débit ciblé. Notre solution améliore ainsi les performances d'autres méthodes de l'état de l'art. D'autre part, notre méthode a été évaluée dans un codeur vidéo moderne (HEVC) où elle est intégrée au système de codage classique et mise en compétition avec les autres modes. La méthode proposée permet d'obtenir des gains significativement meilleurs que les autres méthodes de l'état de l'art de l'échantillonnage compressé mais son efficacité par rapport à la complexité apportée demeure trop faible du fait de la qualité de la prédiction originale de HEVC.

Conclusion et perspectives

Rappel des objectifs

L'objectif de cette thèse était de proposer des outils efficaces pour le codage vidéo permettant d'améliorer les performances de compression des schémas existants, cet objectif s'inscrivant de façon modeste dans le défi offert par l'évolution de la vidéo numérique qui nécessite un ratio de compression de plus en plus élevé.

Deux axes de recherches ont ainsi été définis avec une première phase de travaux associés aux schémas de codage conventionnels et une seconde phase davantage en rupture. Le fil conducteur de l'ensemble de nos travaux devait néanmoins être l'exploitation accrue des capacités d'un décodeur rendu plus actif. En particulier, un effort particulier a été mené pour définir des méthodes s'abstenant de la transmission d'informations de signalisations additionnelles ou permettant de réduire le débit alloué à ce type d'informations. Enfin, nous avons souhaité appuyer nos travaux sur des domaines connexes du traitement du signal afin de proposer des méthodes originales.

Bilan des travaux effectués

Approches conventionnelles pour les schémas de codage vidéo émergents

L'observation de la proportion importante des informations de signalisations au sein des schémas de codage conventionnels a été le point de départ de notre étude sur la transmission d'indices de compétition par une technique issue du tatouage vidéo. L'information résultant des nombreux choix en compétition a en effet une proportion significative dans le flux, en particulier à bas débits, et le sujet de l'amélioration de la transmission de ces informations est peu abordé en dehors du codage entropique. Nous avons tout d'abord réalisé une étude préliminaire approfondie pour définir d'une part l'information de codage à traiter et d'autre part la technique de tatouage la plus appropriée dans le cadre de la compression où les contraintes sont différentes. La solution retenue a été le masquage d'indices de compétition binaires dans les composantes de chrominance et de luminance des résiduels de textures après transformation et quantification. Plus précisément la parité de la somme de ces résiduels a été utilisée dans la mesure où elle

permet de sélectionner habilement la modification optimale à effectuer lorsque cela est nécessaire. Cette modification est choisie selon un critère débit-distorsion pour lequel nous avons utilisé une méthode d'optimisation lagrangienne classique puis proposé une méthode originale basée sur l'optimisation multi-critères de Pareto. Notre méthode a été testée pour la transmission d'indices de compétition de prédicteurs de mouvement en Inter ainsi que pour la transmission des indices most probable mode en Intra. Elle améliore significativement l'efficacité de codage dans le cas Inter et est d'autant plus efficace en Intra où nous disposons de davantage de coefficients résiduels pour le tatouage. La méthode se distingue de plus par la très faible complexité ajoutée au décodeur et a par conséquent été l'objet de nombreuses études menées par différentes compagnies dans le cadre du futur standard HEVC. L'outil a ainsi été adopté dans le HM 6.0 et intégré dans la sixième version du *Working Draft* [Bross *et al.* 12] pour faire partie du principal profil de la future norme HEVC.

Nous avons ensuite isolé le sujet de la transmission des vecteurs de mouvement dont le codage a fait l'objet de nombreuses activités de recherche au cours des dernières années. Les vecteurs de mouvement représentent en effet une part significative du flux, notamment à bas débit où il s'agit de l'information la plus coûteuse. Par ailleurs, nous avons observé la très forte redondance entre les vecteurs de mouvements de blocs voisins spatialement et temporellement qui n'est pas exploitée suffisamment par les méthodes existantes. Ces dernières se basent principalement sur une amélioration de la prédiction des vecteurs à l'aide de vecteurs précédemment codés et nécessitent la transmission d'indices additionnels. En opposition avec ces approches nous avons souhaité, toujours en exploitant les vecteurs causaux, réaliser un pronostic très précis du vecteur du bloc courant afin de privilégier les mouvements les plus probables. Après observation d'une quantité représentative du mouvement dans un voisinage local du bloc courant, nous avons ainsi utilisé une prédiction à l'aide de la méthode de Parzen pour déterminer la probabilité de l'ensemble des vecteurs résiduels et finalement modifier la distribution de ces résiduels avant codage entropique. Par conséquent, cette solution ne nécessite pas de transmission d'informations supplémentaires dans la mesure où le décodeur est capable de reproduire la même redistribution des résiduels et d'ainsi décoder le vecteur correctement. Nous avons montré que la méthode tire particulièrement profit de l'exploitation d'une quantité beaucoup plus élevée des redondances du mouvement et de nouveaux prédicteurs judicieux qui peuvent être ajoutés dans l'information de mouvement considérée et permettre d'anticiper de nombreux mouvements qui sont alors codés efficacement. Il a cependant été nécessaire de réaliser un compromis délicat sur la quantité de vecteurs considérés pour que la méthode soit performante. Les résultats observés ont confirmé la pertinence de la méthode qui se positionne au-dessus d'une méthode par compétition de prédicteurs et dont la combinaison avec cette dernière a été prometteuse bien qu'accompagnée d'une augmentation de la complexité.

Approches en rupture basées sur l'exploitation d'un décodeur intelligent

Les travaux que nous avons réalisés dans un second temps se positionnent plus en amont. Nous nous sommes basés sur une analyse théorique et pratique de l'évolution des derniers standards de compression vidéo pour dégager une limite de ces schémas. Ceux-ci sont basés sur une exploitation des redondances spatiales et temporelles dans le signal à l'aide de multiples choix en compétition qui sont systématiquement augmentés entre chaque génération de standard tandis que la proportion de résiduels n'a cessé de diminuer. Afin de se détacher de cette contrainte liée à la compétition qui nécessite une signalisation de plus en plus coûteuse, nous avons proposé une approche basée sur l'exploitation accrue du décodeur. A l'aide d'une partition causale judicieusement sélectionnée, les paramètres de codage sont dérivés en utilisant cette partition comme référence puis ils sont utilisés pour le traitement de la partition courante qui économise alors la transmission de ces informations. En effet, le décodeur est capable de déterminer à son tour les mêmes paramètres lors du décodage en entraînant cependant une augmentation de la complexité. Nous avons démontré que cette approche ouvrait la voie à deux axes d'améliorations des schémas actuels avec d'une part la réduction des coûts de signalisation et d'autre part la possibilité d'ajouter des choix de codages en compétition sans être contraint par le débit associé.

A titre d'étude préliminaire, cette approche a dans un premier temps été appliquée à la détermination de modes de prédiction Intra. Ceux-ci sont dérivés à l'encodeur comme au décodeur en utilisant un ensemble de partitions voisines prédéfini pour lesquelles les nombreux prédicteurs Intra sont mis en compétition puis sélectionnés selon un critère reproductible lors du décodage. Intégrée dans le futur standard HEVC en compétition avec les modes de codage classique, cette solution a démontré le potentiel de l'approche proposée avec notamment un pourcentage de sélection significatif et la démonstration de l'apport de l'augmentation des modes.

Finalement, nous avons exploré l'utilisation d'un schéma d'échantillonnage compressé dans un but de compression par le biais de l'approche décodeur intelligent. Ce domaine émergent du traitement du signal est en effet une voie de recherche intéressante pour la compression vidéo bien qu'elle soit délicatement combinable avec une étape de quantification. Au-delà de cette question, nous avons souhaité démontrer qu'avec une paramétrisation pointue de la méthode, il était possible d'atteindre des résultats prometteurs. La solution que nous avons proposée consiste tout d'abord à déterminer une partition causale associée à un vecteur de mouvement transmis par l'encodeur. La simulation de l'acquisition et de la reconstruction de cette partition par échantillonnage compressé est alors effectuée avec différents paramètres dont le pas d'échantillonnage et le domaine d'acquisition parmi lesquels le choix optimal est sélectionné pour être appliqué à la partition courante. Seules l'information de mouvement et les mesures résultantes sont alors transmises au décodeur qui reconstruit de façon efficace le signal. Une expérimentation théorique de ce schéma a été réalisée puis un mode de codage basé sur cette approche a été introduit dans l'encodeur HEVC. Les résultats qui ont été obtenus sont encou-

rageants, en particulier pour l'expérimentation théorique qui permet d'atteindre une reconstruction efficace avec un nombre de mesures limitée pour un niveau de qualité donné du fait du caractère adaptatif de la méthode.

Perspectives

Au moment de clore ce manuscrit, des perspectives prometteuses apparaissent suite aux travaux réalisés. Les principales voies de développement concernent l'approche décodeur intelligent mais d'autres débouchés à plus court terme sont envisageables pour les outils proposés dans la première partie du manuscrit.

Revenons tout d'abord sur notre schéma de tatouage d'informations de codage. Un effort particulier doit être mené pour déterminer l'information la mieux adaptée à la méthode tout en ayant l'impact le plus élevé dans le flux comme cela est en train d'être réalisé par divers travaux au sein du JCT-VC dans le cadre du codeur HEVC. Celui-ci est d'autant plus adapté à la méthode qu'il utilise des blocs de tailles supérieures à ceux dont nous disposons dans nos travaux réalisés au sein d'H.264/AVC et où la méthode était particulièrement efficace avec les blocs de taille 16×16 . Au-delà de cette analyse des paramètres à considérer, il serait intéressant de disposer d'une optimisation moins complexe au niveau de l'encodeur pour sélectionner la meilleure modification sans tester l'ensemble des combinaisons. L'exploitation de statistiques pourrait aider dans cette démarche. Enfin, la méthode pourrait être adaptée au pas de quantification en prenant en compte la présence d'un nombre plus élevé de coefficients à haut débit pour masquer davantage d'informations en utilisant, par exemple, un découpage fictif des blocs transformés dont la taille dépendrait du débit. Suite à un récent stage sur le sujet, une partie de ces idées ont été explorées au sein de notre équipe et ont donné lieu à l'acceptation de l'outil dans le futur standard HEVC.

De la même manière, notre méthode de codage des vecteurs de mouvement par redistribution des résiduels pourrait être envisagée dans le cadre du futur standard HEVC dans la mesure où nous avons démontré que notre technique demeurerait efficace lorsqu'elle coexistait avec une méthode de compétition de prédicteurs. La présence du mode Merge nécessite cependant une réflexion supplémentaire dans la mesure où ce dernier capte beaucoup des redondances dans le mouvement et conduit aussi à une réduction de la diversité des vecteurs dont nous avons besoin dans notre outil. L'étude de combinaisons plus efficaces avec d'autres outils de codage du mouvement efficaces serait donc une voie de travail pertinente. Nous avons de plus envisagé d'effectuer notre méthode par composante (horizontale et verticale), en particulier les étapes d'estimation de la densité de probabilité et de modification de la distribution des résiduels. L'utilisation d'une telle solution 1D permettrait de réduire significativement la complexité de la méthode, notamment le mapping qui serait réduit à des permutations. Il serait alors possible de limiter la perte d'efficacité de la méthode en l'appliquant à la seconde composante relativement à la première composante déjà codée : seules les positions de résiduels observés

appartenant aux mêmes vecteurs seraient alors utilisées ce qui limiterait la redistribution aux positions utiles. Enfin, comme nous l'avons évoqué, un apprentissage pourrait être réalisé pour définir la valeur du paramètre de lissage utilisée dans la méthode de Parzen et ainsi optimiser la quantité de résiduels redistribués pour chaque bloc.

Les perspectives relatives à l'approche décodeur intelligent ont été abordées à la fin du chapitre 7. Elles concernent principalement la détermination et la signalisation de la partition causale. En particulier, le développement de techniques de mise en correspondance de blocs dans un domaine différent que le domaine pixellique, comme le domaine transformé ou le domaine des modes de codage semble très prometteur. L'étude de cette approche dans le cadre de travaux liés à la scalabilité en complexité est elle-aussi pertinente et va donner lieu à l'ouverture d'une nouvelle thèse au sein de notre équipe.

La base que nous avons ensuite proposée pour la définition d'un schéma de codage fondé sur l'échantillonnage compressé à partir de cette approche décodeur intelligent semble très encourageante et pourrait être rapidement améliorée. L'intégration des dernières avancées de ce domaine concernant notamment les opérateurs et la quantification des mesures pourrait ainsi faire évoluer notre schéma. A la manière des encodeurs vidéo modernes, une mise en compétition avec d'autres méthodes basées sur des blocs et exploitant les redondances temporelles différemment permettrait aussi d'atteindre des performances significatives. Quant à l'intégration de tels schémas dans un codeur vidéo conventionnel, le challenge s'avère plus délicat mais il serait intéressant d'appliquer directement la méthode sur les blocs sans passer par la prédiction. Le principe d'origine de l'échantillonnage compressé serait ainsi conservé et il serait alors possible d'optimiser la dérivation des paramètres à l'aide de l'approche décodeur intelligent. Le mode de codage ainsi défini pourrait être particulièrement efficace pour des blocs suffisamment parcimonieux dans les bases transformées utilisées.

Annexe

Annexe A

Hybrid-1D : un nouveau mode de prédiction

Nous décrivons dans cette annexe une méthode¹ étudiée au cours de cette thèse faisant suite à un précédent travail réalisé au sein de notre équipe. Un nouveau mode de prédiction et de partitionnement des macroblocs est proposé pour étendre le mode Intra 1D en prédiction Inter. Ce mode de codage tire profit de partitions linéaires permettant d'exploiter habilement une distance plus courte entre signal de référence et signal à prédire. Nous l'appliquons en prédiction Inter en ajoutant tout d'abord une compétition avec une prédiction Skip 1D puis dans un second temps une prédiction Inter 1D.

A la différence du schéma de codage d'H.264/AVC, les moyens de prédiction sont ainsi mis en compétition au sein du macrobloc pour les différentes partitions générées. Nous nous rapprochons ainsi du schéma du futur standard HEVC dont les travaux n'avaient alors pas encore débuté.

A.1 Etat de l'art

Notre travail est basé sur le mode Intra 1D introduit dans [Laroche *et al.* 09] qui propose de modifier le codage Intra à l'aide de quatre changements que nous reprendrons dans la seconde section : partitionnement, ordre de parcours, prédicteurs et transformée. Afin de réduire la distance entre les pixels à coder et le signal prédicteur, les auteurs proposent d'utiliser des partitions 1D. Le mode défini a permis des gains élevés comparé à la prédiction Intra d'H.264/AVC basée sur des blocs. Il a d'ailleurs donné lieu à de nombreuses études réalisées dans le cadre d'HEVC mais n'a pour l'instant pas été accepté au vue de la modification trop importante de la structure de codage malgré des performances significatives.

Par ailleurs, nous avons observé que le standard H.264/AVC dispose de trois types

1. Cette description est basée sur notre article [Conf1].

d'images (I, P et B) mais chaque macrobloc est toujours de type Intra, Inter ou Skip à l'exception de l'Inter 8×8 où deux types de prédiction peuvent être mélangés avec le mode Direct et l'Inter classique. Quelques travaux ont cependant essayé de combiner plusieurs types de prédiction, notamment Intra et Inter. Citons [Xin *et al.* 07] où les auteurs proposent d'exploiter à la fois les corrélations temporelles et spatiales. La méthode réalise une combinaison à l'aide de poids d'une prédiction Inter 16×16 et Intra 16×16 (Horizontale). Les résultats, notamment pour des séquences HD, montrent que la combinaison de ces prédictions est un sujet de recherche pertinent dans la mesure où les énergies dans les résiduels Inter et Intra sont distribuées différemment et peuvent s'avérer complémentaires. Dans [Li *et al.* 09], un outil intitulé *Second Order Prediction* est introduit dans H.264/AVC. Il applique une prédiction Intra sur les résiduels issus d'une prédiction Inter. Bien que la complexité augmente de façon non-négligeable, les gains rapportés confirment que l'utilisation de prédiction Intra et Inter au sein d'un même macrobloc est efficace.

A.2 Méthode proposée : Hybrid 1D

L'environnement Intra 1D défini dans [Laroche *et al.* 09] a montré des résultats probants en prédiction Intra, nous l'étendons par conséquent en prédiction Inter en définissant le mode Hybrid 1D. Ce mode diffère du récent état de l'art par le fait que la signalisation du partitionnement demeure faible et que les prédictions temporelle et spatiale sont combinées au sein du même macrobloc. Chaque partition 1D peut ainsi être codée selon l'un des sous-modes en compétition : Intra 1D, Skip 1D ou Inter 1D.

A.2.1 Intra 1D + Skip 1D

A.2.1.1 Partitionnement 1D

Le partitionnement utilisé est similaire à celui proposé dans [Laroche *et al.* 09]. Nous décrivons le principe général ci-dessous.

Le partitionnement 1D d'un macrobloc est complètement défini à partir d'une partition initiale p_1 qui est signalée au décodeur à l'aide d'un indice. Les P autres partitions, $p_i, \forall i = 1 \dots P$, sont obtenues par propagation de la forme initiale p_1 à l'aide d'un opérateur de dilatation morphologique mathématique utilisant pour élément structurant une forme en croix de taille 3×3 pixels. Des exemples de partitions initiales et les propagations correspondantes sont représentées sur la figure A.1 pour des partitions *Ligne*, *Colonne*, *L'*, *Point* et *Contour*. Dans nos travaux, nous avons utilisé les deux premières formes qui permettent de diviser le macrobloc en seize lignes ou seize colonnes de seize pixels.

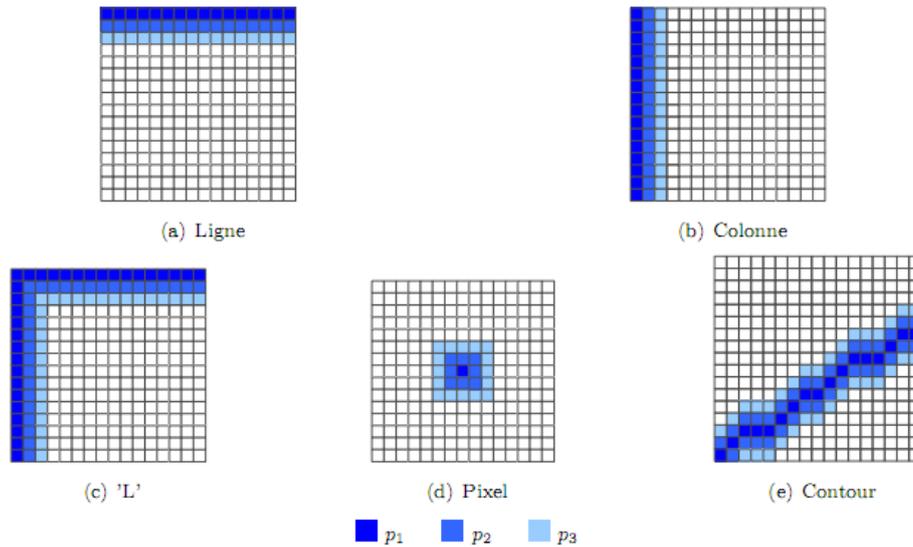


FIGURE A.1: Exemple de partitionnements obtenus à partir d'une partition initiale p_1 .

A.2.1.2 Ordres de parcours

Trois ordres de parcours ont été proposés dans [Laroche *et al.* 09] : parcours lexicographique, bi-directionnel et hiérarchique. Dans le souci de ne pas surcharger la signalisation, les résultats que nous présentons dans cette annexe correspondent à un test réalisé avec le premier ordre de parcours qui consiste à traiter les partitions dans l'ordre lexicographique : p_1, p_2, \dots, p_P . Précisons que les autres ordres de parcours sont plus complexes et permettent d'avoir plusieurs partitions comme référence.

A.2.1.3 Prédiction

Pour chaque partition 1D d'un macrobloc, deux sous-modes sont mis en compétition : Intra 1D et Skip 1D. Le meilleur prédicteur pour chaque sous-mode est sélectionné puis le meilleur sous-mode est sélectionné selon un critère débit-ditorsion décrit dans la sous-section A.2.4.

Intra 1D

L'intra 1D bénéficie de la définition de prédicteurs spécifiques exploitant la distance réduite avec le signal de référence par rapport aux modes Intra par bloc. Nous détaillons ici les prédicteurs utilisés pour le partitionnement Ligne et l'ordre de parcours lexicographique. Ils peuvent être généralisés aux autres partitionnements 1D et ordres de parcours.

Dans ce cas, pour une partition p_i , le signal de référence est composé de la partition p_{i-1} reconstruite et de la zone causale à gauche du macrobloc courant. L'intérêt de

ce mode est que chaque pixel de p_i dispose d'un pixel voisin dans le signal de référence contrairement à l'Intra 16×16 par exemple où le pixel en bas à droite du bloc courant est à une distance de 16 pixels du signal de référence. Ainsi, on retrouve comme prédicteur la copie de la ligne précédente ainsi que diverses combinaisons utilisant les pixels à gauche de la ligne courante.

Dans la mesure où la partition p_{i-1} peut avoir été codée avec le sous-mode Skip 1D dont la qualité de prédiction est plus faible, la partition p_i peut être moins bien prédite du fait d'un signal de référence moins pertinent. Cependant, à la manière des images P et B qui tirent profit d'images précédemment codées en Intra, le macrobloc codé en Hybrid 1D bénéficie judicieusement de partitions prédites en Intra 1D.

Skip 1D

Le second sous-mode est le Skip 1D. C'est le sous-mode que nous avons introduit dans le cadre de ce travail et qui est un cas particulier de l'Inter 1D présenté dans la sous-section suivante. A la manière du mode Skip dans H.264/AVC, ce sous-mode réalise une prédiction temporelle sans transmission d'information (ni vecteurs de mouvement ni résiduels de texture). Un indice est transmis pour signaler que la partition est codée en Skip 1D. La partition décodée correspond alors à la partition de l'image de référence compensée en mouvement à l'aide du vecteur de mouvement prédicteur que nous décrirons dans la sous-section sur l'Inter 1D.

A.2.1.4 Transformée 1D

Le standard H.264/AVC utilise les transformées DCT 4×4 et 8×8 pour traiter les résiduels Intra et Inter. Pour notre schéma, nous utilisons une DCT 1D flottante afin de s'accorder avec les formes linéaires des partitions proposées. Concernant la quantification, nous utilisons celle définie pour la DCT 4×4 en arrangeant les résiduels obtenus au sein d'un bloc 4×4 . Dans le mesure où le sous-mode Skip 1D ne transmet pas de résiduels, seuls les résiduels de l'Intra 1D sont transformés et quantifiés.

A.2.2 Inter 1D

Comme nous le verrons dans les résultats, la première combinaison ajoutant le Skip 1D à l'Intra 1D apportent une amélioration de l'efficacité de codage. Nous avons par conséquent étendu de nouveau le schéma en ajoutant un sous-mode Inter 1D. Dans la prédiction Inter 1D, une estimation de mouvement est réalisée pour chaque partition p_i et délivre un vecteur de mouvement $vm_i(vm_{i_x}, vm_{i_y})$. Les résiduels de pixels (ε_i) et le vecteur de mouvement résiduel (ε_{vm_i}) sont alors transmis. Comme explicité ci-dessous, ε_i est obtenu par soustraction avec la partition originale courante p_i et la partition p'_i compensée en mouvement, ε_{vm_i} est obtenu par soustraction avec le vecteur de mouvement prédicteur vm_{pred_i} :

$$\varepsilon_i = p_i - p'_i(x - mv_{i_x}, y - mv_{i_y}). \quad (\text{A.1})$$

$$\varepsilon_{vm_i} = vm_i - vm_{pred_i}. \quad (\text{A.2})$$

L'estimation de mouvement utilise un algorithme de mise en correspondance adapté aux tailles de partitions définies (figure A.2). Dans le cas du partitionnement Ligne, la recherche permet de trouver la meilleure partition 16×1 pixels. Le demi-pixel et le quart de pixel ont été implémentés en utilisant des filtres identiques à ceux d'H.264/AVC.

A la manière de l'Inter 4×4 , le mode proposé permet d'avoir une meilleure reconstruction de la partition du fait de la précision plus fine qui génère une prédiction efficace des pixels. Cependant, dans la mesure où un vecteur de mouvement vm_i doit être transmis par partition, ce sous-mode est pénalisé par le coût de codage du vecteur, la prédiction du vecteur est par conséquent particulièrement importante. Nous avons implémenté deux moyens de prédiction du mouvement :

- Le premier prédicteur utilise le médian spatial d'H.264/AVC, l'ensemble des partitions du macrobloc ont alors le même prédicteur.
- Le second prédicteur proposé utilise le vecteur de mouvement de la dernière partition codée en Inter 1D au sein du macrobloc courant ou bien le premier prédicteur si aucune partition n'est disponible.

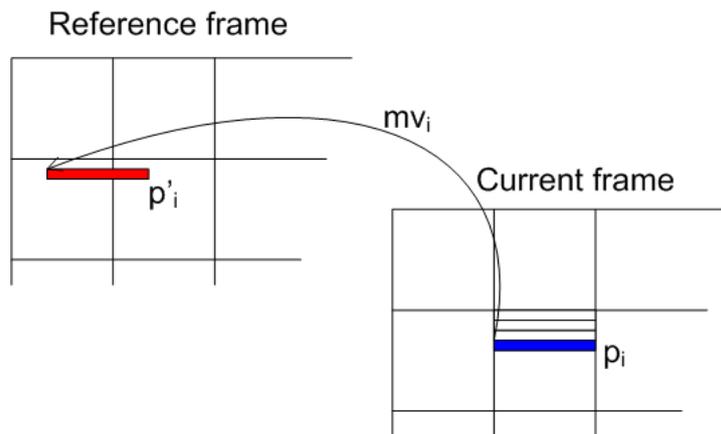


FIGURE A.2: Représentation de la mise en correspondance de partitions 1D.

A.2.3 Signalisation du mode proposé

Pour la signalisation du mode proposé et des sous-modes, nous utilisons une représentation en arbre définie dans la figure A.3. L'Hybrid 1D est signalé au même niveau que les autres modes d'H.264/AVC afin de ne pas biaiser la sélection de l'un des modes, un indice binaire indique alors le type de partition (Ligne ou Colonne). Finalement, un

dernier indice signale le sous-mode optimal pour chaque partition (Skip 1D, Intra 1D ou Inter 1D) avant de transmettre les informations de codage propre à chaque sous-mode (le prédicteur pour l’Intra 1D et le vecteur de mouvement en Inter 1D). La quantité de sous-modes en compétition dans notre schéma étant élevée, leur signalisation est un point clé de la méthode. Un effort particulier a par conséquent été mené pour coder efficacement les sous-modes : inspiré du codage des prédicteurs Intra dans H.264/AVC, nous utilisons l’outil *most probable mode* adapté à notre travail. Cette technique que nous avons détaillée dans le chapitre 3 permet de diminuer le coût de signalisation en indiquant si le sous-mode de la partition courante est le même que celui de la partition précédente. Notons cependant que cette solution peut avoir tendance à favoriser le sous-mode sélectionné en premier. De plus, nous avons utilisé un indice de *raccourci* additionnel en Intra 1D + Skip 1D pour signaler si tous les sous-modes restants sont égaux au sous-mode courant. Ainsi, si les n dernières partitions 1D utilisent le même sous-mode, seulement les $P - n$ premiers indices et un indice de *raccourci* sont transmis au lieu des P indices.

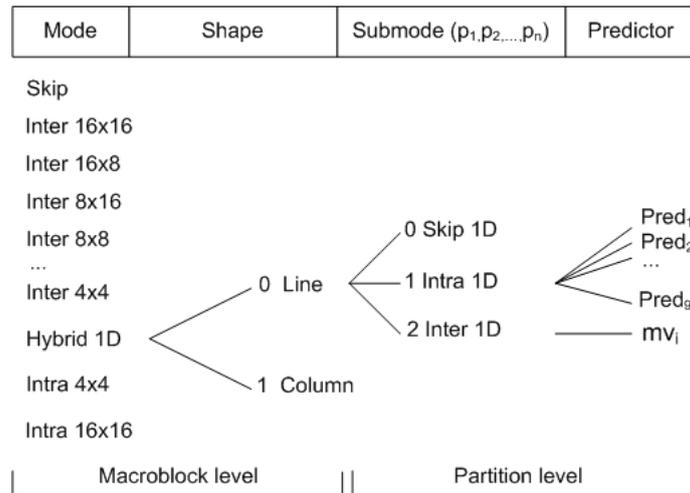


FIGURE A.3: Signalisation Hybrid-1D.

A.2.4 Critère de compétition

Nous avons mis en compétition l’Hybrid 1D avec les autres modes Inter et Intra pour chaque macrobloc à l’aide d’une sélection débit-distorsion classique. Le débit de notre mode est calculé comme suit :

$$R = R_{mode} + R_{coeff} + \sum_{i=1}^P R_{Intra_i} + \sum_{i=1}^P R_{Inter_i}, \quad (A.3)$$

où R_{mode} est le débit associé à la signalisation du mode et des sous-modes, R_{coeff} est le coût des résiduels de texture et R_{intra_i} et R_{inter_i} sont les débits respectifs des sous-modes

Intra et Inter si la partition p_i est codée avec l'un de ces sous-modes. Ils correspondent respectivement au coût des prédicteurs Intra et au coût des vecteurs de mouvements.

Un critère débit-distorsion additionnel $J_i = D_i + \lambda R_i$ est utilisé pour la sélection des sous-modes de chaque partition. D_i représente la distorsion de chaque partition et R_i est le débit calculé comme suit selon le type de sous-mode :

- Intra 1D : coût des résiduels transformés et quantifiés, signalisation du prédicteur Intra 1D et du sous-mode.
- Inter 1D : coût des résiduels transformés et quantifiés, coût du vecteur de mouvement et du sous-mode.
- Skip 1D : coût de la signalisation du sous-mode.

A.3 Résultats expérimentaux

A.3.1 Configuration de test

La configuration utilisée pour les tests est résumée ci-dessous. Nous avons successivement évalué les méthodes Intra 1D, Intra 1D + Skip 1D et Hybrid 1D face à l'encodeur H.264/AVC. Pour ces tests, seules les partitions Ligne et Colonne ont été évaluées avec l'ordre de parcours lexicographique (*Raster*) et les prédicteurs Intra 1D définies dans [Laroche *et al.* 09].

Codeur H.264/AVC, logiciel JSVM (outils scalables désactivés).

GOP IPPP, 12 images.

Modes Inter, Skip et Intra 16×16 et 4×4 + Hybrid 1D.

Transformée DCT 4×4.

Codeur Entropique CABAC.

QPs 22, 27, 32, 37.

Chrominance Désactivée pour cause d'implémentation.

Séquences Cinq séquences classiques des formats CIF à 720p.

Evaluation Pourcentage de réduction du débit moyen calculé par la métrique de Bjontegaard [Bjontegaard 01, Jung & Pateux 07].

A.3.2 Gains sur le débit

Les gains obtenus sont rapportés dans le tableau A.1 pour les trois méthodes. Les modes proposés apportent des gains systématiques par rapport à H.264/AVC. En moyenne, les gains sont de 5.2%, 5.7% et 5.7% respectivement pour l'Intra 1D, l'Intra 1D + Skip 1D et le mode Hybrid 1D. Le meilleur résultat est obtenu pour la séquence Schumacher CIF qui atteint 10.1% de réduction du débit avec la seconde solution.

En moyenne, le schéma avec l’Intra 1D et le Skip 1D donne un gain de 0.5% par rapport à l’Intra 1D seul et jusqu’à 1% pour la séquence SpinCalendar 720p. Le gain apporté par l’Hybrid 1D est approximativement le même du fait du coût des vecteurs de mouvement qui compense le gain en prédiction.

Séquence	Intra 1D	Intra 1D + Skip 1D	Hybrid 1D
Schumacher CIF	9.8	10.1	9.9
City VGA	1.8	2.3	2.4
Exit VGA	2.8	2.9	2.9
Panslow VGA	8.8	9.2	9.5
SpinCalendar 720P	2.9	3.9	3.7
Moyenne	5.2	5.7	5.7

Table A.1: Pourcentages de gains sur le débit pour un encodage avec un GOP de 12 sur les QP 22 à 37 des trois méthodes *Intra 1D*, *Intra 1D+Skip 1D* et *Hybrid 1D* par rapport à la référence H.264/AVC.

A.3.3 Analyse des résultats

Le tableau A.2 donne les pourcentages de sélection de l’ensemble des modes d’H.264/AVC et de l’Hybrid 1D. En moyenne, sur l’ensemble des QPs, l’Hybrid 1D a un pourcentage de sélection de 4.6%, ce qui est plus élevé que les modes Intra ainsi que l’Inter 8×8 et équivalent à la sélection de l’Inter 16×8 et 8×16 . La sélection de notre mode augmente vers les hauts débits pour atteindre 6.5%. Notre mode est en effet coûteux et est donc moins sélectionné à bas débit du fait du compromis débit-distorsion qui privilégie notamment le mode Skip.

Mode	QP				Moyenne
	37	32	27	22	
Intra 16×16	4.8	4.0	3.2	2.7	3.7
Intra 4×4	2.1	3.0	4.0	6.1	3.8
Skip	73.7	66.2	54.9	32.5	56.8
Inter 16×16	12.3	14.8	17.7	26.0	17.7
Inter 16×8	2.0	3.4	5.8	8.8	5.0
Inter 8×16	1.8	3.4	5.4	7.8	4.6
Inter 8×8	0.4	1.4	3.9	9.5	3.8
Hybrid 1D	2.9	3.8	5.1	6.5	4.6

Table A.2: Pourcentages de sélection moyens des modes du standard H.264/AVC et du mode proposé pour quatre QP sur l’ensemble des séquences testées.

Nous avons ensuite rapporté dans le tableau A.3, les pourcentages de sélection de chacun des sous-modes définis. En moyenne, le Skip 1D a un pourcentage de sélection nettement supérieur aux autres sous-modes du fait de son faible coût. Pour le premier schéma, Intra 1D + Skip 1D, le sous-mode Intra 1D est sélectionné en moyenne 22% ce qui permet au mode proposé d'améliorer les gains par rapport à l'Intra 1D seul car la combinaison avec le Skip 1D est efficace : les partitions complexes au sein du macrobloc utilisent le Skip 1D qui est peu coûteux et l'Intra 1D intervient sur certaines partitions plus délicates. Lorsque l'Inter 1D est ajouté, la sélection du Skip 1D demeure équivalente et les deux autres sous-modes se répartissent la sélection restante. La faible proportion de l'Inter 1D confirme cependant que son coût de signalisation trop élevé compense l'amélioration de la reconstruction. Notons enfin que Intra 1D et Inter 1D ont un pourcentage de sélection plus élevé à haut débit où la qualité de la reconstruction est plus importante.

QP	Intra 1D + Skip 1D		Hybrid 1D		
	Intra 1D	Skip 1D	Intra 1D	Skip 1D	Inter 1D
22	25	75	16	74	10
27	22	78	15	77	8
32	20	80	16	78	6
37	20	80	16	80	4
Moyenne	22	78	16	77	7

Table A.3: Pourcentages de sélection de chaque sous-mode proposé dans les schémas *Intra 1D+Skip 1D* and *Hybrid 1D* pour quatre QP en moyenne sur les séquences testées.

Nous avons représenté sur la figure A.4, trois exemples typiques de distribution des sous-modes proposés que nous avons observés lors des tests. La configuration (a) correspond à un macrobloc avec un premier ensemble de partitions codées en Intra 1D et les suivantes en Skip 1D. Dans ce cas, le macrobloc dispose d'une reconstruction de bonne qualité pour sa partie supérieure avec un signal de référence proche tandis que la suite du macrobloc est codée à moindre coût. Ce type de macrobloc peut être plus efficace que certains macroblocs Skip d'H.264/AVC dans la mesure où un faible coût de codage est conservé avec une prédiction efficace de certaines partitions en Intra 1D. Le second cas (b) représente une configuration similaire où une partition du macrobloc est prédite en Inter 1D et la partition suivante, bénéficiant alors d'un signal de référence efficace, est par conséquent codée en Intra 1D. Cet exemple typique peut correspondre à un contour qui couperait le macrobloc et où un mode Skip classique n'aurait pas pu être efficace. Enfin, dans le dernier exemple (c), le sous-mode Skip 1D est suffisamment efficace pour les premières lignes puis dans la mesure où le vecteur de mouvement prédicteur devient de moins en moins précis, le sous-mode Inter 1D est privilégié par la décision débit-distorsion.

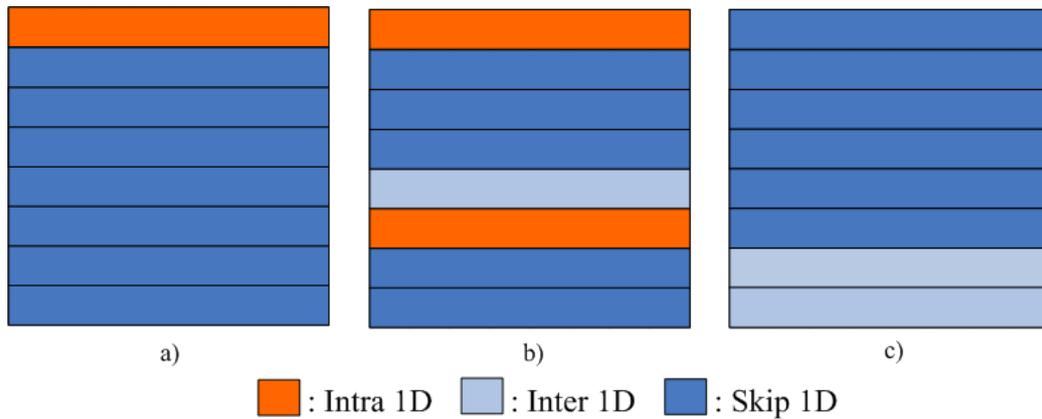


FIGURE A.4: Exemples de distributions des sous-modes proposés.

A.4 Bilan

En conclusion, nous avons proposé dans cette annexe une extension de l’Intra 1D pour le codage Inter. Le mode Hybrid 1D étudié divise le macrobloc en partitions 1D (Ligne ou Colonne) et nous avons tout d’abord mis en compétition les sous-modes Intra 1D et Skip 1D pour chaque partition. Dans un second temps, un sous-mode Inter 1D a été ajouté dans le but d’obtenir une prédiction plus efficace. L’intégration efficace d’un mode Inter 1D nécessite cependant plusieurs modifications de la structure de codage et engendre des difficultés quant au codage des vecteurs de mouvement, la signalisation des sous-modes et la décision. Nous avons proposé des solutions techniques pour ces problèmes et présenté les résultats obtenus par le mode Hybrid 1D dont la qualité de prédiction permet uniquement de compenser le surcoût de codage engendré par rapport à la combinaison Intra 1D + Skip 1D. Cette étude de début de thèse nous a permis de pointer la problématique majeure que représente la signalisation des choix en compétition. Cette conclusion a ainsi orienté la suite de nos travaux vers l’étude de méthodes sans signalisation supplémentaire comme le codage du mouvement présenté au chapitre 5 ou bien nos solutions pour économiser la transmission de certaines informations avec le tatouage au chapitre 4 ainsi que l’approche décodeur intelligent dans la seconde partie.

Néanmoins, la combinaison de la prédiction efficace de l’Intra 1D avec le sous-mode peu coûteux Skip 1D nous a permis de démontrer des gains significatifs en comparaison du standard H.264/AVC avec une réduction du débit de 5.7% en moyenne.

Bibliographie

- [Aaron & Girod 02] A. Aaron & B. Girod. *Compression with side information using turbo codes*. In Data Compression Conference, 2002. Proceedings. DCC 2002, pages 252 – 261, 2002.
- [Aaron *et al.* 03] A. Aaron, E. Setton & B. Girod. *Towards Practical Wyner-Ziv Coding of Video*. In in Proc. IEEE International Conference on Image Processing, pages 869–872, 2003.
- [Agostini & Antonini 08] M.A. Agostini & M. Antonini. *Multiple description video decoding using map*. In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pages 1228 –1231, Octobre 2008.
- [Agostini *et al.* 06] M.A. Agostini, T. Andre, M. Antonini & M. Barlaud. *Modeling the Motion Coding Error for Mcwt Video Coders*. In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, volume 2, page II, may 2006.
- [Agostini-Vautard *et al.* 11] M. A. Agostini-Vautard, M. Cagnazzo, M. Antonini, G. Laroche & J. Jung. *A New Coding Mode for Hybrid Video Coders Based on Quantized Motion Vectors*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 21, no. 7, pages 946 –956, july 2011.
- [Bjontegaard 01] G. Bjontegaard. *Calculation of Average PSNR Differences Between RD-Curves*, Avril 2001.
- [Bross *et al.* 11] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm & T. Wiegand. *WD5 : Working Draft 5 of High-Efficiency Video Coding*, Novembre 2011.
- [Bross *et al.* 12] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm & T. Wiegand. *High Efficiency Video Coding (HEVC) Text Specification Draft 6*, Février 2012.
- [Cagnazzo *et al.* 04] M. Cagnazzo, T. Andre, M. Antonini & M. Barlaud. *A smoothly scalable and fully JPEG2000-compatible video co-*

- der*. In Multimedia Signal Processing, 2004 IEEE 6th Workshop on, pages 91 – 94, Septembre 2004.
- [Cagnazzo *et al.* 07] M.. Cagnazzo, F.. Castaldo, T.. Andre, M.. Antonini & M.. Barlaud. *Optimal Motion Estimation for Wavelet Motion Compensated Video Coding*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 17, no. 7, pages 907–911, Juillet 2007.
- [Campisi *et al.* 02] P. Campisi, D. Kundur, D. Hatzinakos & A. Neri. *Compressive data hiding : An unconventional approach for improved color image coding*. EURASIP Journal on Applied Signal Processing, vol. 2002, no. 1, pages 152–163, 2002.
- [Candes & Wakin 08] E.J. Candes & M.B. Wakin. *An Introduction To Compressive Sampling*. Signal Processing Magazine, IEEE, vol. 25, no. 2, pages 21 –30, march 2008.
- [Candes *et al.* 06] E.J. Candes, J. Romberg & T. Tao. *Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information*. Information Theory, IEEE Transactions on, vol. 52, no. 2, pages 489–509, 2006.
- [Candes 06] E.J. Candes. *Compressive sampling*. In Proceedings of the International Congress of Mathematicians, volume 3, pages 1433–1452. Citeseer, 2006.
- [Chaumont & Puech 06] M. Chaumont & W. Puech. *A DCT-Based Data-Hiding Method to Embed the Color Information in a JPEG Grey Level Image*. In Eusipco, 2006.
- [Chen & Jr 97] Michael C. Chen & Alan N. Willson Jr. *A Spatial and Temporal Motion Vector Coding Algorithm for Low-Bit-Rate Video Coding*. Image Processing, International Conference on, vol. 2, page 791, 1997.
- [Chien 11] Wei-Jung Chien. *CE12 : Summary Report of Adaptive Motion Vector Resolution*, Janvier 2011.
- [Clare *et al.* 12] G. Clare, F. Henry, J. Jung & S. Pateux. *CE11 : Sign Data Hiding*, Février 2012.
- [Cover & Thomas 91] Thomas M. Cover & Joy A. Thomas. Elements of information theory. Wiley-Interscience, New York, NY, USA, 1991.
- [Dai *et al.* 09a] J. Dai, O.C. Au, W. Yang, C. Pang, F. Zou & Y. Liu. *Motion vector coding based on predictor selection and boundary-matching estimation*. In IEEE International Workshop on Multimedia Signal Processing (MMSP), pages 1–5, Brazil, 2009. IEEE.

- [Dai *et al.* 09b] Wei Dai, Hoa Vinh Pham & O. Milenkovic. *Distortion-rate functions for quantized compressive sensing*. In Networking and Information Theory, 2009. ITW 2009. IEEE Information Theory Workshop on, pages 171–175, Juin 2009.
- [de Simone *et al.* 10] F. de Simone, L. Goldmann, J.-S. Lee, T. Ebrahimi & V. Baroncini. *Subjective evaluation of next-generation video compression algorithms : a case study*. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, volume 7798 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, August 2010.
- [de Simone *et al.* 11] F. de Simone, L. Goldmann, J.-S. Lee & T. Ebrahimi. *Performance analysis of VP8 image and video compression based on subjective evaluations*. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, volume 8135 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, September 2011.
- [Deguillaume *et al.* 99] F. Deguillaume, G. Csurka, J.J. O’Ruanaidh & T. Pun. *Robust 3D DFT video watermarking*. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, volume 3657, pages 113–124, 1999.
- [Dittmann *et al.* 98] J. Dittmann, M. Stabenau & R. Steinmetz. *Robust MPEG video watermarking technologies*. In Proceedings of the sixth ACM international conference on Multimedia, pages 71–80. ACM, 1998.
- [Do *et al.* 10] T.T. Do, Xiaoan Lu & J. Sole. *Compressive sensing with adaptive pixel domain reconstruction for block-based video coding*. In Image Processing (ICIP), 2010 17th IEEE International Conference on, pages 3377–3380, 2010.
- [Donoho 06] D.L. Donoho. *For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution*. Communications on pure and applied mathematics, vol. 59, no. 6, pages 797–829, 2006.
- [Dubreil Vall & Stojanovic 11] L. Dubreil Vall & M. Stojanovic. *Towards underwater video transmission*. Master’s thesis, Massachusetts Institute of Technology, 2011.
- [Duda & Hart 73] R.O. Duda & P.E. Hart. *Pattern classification and scene analysis*. A Wiley-Interscience Publication, New York : Wiley, vol. 1, 1973.

- [Dufaux *et al.* 09] F. Dufaux, W. Gao, S. Tubaro & A. Vetro. *Distributed Video Coding : Trends and Perspectives*. EURASIP J. Image and Video Processing, vol. 2009, 2009.
- [Dumitras & Haskell 04] A. Dumitras & B.G. Haskell. *An encoder-decoder texture replacement method with application to content-based movie coding*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 14, no. 6, pages 825 – 840, Juin 2004.
- [Ebrahimi 94] T. Ebrahimi. *A new technique for motion field segmentation and coding for very low bitrate video coding applications*. In Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, volume 2, pages 433 –437 vol.2, nov 1994.
- [Ehrgott 05] M. Ehrgott. Multicriteria optimization, volume 491. Springer Verlag, 2005.
- [Gan 07] Lu Gan. *Block Compressed Sensing of Natural Images*. In Digital Signal Processing, 2007 15th International Conference on, pages 403 –406, Juillet 2007.
- [Giannoula & Hatzinakos 04] A. Giannoula & D. Hatzinakos. *Integrating compression with watermarking on video sequences*. In Information Technology : Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on, volume 2, pages 159–161. IEEE, 2004.
- [Girod 94] B. Girod. *Rate-constrained motion estimation*. In Proc. of SPIE Conf. on Visual Communications and Image Processing (VCIP), volume 2, page 2308, Chicago, Illinois, USA, Septembre 1994.
- [Goyal 01] V.K. Goyal. *Multiple description coding : compression meets the network*. Signal Processing Magazine, IEEE, vol. 18, no. 5, pages 74 –93, sep 2001.
- [Guo *et al.* 10] Liwei Guo, Peng Yin, Yunfei Zheng, Xiaoan Lu, Qian Xu & J. Sole. *Adaptive motion vector resolution with implicit signaling*. In Image Processing (ICIP), 2010 17th IEEE International Conference on, pages 2057 –2060, sept. 2010.
- [Hartung & Girod 98] F. Hartung & B. Girod. *Watermarking of uncompressed and compressed video*. Signal processing, vol. 66, no. 3, pages 283–301, 1998.
- [Haupt & Nowak 06] J. Haupt & R. Nowak. *Signal reconstruction from noisy random projections*. Information Theory, IEEE Transactions on, vol. 52, no. 9, pages 4036–4048, 2006.

- [Joshi *et al.* 95] RL Joshi, TR Fischer & RH Bamberger. *Lossy encoding of motion vectors using entropy-constrained vector quantization*. In Image Processing, 1995. Proceedings., International Conference on, volume 3, pages 109–112. IEEE, October 1995.
- [Jung & Bross 11] J. Jung & B. Bross. *Description of Core Experiment CE13 : Motion data parsing robustness and throughput*, July 2011.
- [Jung & Laroche 06] J. Jung & G. Laroche. *Competition-Based Scheme for Motion Vector Selection and Coding*, VCEG-AC06, Juillet 2006.
- [Jung & Pateux 07] J. Jung & S. Pateux. *An Excel Add-in for Computing Bjontegaard Metric and Its Evolution*, January 2007.
- [Kamp *et al.* 08] S. Kamp, M. Evertz & M. Wien. *Decoder side motion vector derivation for inter frame video coding*. In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on, pages 1120–1123. IEEE, 2008.
- [Kamp *et al.* 09] S. Kamp, J. Ballé & M. Wien. *Multihypothesis prediction using decoder side motion vector derivation in inter frame video coding*. In Proc. SPIE Visual Communications and Image Processing VCIP, volume 9, San José, USA, January 2009.
- [Karczewicz & Lainema 02] M. Karczewicz & J. Lainema. *Analysis and Simplification of Intra Prediction*, Juillet 2002.
- [Karczewicz *et al.* 97] M. Karczewicz, J. Niewgowski & P. Haavisto. *Video coding using motion compensation with polynomial motion vector fields*. Signal Processing : Image Communication, vol. 10, no. 1-3, pages 63–91, 1997.
- [Kim & Ra 99] S.D. Kim & J.B. Ra. *An efficient motion vector coding scheme based on minimum bitrate prediction*. Image Processing, IEEE Transactions on, vol. 8, no. 8, pages 1117–1120, 1999.
- [Kim *et al.* 07] Sung Kim, Sang Kim, Youpyo Hong & Chee Won. *Data Hiding on H.264/AVC Compressed Video*. In Mohamed Kamel & Aurelio Campilho, éditeurs, Image Analysis and Recognition, volume 4633 of *Lecture Notes in Computer Science*, pages 698–707. Springer Berlin Heidelberg, 2007.
- [Kim *et al.* 08] Dae-Yeon Kim, Ki-Hun Han & Yung-Lyul Lee. *Adaptive intra mode bit skip in intra coding*. In Circuits and Systems, 2008. APCCAS 2008. IEEE Asia Pacific Conference on, pages 446–449, 30 2008-dec. 3 2008.

- [Krishnamurthy 97] R. Krishnamurthy. *Compactly-encoded optical flow fields for motion-compensated video compression and processing*. PhD thesis, Rensselaer Polytechnic Institute, 1997.
- [KTA] *KTA software coordination*. <http://iphome.hhi.de/suehring/tml/download/KTA/>.
- [Kuo et al. 08] T.Y. Kuo, Y.C. Lo & C.I. Lin. *Fragile Video Watermarking Technique by Motion Field Embedding with Rate-distortion Minimization*. In International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pages 853–856. IEEE, 2008.
- [Kutter et al. 97] M. Kutter, F. Jordan & T. Ebrahimi. *Proposal of a watermarking technique for hiding/retrieving data in compressed and decompressed video*. Technical report M2881, ISO/IEC document, JTC1/Sc29/WG11, 1997.
- [Lakshman et al. 10a] H. Lakshman, H. Schwarz & T. Wiegand. *Adaptive motion model selection using a cubic spline based estimation framework*. In Image Processing (ICIP), 2010 17th IEEE International Conference on, pages 805 –808, sept. 2010.
- [Lakshman et al. 10b] H. Lakshman, H. Schwarz & T. Wiegand. *Video coding with cubic spline interpolation and adaptive motion model selection*. In Signal Processing and Communications (SPCOM), 2010 International Conference on, pages 1 –5, july 2010.
- [Laroche et al. 08] G. Laroche, J. Jung & B. Pesquet-Popescu. *RD optimized coding for motion vector predictor selection*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 9, pages 1247–1257, Septembre 2008.
- [Laroche et al. 09] G. Laroche, J. Jung & B. Pesquet-Popescu. *Intra Prediction with 1D Macroblocck Partitioning for Image and Video Coding*. In Visual Communications and Image Processing, San Jose, CA, USA, Janvier 2009.
- [Laroche et al. 10] G. Laroche, J. Jung & B. Pesquet-Popescu. *Intra Coding With Prediction Mode Information Inference*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 20, no. 12, pages 1786 –1796, dec. 2010.
- [Laroche 09] G. Laroche. *Modules de codage par competition et suppression de l'information de competition pour le codage de sequences video*. PhD thesis, Telecom ParisTech, Mai 2009.
- [Le Tanou 10] J. Le Tanou. *Optimisation du coût de codage du mouvement, par une redistribution adaptative des résiduels de*

- mouvement. Master's thesis, Telecom ParisTech, Janvier 2010.
- [Lee & Woods 95] Y.Y. Lee & J.W. Woods. *Motion vector quantization for video coding*. IEEE Transactions on Image Processing,, vol. 4, no. 3, pages 378–382, 1995.
- [Lee et al. 09] Hoyoung Lee, Bongsoo Jung, Jooyoung Jung & Byeungwoo Jeon. *Computational complexity scalable scheme for power-aware H.264/AVC encoding*. In Multimedia Signal Processing, 2009. MMSP '09. IEEE International Workshop on, pages 1–6, Octobre 2009.
- [Li et al. 09] S. Li, L. Yu, J. Wang & L. Xiong. *Test results of second order prediction (SOP)*, Janvier 2009.
- [Li et al. 12] B. Li, G.J. Sullivan & J. Xu. *Comparison of Compression Performance of HEVC Working Draft 5 with AVC High Profile*, 2012.
- [Liu et al. 10] Zhaorui Liu, H.V. Zhao & A.Y. Elezzabi. *Block-based adaptive compressed sensing for video*. In Image Processing (ICIP), 2010 17th IEEE International Conference on, pages 1649–1652, sept. 2010.
- [Liu et al. 11] Z. Liu, A. Y. Elezzabi & H. Zhao. *Maximum Frame Rate Video Acquisition Using Adaptive Compressed Sensing*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. PP, no. 99, page 1, 2011.
- [Malvar et al. 03] H.S. Malvar, A. Hallapuro, M. Karczewicz & L. Kerofsky. *Low-complexity transform and quantization in H.264/AVC*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 13, no. 7, pages 598–603, Juillet 2003.
- [Marpe et al. 03] D. Marpe, H. Schwarz & T. Wiegand. *Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 13, no. 7, pages 620–636, Juillet 2003.
- [Marpe et al. 10] D. Marpe, H. Schwarz, S. Bosse, B. Bross, P. Helle, T. Hinz, H. Kirchhoffer, H. Lakshman, T. Nguyen, S. Oudinet et al. . *Video Compression Using Nested Quadtree Structures, Leaf Merging, and Improved Techniques for Motion Representation and Entropy Coding*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 12, pages 1676–1687, 2010.

- [Martin *et al.* 07] A. Martin, J.J. Fuchs, C. Guillemot & D. Thoreau. *Sparse representation for image prediction*. In Proc. European Signal Processing Conference (EUSIPCO), Poznan, Poland. Citeseer, 2007.
- [Martin *et al.* 08] Aurelie Martin, Jean-Jacques Fuchs, Christine Guillemot & Dominique Thoreau. *Atomic decomposition dedicated to AVC and spatial SVC prediction*. In ICIP, pages 2492–2495, 2008.
- [Marvel *et al.* 00] L.M. Marvel, G.W. Hartwig Jr & C. Boncelet Jr. *Compression-compatible fragile and semi-fragile tamper detection*. In PROC SPIE INT SOC OPT ENG, volume 3971, pages 131–139, 2000.
- [Masmoudi *et al.* 11] K. Masmoudi, M. Antonini & P. Kornprobst. *A Biologically Inspired Image Coder with Temporal Scalability*. In Proc. of Advanced Concepts for Intelligent Vision Systems (ACIVS 2011), Ghent, Belgique, 2011.
- [Masmoudi *et al.* 12] K. Masmoudi, M. Antonini & P. Kornprobst. *Frames for Exact Inversion of the Rank Order Coder*. Neural Networks and Learning Systems, IEEE Transactions on, vol. 23, no. 2, pages 353–359, Février 2012.
- [Matsuo *et al.* 09] S. Matsuo, S. Takamura & Y. Yashima. *Intra prediction with spatial gradients and multiple reference lines*. In Picture Coding Symposium, 2009. PCS 2009, pages 1–4, Mai 2009.
- [Maugey 10] Thomas Maugey. *Codage vidéo distribué de séquences multivues*. These, Télécom ParisTech, November 2010.
- [Merkle *et al.* 07] P. Merkle, A. Smolic, K. Muller & T. Wiegand. *Efficient prediction structures for multiview video coding*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, pages 1461–1473, Novembre 2007.
- [Moinard *et al.* 10] M. Moinard, I. Amonou, P. Duhamel & P. Brault. *A set of template matching predictors for intra video coding*. In Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, pages 1422–1425, march 2010.
- [Moinard 11] M. Moinard. *Codage video hybride base contenu par analyse/synthese de donnees*. PhD thesis, Telecom ParisTech, 2011.
- [mpe 94] *Generic coding of moving pictures and associated audio information*, 1994.

- [Mun & Fowler 09] S. Mun & J.E. Fowler. *Block compressed sensing of images using directional transforms*. In Image Processing (ICIP), 2009 16th IEEE International Conference on, pages 3021–3024. IEEE, 2009.
- [Mun & Fowler 11] S. Mun & J.E. Fowler. *Residual Reconstruction for Block-Based Compressed Sensing of Video*. In Data Compression Conference, 2011.
- [Nam *et al.* 10] Hyeong-Min Nam, Jae-Yun Jeong, Keun-Yung Byun, Jong-Ok Kim & Sung-Jea Ko. *A complexity scalable H.264 decoder with downsizing capability for mobile devices*. Consumer Electronics, IEEE Transactions on, vol. 56, no. 2, pages 1025–1033, Mai 2010.
- [Natale & Granelli 01] F.G.B. De Natale & F. Granelli. *Low-Complexity Context-Based Motion Compensation for VLBR Video Encoding*. Multimedia and Expo, IEEE International Conference on, vol. 0, page 113, 2001.
- [Negusse 08] S.A. Negusse. Improving intra pixel prediction for h. 264 video coding. Master’s thesis, Blekinge Institute of Technology, May 2008.
- [Nemethova *et al.* 06] O. Nemethova, G.C. Forte & M. Rupp. *Robust error detection for H. 264/AVC using relation based fragile watermarking*. Citeseer, 2006.
- [Nyquist 24] H. Nyquist. *Certain Factors Affecting Telegraph Speed*. American Institute of Electrical Engineers, Transactions of the, vol. XLIII, pages 412–422, jan. 1924.
- [Park & Wakin 09] J.Y. Park & M.B. Wakin. *A multiscale framework for compressive sensing of video*. In Picture Coding Symposium, 2009. PCS 2009, pages 1–4. IEEE, 2009.
- [Park *et al.* 07] J.-Y. Park, S.W. Park & B.M. Jeon. *Intra prediction with subpel samples*, Octobre 2007.
- [Parzen 62] E. Parzen. *On estimation of a probability density function and mode*. The annals of mathematical statistics, vol. 33, no. 3, pages 1065–1076, 1962.
- [Puri & Ramchandran 02] R. Puri & K. Ramchandran. *PRISM : a new robust video coding architecture based on distributed compression principles*. In Proc. of Allerton Conf. Commun., Ctrl, Computing, Urbana-Champaign, IL, Oct. 2002, 2002.
- [Raad Cisa 11] L. Raad Cisa. Etude de l’approche data hiding pour la compression vidéo. Master’s thesis, Telecom Bretagne, 2011.

- [Rao *et al.* 07] K. R. Rao, Patrick Yip & Vladimir Britanak. Discrete cosine transform : Algorithms, advantages, applications. Academic Press, Inc., Orlando, FL, USA, 2007.
- [Shade *et al.* 98] J. Shade, S. Gortler, L. He & R. Szeliski. *Layered depth images*. In Proceedings of the 25th annual conference on Computer graphics and interactive techniques, pages 231–242, New York, USA, 1998. ACM.
- [Shannon 01] C.E. Shannon. *A mathematical theory of communication*. ACM SIGMOBILE Mobile Computing and Communications Review, vol. 5, no. 1, pages 3–55, 2001.
- [Shimizu *et al.* 11] S. Shimizu, H. Kimata, S. Sugimoto & N. Matsuura. *Decoder-side macroblock information derivation for efficient multiview video plus depth map coding*. In 3DTV Conference : The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2011, pages 1–4, may 2011.
- [Shukla *et al.* 05] R. Shukla, P.L. Dragotti, M.N. Do & M. Vetterli. *Rate-distortion optimized tree-structured compression algorithms for piecewise polynomial images*. Image Processing, IEEE Transactions on, vol. 14, no. 3, pages 343–359, 2005.
- [Slepian & Wolf 73] D. Slepian & J.K. Wolf. *Noiseless Coding of Correlated Information Sources*. IEEE Transactions on Information Theory, vol. 19, no. 4, pages 471–480, 1973.
- [Smolic *et al.* 07] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G.B. Akar, G. Triantafyllidis & A. Koz. *Coding Algorithms for 3DTV-A Survey*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 17, no. 11, pages 1606–1621, nov. 2007.
- [Sole & Salembier 07] J. Sole & P. Salembier. *Generalized Lifting Prediction Optimization Applied to Lossless Image Compression*. Signal Processing Letters, IEEE, vol. 14, no. 10, pages 695–698, Octobre 2007.
- [Stankovic *et al.* 08] V. Stankovic, L. Stankovic & S. Cheng. *Compressive video sampling*. In Proc. Eusipco-2008 16th European Signal Processing Conference, Lausanne, Switzerland. Citeseer, 2008.
- [Su *et al.* 08] P.C. Su, M.L. Li & I.F. Chen. *A Content-Adaptive Digital Watermarking Scheme in H. 264/AVC Compressed Videos*. In Proceedings of the 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pages 849–852. IEEE Computer Society, 2008.

- [Sugimoto *et al.* 04] K. Sugimoto, M. Kobayashi, Y. Suzuki, S. Kato & C.S. Boon. *Inter frame coding with template matching spatio-temporal prediction*. In Image Processing, 2004. ICIP'04. 2004 International Conference on, volume 1, pages 465–468. IEEE, 2004.
- [Swanson *et al.* 97] M.D. Swanson, B. Zhu & A.H. Tewfik. *Data hiding for video-in-video*. In icip, page 676. Published by the IEEE Computer Society, 1997.
- [Swanson *et al.* 98] M.D. Swanson, B. Zhu & A.H. Tewfik. *Multiresolution scene-based video watermarking using perceptual models*. Selected Areas in Communications, IEEE Journal on, vol. 16, no. 4, pages 540–550, 1998.
- [T. K. Tan & Wedi 07] G. J. Sullivan T. K. Tan & T. Wedi. *Recommended Simulation Common Conditions for Coding Efficiency Experiments*, Janvier 2007.
- [Takahashi & Yamaguchi 09] M. Takahashi & M. Yamaguchi. *H. 264/AVC improvement based on adaptive motion vector coding*. In Picture Coding Symposium (PCS), pages 1–4, USA, 2009. IEEE.
- [Tan *et al.* 06] Thiow Keng Tan, Choong Seng Boon & Yoshinori Suzuki. *Intra Prediction by Template Matching*. In Image Processing, 2006 IEEE International Conference on, pages 1693–1696, oct. 2006.
- [Tan *et al.* 07] Thiow Keng Tan, Choong Seng Boon & Yoshinori Suzuki. *Intra Prediction by Averaged Template Matching Predictors*. In Consumer Communications and Networking Conference, 2007. CCNC 2007. 4th IEEE, pages 405–409, jan. 2007.
- [Thoreau *et al.* 10] Dominique Thoreau, Aurelie Martin, Edouard François, Jérôme Vieron & Quan Huynh-Thu. *Sparse shift-DCT spatial prediction*. In ICIP, pages 3385–3388, 2010.
- [Tramel & Fowler 11] E.W. Tramel & J.E. Fowler. *Video Compressed Sensing with Multihypothesis*. In Data Compression Conference, 2011.
- [Turkan & Guillemot 10] M. Turkan & C. Guillemot. *Image prediction : Template matching vs. Sparse approximation*. In Proc. IEEE Int. Conf. on Image Process. (ICIP), Hong Kong, China, sep 2010. best student paper award.
- [Urvoy 11] M. Urvoy. *Les tubes de mouvement : nouvelle représentation pour les séquences d images*. PhD thesis, INSA Rennes, 2011.

- [Van Rullen & Thorpe 01] R Van Rullen & S J Thorpe. *Rate coding versus temporal order coding : what the retinal ganglion cells tell the visual cortex*. Neural Computation, vol. 13, no. 6, pages 1255–1283, 2001.
- [Vatis & Ostermann 06] Y. Vatis & J. Ostermann. *Prediction of P- and B-Frames Using a Two-Dimensional Non-separable Adaptive Wiener Interpolation Filter for H.264/AVC*, Octobre 2006.
- [Wakin et al. 06] M. Wakin, J.N. Laska, M.F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K.F. Kelly & R.G. Baraniuk. *Compressive imaging for video representation and coding*. In Picture Coding Symposium. Citeseer, April 2006.
- [Wedi 00] T. Wedi. *1/8-pel motion vector resolution for H.26L*, Août 2000.
- [Wiegand et al. 98] T. Wiegand, E. Steinbach, A. Stensrud & B. Girod. *Multiple-reference-picture video coding using polynomial motion models*. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, volume 3309, pages 134–145, 1998.
- [Won et al. 09] K. Won, J. Yang & B. Jeon. *Motion vector coding using decoder-side estimation of motion vector*. In IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pages 1–4. IEEE, 2009.
- [Wyner & Ziv 76] A. Wyner & J. Ziv. *The rate-distortion function for source coding with side information at the decoder*. Information Theory, IEEE Transactions on, vol. 22, no. 1, pages 1–10, 1976.
- [Xin et al. 07] J. Xin, K. N. Ngan & G. Zhu. *Combined Inter-Intra prediction for high definition video coding*. In Picture Coding Symposium (PCS), Lisbonne, Portugal, Novembre 2007.
- [Yang & Zhang 09] J. Yang & Y. Zhang. *Alternating Direction Algorithms for l1-Problems in Compressive Sensing*. Arxiv preprint arXiv :0912.1185, 2009.
- [Yann 04] Bodo Yann. *Elaboration d'une technique d'accès conditionnel par tatouage et embrouillage video basee sur la perturbation des vecteurs de mouvement*. PhD thesis, Ecole Nationale Supérieur des Telecommunications, 2004.
- [Yea & Vetro 08] Sehoon Yea & A. Vetro. *View Synthesis Prediction for Rate-Overhead Reduction in FTV*. In 3DTV Conference : The True Vision - Capture, Transmission and Display of 3D Video, 2008, pages 145 –148, may 2008.

- [Yeh *et al.* 95] J. Yeh, M. Vetterli & M. Khansari. *Motion compensation of motion vectors*. Image Processing, International Conference on, vol. 1, page 574, 1995.
- [Yoon *et al.* 07] S.U. Yoon, E.K. Lee, S.Y. Kim & Y.S. Ho. *A framework for representation and processing of multi-view video using the concept of layered depth image*. The Journal of VLSI Signal Processing, vol. 46, no. 2, pages 87–102, 2007.
- [Yu *et al.* 12] X. Yu, J. Wang, D. He, G. Martin-Cocher & S. Campbell. *Multiple Sign Bits Hiding*, Février 2012.
- [Zhang *et al.* 08] Yifu Zhang, Shunliang Mei, Quqing Chen & Zhibo Chen. *A novel image/video coding method based on Compressed Sensing theory*. In Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, pages 1361 –1364, 312008-april4 2008.
- [Zhu & Tewfik 97] B. Zhu & A.H. Tewfik. *Media compression via data hiding*. In Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on, volume 1, pages 647–651. IEEE, 1997.

Publications de l'auteur

Articles de revue internationale

- [Revue1] J.-M. Thiesse, J. Jung & M. Antonini. *Rate Distortion Data Hiding of Motion Vector Competition Information in Chroma and Luma Samples for Video Compression*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 6, pages 729 –741, Juin 2011.
- [Revue2] J.-M. Thiesse, J. Le Tanou, J. Jung & M. Antonini. *Motion Vector Coding with Forecast and Mapping Technique*. IEEE Transactions on Circuits and Systems for Video Technology, *Soumis*.
- [Revue3] J.-M. Thiesse, J. Jung & M. Antonini. *Compressed Sensing for Video Coding based on Smart Decoder for Optimal Derivation of Side Parameters*. IEEE Transactions on Circuits and Systems for Video Technology, *Soumis*.

Articles de conférences internationales

- [Conf1] J.M. Thiesse, J. Jung & M. Antonini. *Hybrid-1D macroblock prediction for video compression*. In Eusipco, 2009.
- [Conf2] J.-M. Thiesse, J. Jung & M. Antonini. *Data hiding of intra prediction information in chroma samples for video compression*. In Image Processing (ICIP), 2010 17th IEEE International Conference on, pages 2861 –2864, sept. 2010.
- [Conf3] J.-M. Thiesse, J. Jung & M. Antonini. *Data hiding of motion information in chroma and luma samples for video compression*. In Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on, pages 217 –221, Octobre 2010.
- [Conf4] J. Le Tanou, J.-M. Thiesse, J. Jung & M. Antonini. *Motion vector forecast and mapping (MV-FMap) method for entropy coding based video coders*. In Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on, pages 206 –211, Octobre 2010.

Articles de conférences nationales

- [ConfFR1] J.M. Thiesse , G. Laroche, et J. Jung. *Prédiction Intra par Partitionnement 1D pour le codage d'Images et de Vidéos*. Proc. Compression et Représentation des Signaux Audiovisuels (CORESA), Toulouse, France, Mars 2009. (*sélectionné au prix jeune chercheur*)
- [ConfFR2] J.-M. Thiesse, J. Jung & M. Antonini. *Compression Vidéo avec Marquage d'Index de Mode Intra dans la Chroma*. Proc. COmpression et Représentation des Signaux Audiovisuels (CORESA), Lyon, France, Octobre 2010. (*second au prix jeune chercheur*)

Brevets

- [Brevet1] J. Jung, G. Laroche et J.-M. Thiesse. Calcul de prédicteurs de vecteurs mouvement. Dépôt FR : Juillet 2009.
- [Brevet2] J. Jung, G. Laroche et J.-M. Thiesse. Calcul de prédicteurs temporels de vecteurs mouvement. Dépôt FR : Juillet 2009.
- [Brevet3] J. Jung et J.-M. Thiesse. Procédé de codage d'index de compétition par signalisation dans des données résiduels. Dépôt FR : Octobre 2010.
- [Brevet4] J. Jung, J. Le Tanou et J.-M. Thiesse. Procédé de traitement d'une information de mouvement, procédés de codage et de décodage, dispositifs, signal et programme d'ordinateur correspondants. Dépôt FR : Avril 2010.
- [Brevet5] J. Jung, J. Le Tanou et J.-M. Thiesse. Procédé d'enrichissement d'une information de mouvement, procédé de codage, dispositifs et programme d'ordinateur correspondants . Dépôt FR : Avril 2010.
- [Brevet6] J. Jung et J.-M. Thiesse. Procédé et dispositifs de codage et de décodage d'au moins une image mettant en oeuvre une sélection des pixels à prédire, programme d'ordinateur correspondant. Dépôt FR : Décembre 2010.
- [Brevet7] J. Jung, S. Pateux et J.-M. Thiesse. Procédés de codage et de décodage d'une valeur n-aire, dispositifs et programme d'ordinateur correspondants. Dépôt FR : Décembre 2010.
- [Brevet8] J. Jung, M. Antonini et J.-M. Thiesse. Procédé de codage par héritage des paramètres de codage à partir de zones causales. Dépôt FR : Juillet 2011. (*non divulgué à ce jour*)
- [Brevet9] J. Jung et J.-M. Thiesse. Calcul de prédicteurs temporels de vecteurs mouvement. Dépôt FR : Septembre 2011. (*non divulgué à ce jour*)



Résumé

Cette thèse est dédiée à l'amélioration des performances de compression vidéo. Deux types d'approches, conventionnelle et en rupture, sont explorées afin de proposer des méthodes efficaces de codage Intra et Inter pour les futurs standards de compression.

Deux outils sont étudiés pour la première approche. Tout d'abord, des indices de signalisations sont habilement traités par une technique issue du tatouage permettant de les masquer dans les résiduels de luminance et de chrominance de façon optimale selon le compromis débit-distorsion. La forte redondance dans le mouvement est ensuite exploitée pour améliorer le codage des vecteurs de mouvement. Après observation des précédents vecteurs utilisés, un fin pronostic permet de déterminer les vecteurs résiduels à privilégier lors d'une troisième étape de modification de la distribution des résiduels. 90% des vecteurs codés sont ainsi pronostiqués, ce qui permet une nette réduction de leur coût.

L'approche en rupture vient de la constatation que H.264/AVC et son successeur HEVC sont basés sur un schéma prédictif multipliant les choix de codage, les améliorations passent alors par un meilleur codage de la texture à l'aide d'une compétition accrue. De tels schémas étant bornés par la signalisation engendrée, il est alors nécessaire de transférer des décisions au niveau du décodeur. Une approche basée sur la détermination conjointe au codeur et au décodeur de paramètres de codage à l'aide de partitions causales et ainsi proposée et appliquée aux modes de prédiction Intra et à la théorie émergente de l'échantillonnage compressé. Des performances encourageantes sont reportées et confirment l'intérêt d'une telle solution innovante.

Mots clés : Compression vidéo, H.264/AVC, HEVC, Codage Intra, Codage Inter, Tatouage, Vecteurs de mouvement, Décodeur intelligent, Echantillonnage compressé.

Abstract

This Ph.D. thesis deals with the improvement of video compression efficiency. Both conventional and breakthrough approaches are investigated in order to propose efficient methods for Intra and Inter coding dedicated to next generations video coding standards.

Two tools are studied for the conventional approach. First, syntax elements are cleverly transmitted using a data hiding based method which allows embedding indices into the luminance and chrominance residuals in an optimal way, rate-distortion wise. Secondly, the large motion redundancies are exploited to improve the motion vectors coding. After a statistical analysis of the previously used vectors, an accurate forecast is performed to favour some vector residuals during a last step which modifies the original residual distribution. 90% of the coded vectors are efficiently forecasted by this method which helps to significantly reduce their coding cost.

The breakthrough approach comes from the observation of the H.264/AVC standard and its successor HEVC which are based on a predictive scheme with multiple coding choices, consequently future improvements shall improve texture by extensively using the competition between many coding modes. However, such schemes are bounded by the cost generated by the signalling flags and therefore it is required to transfer some decisions to the decoder side. A framework based on the determination of encoding parameters at both encoder and decoder side is consequently proposed and applied to Intra prediction modes on the one hand, and to the emerging theory of compressed sensing on the other hand. Promising results are reported and confirm the potential of such an innovative solution.

Keywords : Video compression, H.264/AVC, HEVC, Intra coding, Inter coding, Data Hiding, Motion vector, Smart decoder, Compressed sensing.