



**HAL**  
open science

# Modules de codage par compétition et suppression de l'information de compétition pour le codage de séquences vidéo

Guillaume Laroche

► **To cite this version:**

Guillaume Laroche. Modules de codage par compétition et suppression de l'information de compétition pour le codage de séquences vidéo. domain\_other. Télécom ParisTech, 2009. English. NNT : . pastel-00005379

**HAL Id: pastel-00005379**

**<https://pastel.hal.science/pastel-00005379>**

Submitted on 19 May 2010

**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.



# Thèse

présentée pour obtenir le grade de docteur  
de l'Ecole Nationale Supérieure des Télécommunications

Spécialité : **Signal et Images**

**Guillaume LAROCHE**

Modules de codage par compétition et suppression  
de l'information de compétition pour le codage de  
séquences vidéo

Soutenue le 19 mai 2009 devant le jury composé de :

Benoit Macq  
Marc Antonini  
Philippe Salembier  
Patrice Onno  
Béatrice Pesquet-Popescu  
Joël Jung

Président  
Rapporteurs  
Examineur  
Directeurs de thèse



# Remerciements

Je remercie Joël Jung, Ingénieur de recherche à Orange Labs, d'avoir assuré la direction et l'orientation de mes travaux de recherche. Je le remercie pour sa patience, sa rigueur, le partage de son expérience dans de nombreux domaines dépassant celui de la recherche, et aussi pour l'énergie qu'il a mise en œuvre pour valoriser nos travaux sur le plan international et enfin pour toutes les collaborations dont il m'a fait bénéficier.

Je remercie également Béatrice Pesquet-Popescu, Professeure à l'Ecole Nationale Supérieure des Télécommunications de Paris, d'avoir dirigé ces travaux de thèse durant ces trois années, pour les conseils toujours pertinents et avisés, mais aussi pour la transmission d'une partie de sa vaste connaissance du domaine des télécommunications.

Je remercie Benoit Macq, Professeur et Vice-Recteur de l'Université Catholique de Louvain, d'avoir accepté de présider le jury de cette thèse.

Je remercie également Philippe Salembier, Professeur à l'Université Polytechnique de Catalogne et Marc Antonini, Directeur de recherche au Laboratoire Informatique des Signaux et Systèmes de Sophia Antipolis (I3S) du CNRS, d'avoir accepté la charge de rapporteur.

Je remercie également Patrice Onno, Ingénieur de recherche à Canon Research Center France, d'avoir bien voulu juger ce travail.

Je souhaite remercier sincèrement Marco Cagnazzo, Maître de Conférences à l'Ecole Nationale Supérieure des Télécommunications de Paris, d'avoir partagé son professionnalisme et sa rigueur scientifique durant nos deux collaborations. Je remercie aussi Nicolas Loménie, Maître de Conférences de l'Université Paris 5, de m'avoir fait confiance pour assurer le cours de compression vidéo au sein du master recherche IPCC.

Je tiens à remercier Jean-Claude Schmitt de m'avoir accueilli, pour mon stage de Master et pour mon Doctorat au sein du Service des Images Animées. Un grand merci à toute cette équipe, Erwan, Francis, Guillaume, Jérôme, Marc, Olivier, les trois Patrick, Stéphane, pour leur accueil, les débats animés du déjeuner, en particulier Patrick et Olivier pour leurs soutiens matériels et Thomas pour toutes les pauses partagées, sans oublier Chantal pour son efficacité et sa gentillesse. Je remercie

aussi tous les stagiaires avec qui j'ai eu l'occasion de travailler, Etienne, Ting, Diane, Hassna, Joselin et Jean-Marc.

Enfin un grand merci à Emmanuelle, à mes parents et à mon frère pour leur soutien et leurs encouragements pendant ce doctorat. Je remercie aussi tous mes amis et mes proches qui ont contribué au bon déroulement de ma thèse.

# Table des matières

<b>Remerciements</b>	<b>i</b>
<b>Liste des tableaux</b>	<b>viii</b>
<b>Liste des figures</b>	<b>xv</b>
<b>Glossaire</b>	<b>xxi</b>
<b>Introduction</b>	<b>1</b>
<b>I Codage vidéo : un état de l'art</b>	<b>7</b>
<b>1 Le codage vidéo et la norme H.264/AVC</b>	<b>9</b>
1.1 Description du schéma global de codage de la norme H.264/AVC . . . . .	9
1.1.1 Calcul du résiduel . . . . .	10
1.1.2 Codage du résiduel . . . . .	12
1.1.2.1 Transformation . . . . .	12
1.1.2.2 Quantification . . . . .	13
1.1.2.3 Codage entropique et arithmétique . . . . .	14
1.2 Normes de compression en codage d'images fixes . . . . .	14
1.2.1 Codage par prédiction . . . . .	15
1.2.1.1 Partitionnement . . . . .	15
1.2.1.2 Ordre de parcours des partitions . . . . .	16
1.2.1.3 Prédicteurs Intra . . . . .	17
1.2.1.4 Transformée . . . . .	18
1.2.2 Les standards de codage d'images . . . . .	19
1.3 Etat de l'art sur le codage de l'information de mouvement . . . . .	22
1.3.1 Représentation du mouvement . . . . .	22
1.3.2 Codage des vecteurs mouvement avec pertes . . . . .	23
1.3.3 Codage prédictif des vecteurs mouvement . . . . .	23
1.3.4 Codage de l'information de mouvement dans H.264/AVC . . . . .	24
1.3.4.1 Le prédicteur médian . . . . .	24

1.3.4.2	L'information de mouvement pour le mode Skip . . . . .	25
1.3.4.3	L'information de mouvement pour le mode Direct . . . . .	25
1.4	H.264/AVC et après ? . . . . .	26
1.5	Résumé du chapitre . . . . .	29
<b>2</b>	<b>Méthodes de codage par compétition</b>	<b>31</b>
2.1	Définition . . . . .	31
2.2	Sélection du meilleur codage . . . . .	33
2.2.1	Critères exhaustifs . . . . .	33
2.2.2	Critères "a priori" . . . . .	35
2.3	Possibilités de codage par compétition . . . . .	36
2.3.1	Compétition de codeurs . . . . .	36
2.3.2	Compétition de prédictions . . . . .	37
2.3.3	Compétition de transformées . . . . .	37
2.3.4	Autres compétitions . . . . .	38
2.4	Ajout d'une nouvelle possibilité de codage . . . . .	38
2.5	Conclusion . . . . .	39
<b>3</b>	<b>Etude comportementale du codec de référence de la norme H.264/AVC</b>	<b>41</b>
3.1	Analyse des slices I, P et B . . . . .	41
3.1.1	Efficacité de codage . . . . .	42
3.1.2	Proportion des images I P B dans un flux . . . . .	43
3.1.3	Représentation visuelle des coûts . . . . .	45
3.2	Répartition des informations de codage dans le flux . . . . .	47
3.3	Sélection des modes de codage . . . . .	50
3.3.1	Slices Inter . . . . .	50
3.3.2	Slices Intra . . . . .	52
3.4	H.264/AVC et JPEG 2000 . . . . .	54
3.5	Conclusion et résumé du chapitre . . . . .	55
<b>II</b>	<b>Modules de codage par compétition</b>	<b>57</b>
<b>4</b>	<b>Codage de l'information de mouvement par compétition de prédicteurs</b>	<b>59</b>
4.1	Prédicteurs . . . . .	59
4.1.1	Prédicteurs Spatiaux . . . . .	60
4.1.2	Prédicteurs Temporels . . . . .	60
4.1.3	Prédicteur spatio-temporel . . . . .	61
4.1.4	Adaptation des prédicteurs pour les images de référence multiple . . . . .	62
4.1.5	Adaptation des prédicteurs pour les slices bidirectionnels . . . . .	63
4.1.6	Adaptation des prédicteurs pour les slices B hiérarchiques . . . . .	65
4.2	Compétition exhaustive . . . . .	66
4.2.1	Sélection des prédicteurs de vecteurs mouvement des modes Inter . . . . .	66

4.2.2	Sélection des prédicteurs du mode Skip . . . . .	67
4.2.3	Codage de l'information de compétition . . . . .	67
4.2.4	Complexité . . . . .	68
4.2.4.1	Complexité en mémoire . . . . .	68
4.2.4.2	Complexité de calcul . . . . .	69
4.3	Résultats expérimentaux . . . . .	70
4.3.1	Sélection des ensembles de prédicteurs . . . . .	70
4.3.1.1	Efficacité de codage des prédicteurs proposés . . . . .	70
4.3.1.2	Ensembles $\mathcal{P}$ et $\mathcal{P}_s$ des slices P . . . . .	71
4.3.1.3	Ensemble $\mathcal{P}$ des slices B . . . . .	72
4.3.2	Réduction de débit et analyses . . . . .	74
4.3.2.1	Configurations IPP . . . . .	74
4.3.2.2	Configuration High IBBP . . . . .	80
4.3.2.3	Configuration High BH . . . . .	82
4.3.3	Résultats MVComp dans la JM KTA . . . . .	84
4.4	Conclusion du chapitre . . . . .	85
<b>5</b>	<b>Compétition a priori pour le mode Skip</b>	<b>87</b>
5.1	Description de la méthode . . . . .	87
5.1.1	Adaptation d'un ensemble de prédicteurs . . . . .	88
5.1.2	Classification des macroblocs . . . . .	89
5.2	Résultats expérimentaux . . . . .	90
5.2.1	Augmentation de l'occurrence du mode Skip . . . . .	90
5.2.2	Analyse des ensembles $\mathcal{P}_s^M$ et $\mathcal{P}_s^S$ . . . . .	90
5.2.3	Analyse de la classification des macroblocs . . . . .	94
5.2.4	Gains sur le débit total . . . . .	95
5.3	Conclusion . . . . .	97
<b>6</b>	<b>Intra 1D</b>	<b>99</b>
6.1	Description de la méthode . . . . .	99
6.1.1	Partitionnement 1D . . . . .	100
6.1.2	Ordre de parcours des partitions 1D . . . . .	101
6.1.3	Prédicteurs . . . . .	102
6.1.3.1	Ordre de parcours lexicographique . . . . .	102
6.1.3.2	Ordre de parcours bidirectionnel . . . . .	103
6.1.3.3	Ordre de parcours hiérarchique . . . . .	104
6.1.4	Transformée DCT 1D . . . . .	105
6.2	Résultats expérimentaux . . . . .	106
6.2.1	Conditions de tests . . . . .	106
6.2.1.1	Partitionnements 1D actifs . . . . .	106
6.2.1.2	Ordre de parcours actifs . . . . .	106
6.2.1.3	Prédicteurs actifs . . . . .	107

6.2.2	Analyses de l'efficacité des ordres de parcours . . . . .	109
6.2.3	Configuration Intra . . . . .	111
6.2.3.1	Gain sur le débit total . . . . .	111
6.2.3.2	Sélection des Modes Intra . . . . .	111
6.2.3.3	Sélection des partitions 1D . . . . .	114
6.2.3.4	Sélection des prédicteurs Intra . . . . .	115
6.2.3.5	Sélection spatiale des modes Intra 1D . . . . .	116
6.2.3.6	Coût du codage de la compétition Intra 1D . . . . .	117
6.2.3.7	Intra 1D par rapport à JPEG 2000 . . . . .	120
6.2.4	Configuration Broadcast TV . . . . .	121
6.3	Conclusion et perspectives . . . . .	122
<b>III Que doit-on vraiment transmettre ?</b>		<b>125</b>
<b>7</b>	<b>Suppression d'indices de prédictions pour la méthode MVComp</b>	<b>127</b>
7.1	Description de la méthode . . . . .	127
7.1.1	Deux prédicteurs . . . . .	128
7.1.2	$N$ prédicteurs . . . . .	128
7.1.3	Remarques et discussions . . . . .	129
7.2	Résultats expérimentaux . . . . .	129
7.2.1	Implémentation . . . . .	129
7.2.2	Réduction de débit des indices de prédicteurs . . . . .	130
7.2.3	Analyse . . . . .	130
7.2.4	Elimination et fonctions de choix . . . . .	132
7.3	Conclusion . . . . .	133
<b>8</b>	<b>Suppression de prédicteurs pour le codage Intra</b>	<b>137</b>
8.1	Description de la méthode . . . . .	137
8.1.1	Critère idéal . . . . .	138
8.1.2	Critère d'égalité des résiduels reconstruits . . . . .	138
8.1.3	Approximation du critère optimal . . . . .	140
8.2	Résultats expérimentaux . . . . .	142
8.2.1	Implémentation de l'algorithme . . . . .	142
8.2.2	Sélection de $QP_{opt}$ . . . . .	142
8.2.3	Comparaison des critères . . . . .	144
8.2.4	Application à H.264/AVC . . . . .	145
8.2.4.1	Gain sur le débit total . . . . .	145
8.2.4.2	Analyses . . . . .	145
8.2.5	Application à l'Intra 1D . . . . .	148
8.2.5.1	Gain sur le débit total . . . . .	148
8.2.5.2	Analyses . . . . .	149

8.3	Extension de la méthode . . . . .	149
8.4	Conclusion et perspectives . . . . .	153
<b>9</b>	<b>Estimation de mouvement conjointe à l'encodeur et au décodeur</b>	<b>155</b>
9.1	Description de la méthode . . . . .	155
9.1.1	Critère d'appariement . . . . .	156
9.1.2	Partitionnements 1D . . . . .	158
9.1.2.1	Inter 1D . . . . .	160
9.1.2.2	Inter 1D sans vecteur . . . . .	160
9.1.3	Sélection des modes sans vecteur . . . . .	160
9.2	Résultats expérimentaux . . . . .	161
9.2.1	Implémentation . . . . .	161
9.2.2	Partitionnements H.264/AVC . . . . .	162
9.2.2.1	EMD seule . . . . .	163
9.2.2.2	Compétition des estimations de mouvement . . . . .	163
9.2.3	Inter 1D . . . . .	166
9.2.3.1	Estimation de mouvement classique . . . . .	167
9.2.3.2	Inter 1D et EMD . . . . .	169
9.2.4	Inter bloc et Inter 1D . . . . .	170
9.2.5	MVComp et EMD . . . . .	172
9.3	Conclusion et perspectives . . . . .	175
<b>IV</b>	<b>Résultats globaux</b>	<b>177</b>
<b>10</b>	<b>Intra 1D, MVComp et suppression de prédicteurs</b>	<b>179</b>
10.1	Configurations de tests . . . . .	179
10.2	Gains séquence par séquence . . . . .	180
10.3	Gains par résolution . . . . .	180
10.4	Efficacité de codage par slices . . . . .	182
10.5	Courbe de débit-distorsion . . . . .	183
10.6	Conclusion . . . . .	183
	<b>Conclusion</b>	<b>187</b>
<b>A</b>	<b>Configuration de Test</b>	<b>193</b>
A.1	Conditions VCEG . . . . .	193
A.1.1	Baseline . . . . .	193
A.1.2	High . . . . .	194
A.1.3	Liste des séquences . . . . .	194

<b>B</b>	<b>Sélection des prédicteurs de l’Intra 1D</b>	<b>199</b>
B.1	Partitionnement Ligne . . . . .	199
B.1.1	Ordre lexicographique . . . . .	199
B.1.2	Ordre bidirectionnel . . . . .	200
B.1.3	Ordre hiérarchique . . . . .	200
B.1.4	Deux ensembles, ordre lexicographique . . . . .	201
B.2	Partitionnement Colonne . . . . .	202
B.2.1	Ordre lexicographique . . . . .	202
B.2.2	Ordre bidirectionnel . . . . .	202
B.2.3	Ordre hiérarchique . . . . .	202
B.2.4	Deux ensembles, ordre lexicographique . . . . .	202
<b>C</b>	<b>Autres méthodes et perspectives étudiées</b>	<b>205</b>
C.1	Demi images de référence . . . . .	205
C.1.1	Implémentation . . . . .	206
C.1.2	Résultats . . . . .	206
C.1.3	Conclusion . . . . .	207
C.2	Estimation de mouvement récursive conjointe à l’encodeur et au décodeur . . . . .	207
C.2.1	Estimation de mouvement récursive de type Cafforio-Rocca . . . . .	209
C.2.2	Amélioration du prédicteur de bloc . . . . .	209
C.2.3	Second résiduel de texture . . . . .	210
C.2.4	Conclusion . . . . .	210
C.3	Adaptation d’ensembles de prédicteurs de vecteurs mouvement . . . . .	211
C.4	Perspectives de l’estimation de mouvement conjointe à l’encodeur et au décodeur . . . . .	211
	<b>Bibliographie</b>	<b>212</b>
	<b>Publications de l’auteur</b>	<b>224</b>

# Liste des tableaux

<b>Codage vidéo : un état de l'art</b>	<b>7</b>
3.1 Pourcentage de réduction de débit moyen d'une image P par rapport à une image I, d'une image B par rapport à une image I et d'une image B par rapport à une image P. Les débits et PSNR moyen ont été calculés sur l'ensemble des séquences VCEG pour le profil High IBBP GOP 12. . . . .	43
3.2 Pourcentage de sélection des modes de codage pour les slices P et B de la configuration High IBBP calculé en moyenne sur l'ensemble des séquences VCEG. . . . .	51
3.3 Pourcentage de sélection des modes Intra 16×16, 8×8 et 4×4 en fonction des résolutions pour la configuration High III, en moyenne pour l'ensemble des séquences VCEG pour six QP (17-42). . . . .	53
3.4 Pourcentage de sélection des modes Intra 16×16, 8×8 et 4×4 pour la séquence <i>CrowdRun</i> 50Hz à différentes résolutions spatiales (même nombre d'images) pour la configuration High III, en moyenne pour six QP (17-42). . . . .	53
3.5 Pourcentage de réduction de débit des configurations Baseline III et High III par rapport à JPEG 2000 en fonction des résolutions, pour trois ensembles de QP calculés sur l'ensemble des séquences VCEG. . . . .	54
<b>Modules de codage par compétition</b>	<b>56</b>
4.1 Sélection de chaque prédicteur inclus dans l'ensemble $\mathcal{P}$ pour neuf séquences CIF pour le profil Baseline IPP. . . . .	70
4.2 Pourcentage de réduction de débit de chaque compétition d'un prédicteur et du prédicteur $mv_{H.264}$ inclus dans l'ensemble $\mathcal{P}$ pour neuf séquences CIF pour le profil Baseline IPP. . . . .	71
4.3 Pourcentage de réduction de débit pour la compétition des prédicteurs de vecteurs mouvement et la compétition de prédicteurs de vecteurs mouvement pour le mode Skip pour différents ensembles de prédicteurs incluant 1, 2, ou 4 prédicteurs pour neuf séquences CIF pour le profil Baseline IPP. . . . .	72

4.4	Pourcentage de réduction de débit pour les images P en fonction de chaque combinaison de prédicteurs pour le codage de vecteurs mouvement et du mode Skip pour neuf séquences CIF pour le profil High IBBP. . . . .	73
4.5	Pourcentage de réduction de débit pour la première et la seconde image B en fonction de différentes combinaisons de prédicteurs pour le codage de vecteurs mouvement et du mode Skip pour neuf séquences CIF pour le profil High IBBP. . . . .	74
4.6	Pourcentage de réduction de débit de chaque séquence pour le profil Baseline IPP et le profil High IPP, pour trois ensembles de QP. . . . .	76
4.7	Pourcentage de réduction de débit de chaque séquence pour la configuration Baseline IPP avec un ensemble $\mathcal{P}_s$ pour le mode Skip composé des prédicteurs $mv_{Scol_{R_0}}$ et $mv_{H.264}$ pour trois ensembles de QP. . . . .	77
4.8	Pourcentage de réduction de débit de la séquence <i>Soccer</i> pour les résolutions QCIF, CIF et 4CIF codée en fonction de différents nombres d'images par seconde. La configuration Baseline IPP avec l'ensemble des QP 27-42 est utilisée. . . . .	77
4.9	Pourcentage de réduction du débit de l'information de mouvement de la référence par rapport à celui de la méthode proposée incluant les indices de prédicteurs, pour la configuration Baseline IPP, en fonction des QP, pour l'ensemble des séquences et des six QP utilisés. . . . .	78
4.10	Proportion du débit par rapport au débit total des résiduels de vecteurs mouvement pour la référence et du débit des résiduels et des indices de prédicteurs de la méthode proposée pour la configuration Baseline IPP, en fonction des QP, pour l'ensemble des séquences et des six QP utilisés. . . . .	79
4.11	Proportion de sélection des prédicteurs $mv_{Scol_{R_0}}$ et $mv_{H.264}$ et de leur égalité de l'ensemble $\mathcal{P}$ en fonction de la taille des blocs Inter (sauf Skip) de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences et des six QP utilisés. . . . .	81
4.12	Proportion de sélection des prédicteurs $mv_{Scol_{R_0}}$ et $mv_{H.264}$ de l'ensemble $\mathcal{P}$ et de leur égalité en fonction des images de référence de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences et des six QP utilisés. . . . .	82
4.13	Pourcentage de réduction de débit de chaque séquence et de leurs images P et B du profil High IBBP, pour trois ensembles de QP. . . . .	83
4.14	Pourcentage de réduction de débit moyen de la configuration du profil High composée de sept images B hiérarchiques, en fonction de chaque résolution, pour deux ensembles de QP. . . . .	84
4.15	Pourcentage de réduction de débit de chacune des méthodes intégrées dans le JM KTA, et de leur combinaison avec MVComp, pour l'ensemble des séquences 720p et les quatre QP VCEG dans la configuration High IPP. . . . .	85
5.1	Pourcentage d'appartenance à l'ensemble $\mathcal{P}_s^S$ de chaque prédicteur pour la méthode ADAPT 2 SETS en fonction de chaque séquence en moyenne sur les 4 QP. . . . .	93
5.2	Pourcentage d'appartenance à l'ensemble $\mathcal{P}_s^M$ de chaque prédicteur pour la méthode ADAPT 2 SETS en fonction de chaque séquence en moyenne sur les 4 QP. . . . .	93

6.1	Prédicteurs utilisés pour chacun des ordres de parcours de chacun des modes Intra 1D et chacun des ensembles de prédicteurs. . . . .	108
6.2	Entropie de chaque $Par_i$ pour le mode Intra $4 \times 4$ , le mode Ligne et ses trois ordres de parcours, le mode Colonne et ses trois ordres de parcours pour l'ensemble de la première image de la séquence <i>Tempete</i> CIF codée avec le profil Baseline pour le QP 27. Chaque mode et chacun de ces ordres de parcours a neuf prédicteurs. . . . .	110
6.3	Pourcentage de réduction de débit pour l'ensemble des séquences proposées incluant l'ensemble des séquences VCEG pour le profil Baseline III, High III et High III- $\{8 \times 8\}$ (sans Intra $8 \times 8$ ) pour la configuration Intra. . . . .	112
6.4	Pourcentage de sélection, en fonction des QP, des modes Intra $16 \times 16$ et $4 \times 4$ de la référence H.264/AVC et des modes Intra $16 \times 16$ , $4 \times 4$ et 1D, pour la méthode proposée, codée avec le profil Baseline III, calculé pour l'ensemble des séquences. . . . .	113
6.5	Pourcentage de sélection, en fonction des QP, des modes Intra $16 \times 16$ , $8 \times 8$ et $4 \times 4$ de la référence H.264/AVC et des modes Intra $16 \times 16$ , $8 \times 8$ , $4 \times 4$ et 1D pour la méthode proposée, codée avec le profil High III, calculé pour l'ensemble des séquences. . . . .	114
6.6	Pourcentage de sélection des ordres de parcours des partitionnements Ligne et Colonne en fonction des QP pour le profil Baseline III, calculé pour l'ensemble des séquences. Le parcours lexicographique est noté L. le symbole $\times 2$ signifie que deux ensembles de prédicteurs sont utilisés pour ce parcours. Le parcours bidirectionnel est noté B et le parcours hiérarchique est noté H. . . . .	115
6.7	Pourcentage de sélection de chaque prédicteur Intra $16 \times 16$ pour la référence et pour la méthode proposée en profil Baseline III, calculé pour l'ensemble des séquences. . . . .	115
6.8	Pourcentage de sélection de chaque prédicteur Intra $4 \times 4$ pour la référence et pour la méthode proposée en profil Baseline III, calculé pour l'ensemble des séquences. . . . .	116
6.9	Pourcentage de sélections de chaque prédicteur Intra Ligne et Intra Colonne de l'ordre de parcours lexicographique pour le profil Baseline III, calculé pour l'ensemble des séquences. . . . .	117
6.10	Proportion de débit des indices de modes et des indices de prédicteurs dans le débit total pour la référence et la méthode proposée pour le profil Baseline III. . . . .	118
6.11	Proportion de débit des indices de modes et des indices de prédicteurs dans le débit total pour la référence et la méthode proposée pour le profil High III. Le nombre de bits par macrobloc ( <i>bpm</i> ), relatif au codage des indices de prédicteurs, par mode Intra est indiqué en gras. . . . .	120
6.12	Pourcentage de réduction de débit de la configuration Baseline III et Baseline III Intra 1D par rapport à JPEG 2000 pour l'ensemble des séquences VCEG. . . . .	121
6.13	Pourcentage de réduction de débit pour l'ensemble des séquences proposées incluant l'ensemble des séquences VCEG, pour les profils Baseline GOP 12 IPP et High GOP 12 IBBP. . . . .	123

<b>Que doit-on vraiment transmettre ?</b>	<b>125</b>
7.1 Proportion de sélection des prédicteurs $mv_{Scol_{R_0}}$ et $mv_{H.264}$ , et de leur égalité de l'ensemble $\mathcal{P}$ lorsque la méthode proposée est activée, de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences et des six QP utilisés. . . . .	132
7.2 Proportion de suppression des indices des prédicteurs avec l'égalité des prédicteurs et la méthode proposée et le pourcentage total d'indices non transmis en fonction des images de référence de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences et des six QP utilisés. . . . .	133
7.3 Proportion de suppression des indices des prédicteurs obtenue avec la méthode proposée, en fonction de plusieurs fonctions de sélection du meilleur prédicteur, pour la séquence <i>Foreman</i> QCIF 15Hz codée avec le profil Baseline à QP 32. . . . .	134
8.1 Pourcentage de réduction de débit avec les critères $C_{ideal}$ , $C_{err}$ et $C_{opt}$ appliqués sur les prédicteurs de l'Intra 4×4, pour le profil Baseline sur l'ensemble des séquences VCEG. . . . .	145
8.2 Pourcentage de réduction de débit, avec le critère $C_{opt}$ , de l'ensemble des séquences, pour le profil Baseline III et le profil High III, pour trois ensembles de QP, lorsque la méthode est activée pour les modes Intra 4×4 et 8×8. . . . .	146
8.3 Pourcentage de sélection, en fonction des QP, des modes Intra 16×16, 4×4 et 8×8 pour la méthode proposée, codée avec le profil Baseline III et le profil High III, calculé pour l'ensemble des séquences. . . . .	147
8.4 Nombre de bits moyen par macrobloc ( $bpm$ ) nécessaire au codage des prédicteurs, des modes Intra, 4×4 et 8×8, en fonction des QP, pour la référence et la méthode proposée, codées avec le profil Baseline III et le profil High III, calculé pour l'ensemble des séquences. . . . .	148
8.5 Pourcentage de réduction de débit de l'ensemble des séquences avec le critère $C_{opt}$ et la méthode Intra 1D, comparée à la méthode Intra 1D seule et la référence, pour le profil Baseline III et le profil High III, pour deux ensembles de QP, lorsque la méthode est activée pour les modes Intra 4×4, 8×8 et 1D. . . . .	150
8.6 Pourcentage de sélection, en fonction des QP, des modes Intra 16×16, 4×4, 8×8 et Intra 1D pour la méthode proposée, codée avec le profil Baseline III et le profil High III, calculé pour l'ensemble des séquences. . . . .	151
8.7 Nombre de bits moyen par macrobloc nécessaire au codage des prédicteurs, des modes Intra, 4×4, en fonction des QP, pour la référence et la méthode proposée, codées avec le profil Baseline III, calculé pour l'ensemble des séquences. . . . .	151
8.8 Nombre de bits moyen par macrobloc nécessaire au codage des prédicteurs, des modes Intra 4×4 et 8×8, en fonction des QP, pour la référence et la méthode proposée, codées avec le profil High III, calculé pour l'ensemble des séquences. . . . .	152
8.9 Pourcentage de réduction de débit calculé sur l'ensemble des séquences avec le critère $C_{opt}$ pour la suppression de résolutions sous-pixelliques de la compensation de mouvement, pour le profil Baseline IPP l'ensemble des QP 27-42. . . . .	153

9.1	Pourcentage de réduction de débit obtenu lorsque seule la méthode EMD est utilisée au pixel entier et au $\frac{1}{4}$ de pixel, pour l'Inter $4 \times 4$ , l'Inter $16 \times 16$ et la combinaison de 3 modes Inter ( $4 \times 4$ , $8 \times 8$ , $16 \times 16$ ) en fonction du nombre $N$ de lignes ( $l_a^i, \forall i = 1..N$ ) et $M$ de colonnes ( $c_a^j, \forall j = 1..M$ ) utilisées pour l'appariement, pour l'ensemble des QP 22-37. . . . .	163
9.2	Pourcentage de réduction de débit de la compétition entre l'estimation de mouvement au décodeur avec $N = 4$ et $M = 4$ et l'estimation de mouvement classique, pour la résolution au pixel entier et au $\frac{1}{4}$ de pixel, pour chaque mode Inter, en moyenne sur l'ensemble des QP 22-37. . . . .	164
9.3	Pourcentage de réduction de débit de la compétition entre l'estimation de mouvement au décodeur pour 4 lignes et 4 colonnes utilisées pour l'appariement et l'estimation de mouvement classique, pour la résolution au pixel entier et au $\frac{1}{4}$ de pixel, pour trois ensembles de QP. . . . .	165
9.4	Pourcentage de sélection de l'estimation de mouvement au décodeur (Sans MV) pour 4 lignes et 4 colonnes utilisées pour l'appariement et l'estimation de mouvement classique (Avec MV), pour la résolution au pixel entier et au $\frac{1}{4}$ de pixel, pour chaque configuration présentée dans le tableau 9.2 et pour la configuration incluant l'ensemble des modes Inter, en moyenne pour l'ensemble des QP 22-37. . . . .	166
9.5	Proportion du débit des résiduels de vecteurs mouvement pour la référence et du débit des résiduels et des indices de signalisation de l'estimation de mouvement (EM) choisie, en fonction des six QP utilisés. Les résultats concernent la compensation de mouvement au $\frac{1}{4}$ de pixel. . . . .	167
9.6	Pourcentage de réduction de débit de la compétition entre l'Inter 1D et les autres modes Inter de la norme H.264/AVC, pour une compensation de mouvement au pixel entier et au $\frac{1}{4}$ de pixel, pour chaque séquence, pour trois ensembles de QP. . . . .	168
9.7	Pourcentage de sélection de l'Inter 1D, pour la résolution au pixel entier et au $\frac{1}{4}$ de pixel, pour la configuration présentée dans le tableau 9.6 en fonction des QP. . . . .	169
9.8	Pourcentage de réduction de débit de la compétition débit-distorsion entre la méthode EMD lorsque 1, 2, 4, 6 et 8 lignes sont utilisées pour l'appariement, et l'estimation de mouvement classique, pour la résolution au pixel entier et au $\frac{1}{4}$ de pixel, pour chaque séquence, pour l'ensemble des QP 22-37. . . . .	170
9.9	Pourcentage de sélection de l'estimation de mouvement au décodeur (Sans MV) pour 4 lignes utilisées pour l'appariement et l'estimation de mouvement classique (Avec MV), de la configuration Inter 1D du tableau 9.8 pour la résolution au pixel entier et au $\frac{1}{4}$ de pixel, en moyenne pour l'ensemble des QP 22-37. . . . .	171
9.10	Pourcentage de réduction de débit lorsque l'ensemble des modes Inter H.264/AVC et l'Inter 1D sont activés, pour la compétition entre la méthode EMD ( $N = 4$ et $M = 4$ ) et l'estimation de mouvement classique, pour la résolution au pixel entier et celle au $\frac{1}{4}$ de pixel, pour chaque séquence pour trois ensembles de QP. . . . .	171

9.11	Pourcentage de sélection de l'Inter 1D lorsque l'ensemble des modes Inter H.264/AVC sont activés pour la compétition débit-distorsion entre la méthode EMD ( $N = 4$ et $M = 4$ ) et l'estimation de mouvement classique, pour la résolution au pixel entier et celle au $\frac{1}{4}$ de pixel, pour la configuration présentée dans le tableau 9.10 en fonction des QP. . . . .	172
9.12	Pourcentage de réduction de débit de l'utilisation conjointe de la méthode proposée (RD EMD) et de la méthode MVComp, pour trois ensembles de QP. . . . .	173
<b>Résultats globaux</b>		<b>177</b>
10.1	Pourcentage de réduction de débit de chaque séquence pour la configuration Baseline GOP 13 IPP et la configuration High GOP 13 IBBP, pour trois ensembles de QP. . . . .	181
10.2	Pourcentage de réduction de débit de chaque type de slice pour la configuration Baseline GOP 13 IPP et la configuration High GOP 13 IBBP, pour trois ensembles de QP. . . . .	183
B.1	Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF pour 1, 2, 3, 5, 9, 17 prédicteur(s) dans l'ensemble des prédicteurs du mode Intra 1D Ligne lexicographique. . . . .	200
B.2	Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF en fonction du nombre de prédicteurs de chacun des ensembles utilisés pour le codage des partitions $Par^P$ et $Par^B$ du mode Intra 1D Ligne bidirectionnel. . . . .	200
B.3	Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF pour 1, 2, 3, 5, 9 prédicteur(s) dans l'ensemble des prédicteurs du mode Intra 1D Ligne hiérarchique. . . . .	201
B.4	Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF en fonction du nombre de prédicteurs des deux ensembles du mode Intra 1D Ligne lexicographique. . . . .	201
B.5	Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF pour 1, 2, 3, 5, 9, 17 prédicteur(s) dans l'ensemble des prédicteurs du mode Intra 1D Colonne lexicographique. . . . .	202
B.6	Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF en fonction du nombre de prédicteurs de chacun des ensembles utilisés pour le codage des partitions $Par^P$ et $Par^B$ du mode Intra 1D Colonne bidirectionnel. . . . .	203
B.7	Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF pour 1, 2, 3, 5, 9 prédicteur(s) dans l'ensemble des prédicteurs du mode Intra 1D Colonne hiérarchique. . . . .	203

B.8 Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF en fonction du nombre de prédicteurs des deux ensembles du mode Intra 1D Colonne lexicographique. . . . . 204



# Liste des figures

<b>Codage vidéo : un état de l'art</b>	<b>7</b>
1.1 Schéma global d'un codeur H.264/AVC. . . . .	11
1.2 Parcours zigzag par ordre de fréquences d'un bloc de coefficients DCT 8×8. . . . .	15
1.3 Partitionnements de l'Intra de la norme H.264/AVC. Le signal de référence pour le premier bloc est représenté en gris. Les numéros à l'intérieur des blocs correspondent à l'ordre de parcours des partitionnements. . . . .	16
1.4 Ordre de parcours de l'Intra 8×8 et l'Intra 4×4 dans la méthode BIP. . . . .	17
1.5 Les neuf prédicteurs des blocs Intra 4×4 de la norme H.264/AVC. . . . .	19
1.6 Décomposition en ondelettes de l'image <i>Lena</i> sur trois niveaux avec un banc de filtres 9/7. . . . .	20
1.7 Localisation des vecteurs spatiaux et temporels utilisés pour la prédiction du vecteur courant <i>mv</i> . . . . .	25
1.8 Prédicteurs du mode Direct temporel des slices B. . . . .	26
2.1 Schéma d'une méthode de compétition pour $N$ codages. . . . .	32
3.1 Courbes débit-distorsion de la deuxième image de la séquence <i>Foreman</i> CIF 30Hz codée en I, P et B. Ces trois types d'images proviennent respectivement de la configuration High III, High IPP et High IBBP. . . . .	42
3.2 Histogramme cumulatif de la proportion de débit de chaque type d'image I, P et B pour la séquence <i>Foreman</i> CIF 30Hz codée à l'aide de la configuration High IBBP GOP 12, et de la proportion du nombre d'images. . . . .	44
3.3 Représentation visuelle des coûts J, D et R pour 3 images successives de la séquence <i>Foreman</i> CIF 30Hz codées respectivement en I, B, P avec les QP respectivement de 36, 37, 38. . . . .	46
3.4 Evolution de la proportion des débits des composantes du flux pour les slices I en fonction des QP de la configuration High III calculée en moyenne sur l'ensemble des séquences VCEG. . . . .	47

3.5	Evolution de la proportion des débits des composantes du flux pour les slices P en fonction des QP de la configuration Baseline IPP calculée en moyenne sur l'ensemble des séquences VCEG. . . . .	48
3.6	Evolution de la proportion des débits des composantes du flux pour les slices B en fonction des QP de la configuration High IBBP calculée en moyenne sur l'ensemble des séquences VCEG. . . . .	49
3.7	Représentation visuelle des valeurs de la composante $X$ des vecteurs mouvement des images 3 et 4 de la séquence <i>Foreman</i> CIF 30Hz codée avec le profil Baseline IPP. Pour la représentation visuelle les valeurs de la composante horizontale sont centrées en 128. . . . .	51
3.8	Représentation spatiale de la sélection de 3 images de la séquence <i>Foreman</i> CIF 30Hz, codées respectivement en I, P, B avec les QP respectivement de 36, 37, 38. . . . .	52
<b>Modules de codage par compétition</b>		<b>56</b>
4.1	Localisation des vecteurs spatiaux et temporels utilisés. . . . .	60
4.2	Exemple de localisation des vecteurs $mv_{H.264}$ , $mv_{col}$ et $mv_{trans}$ pour une image de référence. . . . .	61
4.3	Vecteur mouvement collocated de l'image de référence $Ref_0$ pointant sur l'image $Ref_j$ et le vecteur courant $mv$ pointant sur l'image $Ref_i$ . . . . .	62
4.4	Vecteur mouvement collocated $mv_{Scol_i}$ et temporel translaté $mv_{trans_i}$ . . . . .	63
4.5	Prédicteurs du mode Direct Temporel des slices B, et les distances temporelles associées à la mise à l'échelle. . . . .	64
4.6	Vecteurs mouvement collocated de l'image $B - 1$ et les distances temporelles associées à la mise à l'échelle. . . . .	66
4.7	Augmentation de sélections du mode Skip avec la méthode proposée, pour les QP 17, 22, 37 et 42 pour huit séquences CIF codées avec le profil Baseline IPP. . . . .	80
4.8	Evolution de la proportion de sélection des prédicteurs $mv_{Scol_{R_0}}$ et $mv_{H.264}$ et de leur égalité de l'ensemble $\mathcal{P}$ en fonction des QP de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences. . . . .	81
5.1	Pourcentage d'augmentation moyen du nombre de macroblocs codés avec le mode Skip par séquence, pour MVComp, ADAPT 1 SET et ADAPT 2 SETS par rapport à la référence H.264/AVC en moyenne sur les 4 QP. . . . .	91
5.2	Nombre moyen de prédicteurs utilisés pour chacun des ensembles de prédicteurs $\mathcal{P}_s^M$ et $\mathcal{P}_s^S$ de la méthode ADAPT 2 SETS, pour chaque séquence, en moyenne sur les 4 QP. . . . .	92
5.3	Nombre de bits moyen par macrobloc nécessaire au codage de l'indice du prédicteur du mode Skip, pour chaque séquence, pour la méthode MVComp et pour chacun des ensembles $\mathcal{P}_s^M$ et $\mathcal{P}_s^S$ de la méthode d'adaptation d'ensemble de prédicteurs, en moyenne sur les 4 QP. . . . .	94

5.4	Représentaion spatiale de la classification des macroblocs en zones statiques ou en zones de mouvement, de l'image 3 de la séquence <i>Ice</i> 4CIF. . . . .	95
5.5	Proportion de macroblocs sautés sélectionnés dans l'ensemble $\mathcal{P}_s^M$ ou l'ensemble $\mathcal{P}_s^S$ par rapport au nombre total de macroblocs codés en mode Skip, pour chaque séquence en moyenne sur les 4 QP. . . . .	96
5.6	Réduction du débit pour la méthode MVComp, l'adaptation d'un ensemble de prédicteurs et l'adaptation de deux ensembles de prédicteurs pour le mode Skip, pour chaque séquence en moyenne sur les 4 QP. . . . .	97
6.1	Exemple de partition initiale $Par_1$ et les partitions $Par_2$ et $Par_3$ engendrées par la dilatation au rang 2 et 3. . . . .	101
6.2	Ordres de parcours bidirectionnel et hiérarchique pour un partitionnement Ligne. . .	102
6.3	Pixels du signal de référence utilisés pour le calcul des prédicteurs du partitionnement Ligne pour le parcours lexicographique. . . . .	103
6.4	Pixels du signal de référence utilisés pour le calcul des prédicteurs du partitionnement Ligne pour le parcours bidirectionnel dans le cas des partitions utilisant les partitions $Par_i^B$ . . . . .	104
6.5	Pixels du signal de référence utilisés pour le calcul des prédicteurs du partitionnement Ligne pour le parcours hiérarchique dans le cas des partitions $Par_i^{BH}$ . . . . .	105
6.6	Représentation spatiale de la sélection des modes Intra pour l'image 51 de la séquence <i>Panslow</i> 720p codée avec le profil Baseline III à QP 37 pour la méthode proposée. .	118
6.7	Représentation spatiale de la sélection des modes Intra pour l'image 2 de la séquence <i>Stockholm</i> 720p codée avec le profil Baseline III à QP 22 pour la méthode proposée. .	119
6.8	Représentation spatiale de la sélection des modes Intra pour l'image 91 de la séquence <i>Harbour</i> 4CIF codée avec le profil Baseline III à QP 32 pour la méthode proposée. .	119
6.9	Courbes débit-distorsion pour la référence JPEG 2000, H.264/AVC Baseline III et High III et la méthode proposée en Baseline III et High III pour la séquence Schumacher CIF 30Hz. . . . .	122

## Que doit-on vraiment transmettre ?

125

7.1	Proportion des indices de prédicteurs des vecteurs mouvement transmis et non transmis grâce au test d'égalité et ceux éliminés par le critère des prédicteurs implicites, pour la configuration Baseline IPP, en fonction des six QP utilisés. . . . .	131
7.2	Représentation graphique des fonctions $Golomb(x)$ , $ x $ , $x^2$ , $e^x$ , $floor(\log( x ))$ . . . .	134
8.1	Représentation graphique de $X$ , $P_i$ , $\varepsilon_{P_i}$ , $\tilde{\varepsilon}_{P_i}$ et des zones où le critère de l'équation (8.11) est vérifié lorsque $E$ correspond à un pixel. . . . .	140
8.2	Histogramme des valeurs de $ Q^{-1}Q(DCT(X - P_i)) - DCT(X - P_i) $ de l'ensemble des prédicteurs mis en compétition pour le mode Intra 1D Ligne de la première image de la séquence <i>Foreman</i> CIF codée avec le profil Baseline à QP 42. . . . .	141

8.3	Pourcentage de réduction de débit obtenu avec l'utilisation du critère $C_{opt}$ sur les prédicteurs de l'Intra $4 \times 4$ , pour plusieurs valeurs de $QP_{opt}$ , en moyenne pour l'ensemble des séquences VCEG codées avec le profil Baseline pour l'ensemble de QP 27-42. . . . .	143
8.4	Pourcentage de réduction de débit obtenu avec l'utilisation du critère $C_{opt}$ sur les prédicteurs de l'Intra $8 \times 8$ , pour plusieurs valeurs de $QP_{opt}$ , en moyenne pour l'ensemble des séquences VCEG codées avec le profil High pour l'ensemble de QP 27-42. . . . .	143
8.5	Pourcentage de réduction de débit obtenu avec l'utilisation du critère $C_{opt}$ sur les prédicteurs de l'Intra 1D, pour plusieurs valeurs de $QP_{opt}$ , en moyenne pour l'ensemble des séquences VCEG codées avec le profil Baseline pour l'ensemble de QP 27-42. . . . .	144
9.1	Schéma d'encodage et de décodage de la prédiction Inter de type H.264/AVC. . . . .	157
9.2	Schéma d'encodage et de décodage de la prédiction Inter pour la méthode d'estimation de mouvement au décodeur proposée. . . . .	157
9.3	Bloc courant et bloc prédicteur avec les lignes et colonnes voisines utilisées pour l'appariement dans la méthode d'estimation de mouvement au décodeur. . . . .	159
9.4	Proportion du débit des résiduels de vecteurs mouvement, des résiduels et des indices de l'estimation de mouvement (EM) et de prédicteurs, pour la référence H.264/AVC, la méthode MVComp seule, la méthode proposée (RD EMD) et l'utilisation conjointe de MVComp et la méthode proposée, en moyenne pour l'ensemble des séquences et des six QP utilisés. . . . .	174
<b>Résultats globaux</b>		<b>177</b>
10.1	Pourcentage de réduction de débit en fonction des résolutions pour les deux configurations testées pour l'ensemble des QP 27-42. . . . .	182
10.2	Courbes débit-distorsion pour la référence H.264/AVC Baseline GOP 13 IPP et High GOP 13 IBBP et pour la combinaison des méthodes Baseline GOP 13 IPP et High GOP 13 IBBP pour six séquences pour les QP 27, 32, 37, 42. . . . .	184
C.1	Représentation des demi-images de référence et des images de référence classiques. . . . .	206
C.2	Image $Ref_1$ et les interpolations $Ref_{\frac{1}{2}}^{V1}$ , $Ref_{\frac{1}{2}}^{V2}$ et $Ref_{-1}^{Estimated}$ obtenues dans la méthode des demi images de référence pour la séquence Irene CIF 30Hz. . . . .	208

# Glossaire

4CIF	<i>four CIF</i> - Format de résolution de 704×576 pixels
ADSL	<i>Asymmetrical Digital Subscriber Line</i> - Technique qui permet la transmission et la réception de signaux numériques sur une ligne téléphonique
ARCEP	<i>Autorité de Régulation des Communications Electroniques et des Postes</i>
AVC	<i>Advanced Video Coding</i> - Codec vidéo communément appelé H.264/AVC ou MPEG-4 part 10
BMA	<i>Block Matching Algorithm</i> - Algorithme de mise en correspondance de blocs
CABAC	<i>Context Adaptive Binary Arithmetic Coding</i> - Codage arithmétique de la norme H.264/AVC
CAVLC	<i>Context Adaptive Variable Length Coding</i> - Codage entropique de la norme H.264/AVC
CIF	<i>Common Interchange Format</i> ou <i>Common Intermediate Format</i> - Format de résolution de 352×288 pixels
DCT	<i>Discrete Cosine Transform</i> - Transformée en cosinus discrète
DVD	<i>Digital Versatile Disc</i> - Standard de disque optique
DWT	<i>Discrete Wavelet Transform</i> - Transformée en ondelette discrète
GOP	<i>Group of Pictures</i> - Groupe d'images consécutives
GSM	<i>Global System for Mobile communications</i> - Norme de communication radio utilisée pour le téléphone portable
HD	<i>Haute Définition</i> - Format de résolution de 1280×720 pixels ou 1920×1080 pixels
HDDVD	<i>High Density Digital Versatile Disc</i> - Standard de disque optique remplaçant du DVD
ISO	<i>International Organization for Standardization</i> - Organisme de normalisation non gouvernemental
ITU	<i>International Telecommunications Union</i> - Organisme de normalisation de l'ONU
JM	<i>Joint Model</i> - Logiciel de référence associé à une norme
JPEG	<i>Joint Photographic Experts Group</i> - Groupe de travail de l'ISO chargé de la normalisation de format de compression d'images fixes

JVT	<i>Joint Video Team</i> - Groupe formé des groupes VCEG et MPEG
KTA	<i>Key Technical Area</i> - Logiciel contenant les outils jugés efficaces par le groupe VCEG depuis la fin de la normalisation d'H.264/AVC
MPEG	<i>Moving Picture Experts Group</i> - Groupe de travail "SC 29/WG 11" de l'ISO et de la CEI chargé de la normalisation de format de compression vidéo
PSNR	<i>Peak Signal Noise Ratio</i> - Rapport de signal à bruit crête à crête, mesure utilisée pour l'évaluation objective de la distorsion
QCIF	<i>Quarter CIF</i> - Format de résolution de 176×144 pixels
QP	<i>Quantization Parameter</i> - Paramètre utilisé pour fixer le degré de pertes introduites lors du processus de quantification dans un codage avec pertes
RD	<i>Rate Distortion</i> - Qualificatif attribué au critère de sélection prenant en compte à la fois le débit et la distorsion
RLC	<i>Run Length Coding</i> - Technique de compression consistant à affecter à un mot un facteur de répétition suivi de la valeur de ce mot
SD	<i>Standard Definition</i> - Format de résolution de 720×576 pixels
SMPTE	<i>Society of Motion Picture and Television Engineers</i> - Organisme de normalisation non gouvernemental
SVC	<i>Scalable Video Coding</i> - Extension scalable de la norme H.264/AVC
TNT	<i>Télévision Numérique Terrestre</i>
VC1	<i>Video Codec 1</i> - Norme de compression vidéo de l'organisme SMPTE initialement développé par Microsoft et basé sur la technologie de H.264/AVC
VCEG	<i>Video Coding Experts Group</i> - Groupe de travail "SG16-Q6" de l'ITU chargé de la normalisation de format de compression d'image et vidéo
VOD	<i>Video On Demand</i> - Technique de diffusion de vidéo numérique

# Introduction

## Vidéo numérique et applications

Durant ces dernières années, le domaine de la vidéo numérique a transformé le monde du multimédia avec de nouveaux supports de capture, de visionnage et de transmission. Ces avancées technologiques ont bouleversé le paysage médiatique en multipliant l'offre de contenu, de visionnage et de capture.

Les écrans cathodiques ont laissé place aux écrans plats qui sont vite devenus "HD Ready" puis "Full HD", multipliant ainsi par 20 le nombre de pixels dans une image. Parallèlement, les résolutions des chaînes de télévision sont en train d'évoluer, passant de la résolution SD ( $720 \times 576$ ) aux résolutions HD ( $1280 \times 720$  ou  $1920 \times 1080$ ). De même, le format DVD est remplacé par son successeur offrant de plus grandes résolutions : le Blue Ray. L'autre révolution est la mobilité des supports de visionnage avec les lecteurs de DVD portables, les baladeurs vidéo, la télévision sur mobile et bien entendu les ordinateurs portables.

L'offre de contenu télévisuel s'est accrue avec l'arrivée de la télévision numérique par câble et satellite, et plus récemment sur le réseau hertzien. La TNT a ainsi modifié le paysage de la télévision française gratuite, en offrant 18 chaînes au lieu des 6 du réseau classique. L'accroissement du nombre de contenus est aussi lié au développement d'Internet. Le nombre d'abonnements a doublé en 3 ans, passant de 8 millions au troisième trimestre 2005 à 16 millions au 31 mars 2008 d'après l'ARCEP [ARCEP]. Les fournisseurs d'accès Internet (FAI) offrent désormais un accès à plusieurs dizaines de chaînes de télévision, dont celles de la TNT et des chaînes des bouquets satellites, en plus de la connexion au réseau Internet. Le principal apport d'Internet, dans l'élargissement de l'offre de contenus vidéo, est apparu avec le "streaming". Cette méthode de diffusion est utilisée pour les services de vidéo à la demande (VOD) classique inclus dans toutes les "box" des fournisseurs d'accès, mais aussi pour le partage de vidéos non commerciales entre utilisateurs comme les sites de streaming vidéo tels que Google video, You Tube et Daily Motion.

Tous ces nouveaux outils entrés récemment dans notre quotidien ont modifié nos manières de consommer et de partager l'information visuelle. La consommation des vidéos ne se fait plus uniquement assis passivement devant un téléviseur mais devant un ordinateur, un mobile, un baladeur, dans une voiture, etc. Cette consommation n'est plus subie mais choisie : on peut regarder à tout moment les journaux télévisés, les séries, les émissions de toutes les chaînes du monde entier. Les sites de streaming vidéo (Google video, You Tube ou Daily Motion) qui exploitent ce nouveau mode de

consommation sont devenus des incontournables de l'actualité en proposant des vidéos en intégralité, des vidéos censurées par les comités de rédaction, des vidéos amateurs d'événements parfois repris ensuite par les médias classiques. La nature des contenus a, elle aussi, évolué avec l'intégration de caméras dans les téléphones mobiles qui permettent d'immortaliser des moments de vie avec un objet du quotidien toujours à portée de main. D'ailleurs des festivals dédiés à ces minis films ont commencé à apparaître comme le "Mobile Film Festival" [MFF09] ou le "Festival Pocket Films" [FPF09]. Ces changements dans le monde du multimédia ont permis à une entreprise de "transmission" de flux (le groupe Orange) de devenir également un fournisseur de contenus.

Les évolutions technologiques de transmission, de vitesse de calcul et de compression ont engendré un changement culturel de l'information et de la communication. Ces systèmes d'information et de communication vont certainement continuer à évoluer. Mais de quelle manière ? Est-ce que les réseaux filaires ou sans fil offriront des débits suffisamment élevés pour transporter toutes ces vidéos ? Enfin, est-ce que de nouveaux algorithmes de compression seront nécessaires ?

## Compression vidéo : un outil toujours nécessaire ?

L'utilité de la compression n'est plus à prouver : les mobiles GSM (nos téléphones portables) ne pourraient exister sans compression du signal audio faute de débit suffisant, de même pour la télévision sur ADSL. En effet, si l'on considère une séquence de télévision en SD contenant 4 composantes de luminance pour 4 composantes de chrominance (format 4:2:2) dont chacune d'entre elles est représentée par 8 bits (soit 16 bits par pixel) avec une fréquence de 24 images seconde, le débit nécessaire sans compression est de  $720 \times 576 \times 16 \times 24 = 159,25$  Mbits/s alors que le débit maximum proposé par les fournisseurs d'accès, pour une connexion ADSL, est de 30 Mbits/s. Notons que ce débit est le débit maximum théorique qu'il est possible d'atteindre (codage canal et encapsulation compris).

Les taux de compression atteints actuellement permettent d'obtenir une qualité convenable pour une majorité d'utilisateurs. On notera qu'une distance importante à une borne ADSL entraîne une perte de qualité sur la télévision mais aussi, parfois, une impossibilité d'utiliser conjointement l'accès Internet pour d'autres applications. Cette qualité de la télévision reste néanmoins supportable comparée à celles de vidéos émises depuis les sites de streaming. Certes ces qualités peuvent dépendre, pour les vidéos non commerciales, des capteurs utilisés, cependant les vidéos commerciales mises à la disposition des usagers sur les sites des chaînes de télévision sont elles aussi inférieures aux qualités des diffusions classiques. Les débits ou les capacités de compressions actuelles paraissent donc insuffisantes pour générer des séquences de bonne qualité de visionnage.

Cette nécessité de qualité pourrait être compensée par l'arrivée de la fibre optique. Certains fournisseurs d'accès annoncent des débits de l'ordre de 3 Gbits/s, soit un débit théorique multiplié par 100 par rapport à l'ADSL. Cependant, les tailles des images capturées et affichées s'agrandissent et vont donc augmenter la quantité d'information des contenus vidéo. Actuellement les constructeurs de téléviseurs proposent en démonstration des écrans 2160p, composés de 2160 lignes de 4096 pixels. Cette nouvelle technologie s'accompagne d'une capture et d'une restitution plus fine des couleurs avec 4 composantes de luminance et 8 composantes de chrominance (format 4:4:4), lesquelles sont

représentées sur 12 bits (au lieu de 8 actuellement). De plus, les séquences utilisées ont une fréquence de 120Hz. Le débit d'une telle séquence est de  $4096 \times 2160 \times 36 \times 120 = 38,2$  Gbits/s. L'utilisation de la résolution Ultra HD (7680×4320) est envisagée pour 2020 [SFK09]. L'entreprise de diffusion NHK prévoit des premiers tests de diffusion en 2011 [Sug08]. Cette résolution engendrerait, sans aucune autre augmentation du nombre de bits par pixel et de la fréquence entre les images, un débit de 143,3 Gbits/s, soit quasiment 1000 fois plus de données que dans un flux de télévision actuel. L'augmentation du nombre d'utilisateurs d'Internet va continuer à augmenter, réduisant ainsi la bande passante. De plus, les usages de la consommation de contenus multimédia sont en constante évolution. Chaque minute, plus de dix heures de vidéo sont ajoutées sur YouTube dans le monde. On peut aussi imaginer que la part de l'accès mobile, pour lequel le débit de la fibre optique ne sera d'aucun secours, va encore s'accroître dans l'avenir. L'augmentation des résolutions, des fréquences de visionnage et les meilleures qualités de restitution de couleurs ne sont pas les uniques avancées technologiques probables des prochaines années. Les vidéos en relief sur téléphones mobiles sont déjà une réalité pour les sud coréens. L'industrie du cinéma commence à utiliser cette technologie. En effet, certains films 2D sont convertis en 3D, d'autres sont entièrement filmés avec des caméras 3D ou proviennent d'images de synthèse tel que Shrek 4. La chaîne de télévision japonaise NHK diffuse une heure par jour des émissions en relief. Les constructeurs de téléviseurs travaillent actuellement sur des écrans capables de diffuser des flux 3D sans lunettes. Le "format" d'un flux 3D dépend de l'application et du système d'affichage utilisé. Ces flux sont composés, soit de plusieurs flux 2D, soit d'images classiques et d'images de profondeur. Cette technologie va encore augmenter la quantité d'information à transmettre.

Enfin, la consommation énergétique des appareils mobiles de capture et de visionnage est un enjeu majeur. Les avancées technologiques sur les batteries permettent une autonomie de plus en plus longue, mais qui reste encore insuffisante. La compression vidéo est essentielle dans ce domaine. En effet, la consommation liée à un encodage ou un décodage est moindre comparée à celle d'une réception ou d'une transmission d'un signal vidéo sur un réseau sans fil. Des taux de compression plus élevés permettent et permettront donc d'augmenter cette autonomie.

## Contexte de la thèse

Ce travail a été effectué au sein du service des images animées du département Orange Labs du groupe Orange. Cette équipe a en charge le développement de la brique logiciel de visiophonie (e-conf) intégrée dans les produits entreprise et grand public de vidéo conférence. Dans ce service, une équipe est chargée de participer à la normalisation des prochains standards vidéo au sein du groupe VCEG (Video Coding Expert Group) ITU-T SG16-Q6 et du groupe MPEG (Moving Picture Expert Group) ISO/IEC JTC 1/SC 29/WG 11. Le but de ces participations est dans un premier temps de contribuer à la normalisation d'un nouveau standard et dans un second temps d'anticiper ces futures normes pour une intégration rapide dans le logiciel e-conf.

La norme H.264/MPEG-4 AVC issue des travaux du groupe JVT (Joint Video Team) (réunion des groupes VCEG et MPEG), a été finalisée en mai 2005. Ce standard apporte des gains significatifs par rapport à ses prédécesseurs H.263 et MPEG-4 part 2. Cette efficacité de codage provient de

l'amélioration d'outils existants et de l'introduction de nouvelles méthodes de codage. Parallèlement, un travail conséquent a été effectué sur la sélection non normative des meilleures possibilités de codage. A la fin de cette normalisation, les groupes de travail se sont concentrés sur l'annexe scalable de ce standard, H.264/SVC qui a été normalisée en juillet 2007, puis sur l'annexe multivue, H.264/MVC, finalisée en juillet 2008. Cependant, les activités classiques sur l'amélioration du codage vidéo se sont poursuivies durant cette période. Afin d'encourager de nouvelles contributions, il a été décidé à la 26<sup>ième</sup> réunion du groupe VCEG, en avril 2005, la création d'un logiciel nommé KTA (Key Technical Area) qui regroupe tous les outils apportant des gains jugés significatifs par ce groupe et proposés depuis la fin de la normalisation d'H.264/AVC [H.264/AVC]. Dans ce contexte, l'objectif principal de la thèse est de fournir des méthodes normalisables, suffisamment efficaces et réalistes pour être intégrées dans une éventuelle future norme. Dans ce but, l'approche définie au sein d'Orange Labs est l'étude de méthodes de codage par compétition pour l'amélioration incrémentale de la référence. Cette étude comprend à la fois la mise en place de nouveaux modules de codage par compétition et la sélection optimale parmi ces nouvelles possibilités de codage, mais aussi la suppression de l'information de compétition. Dans ce manuscrit nous détaillerons tout particulièrement les modules de compétition qui ont fourni des améliorations significatives de la référence dont :

- un module de compétition pour le codage de l'information de mouvement permettant d'exploiter à la fois les redondances spatiales et les redondances temporelles entre les champs de vecteurs mouvement
- une sélection automatique d'ensembles de prédicteurs pour le mode Skip basée sur le contenu des séquences
- un nouveau mode Intra reposant sur un partitionnement 1D des macroblocs afin de réduire la distance spatiale entre le signal à coder et le signal de référence utilisé pour le calcul des prédicteurs

La deuxième partie de cette thèse a été consacrée à la réduction de débit de l'information de compétition. En ce sens, nous avons développé des méthodes originales de suppression de cette information à la fois pour les outils de la référence et pour les modules de codage par compétition développés. Parmi ces méthodes, nous présentons :

- une suppression des prédicteurs de vecteurs mouvement implicites
- une élimination de prédicteurs Intra, adaptée à la fois au contenu des images et aux débits visés pour le codage
- une estimation de mouvement au décodeur, permettant d'éliminer une partie de l'information de mouvement.

Ces méthodes de suppression sont fondées sur une augmentation de l'intelligence du décodeur pour améliorer l'efficacité de codage. Les résultats obtenus montrent que cette voie de recherche présente un fort potentiel pour un éventuel remplaçant de la norme H.264/AVC.

## Plan du manuscrit

Ce manuscrit s'articule en quatre parties. La première est un état de l'art du codage vidéo dans laquelle nous examinerons la norme H.264/AVC d'un point de vue théorique et d'un point de vue

pratique. La deuxième partie présente nos travaux sur les modules de codage par compétition et la sélection des nouvelles possibilités de codage. Dans la troisième partie, les méthodes de suppression de l'information de mouvement sont décrites, enfin la quatrième partie donne les résultats de l'utilisation conjointe de méthodes développées dans ce manuscrit.

### **Partie 1 :**

- Le chapitre 1 énonce les modules du schéma global de la norme H.264/AVC. Nous étudions tout particulièrement le codage Intra prédictif et la norme de référence du codage d'image : JPEG 2000. Puis nous analysons le codage de l'information de mouvement, avant d'exposer brièvement les méthodes proposées en normalisation qui sont déjà intégrées dans le JM KTA.
- Le chapitre 2 est dédié au codage par compétition. Nous évoquons notamment les étapes de sélection basées sur le critère de débit-distorsion ou sur des critères a priori et les possibilités de codage qui ont été mises en compétition dans l'état de l'art.
- Le chapitre 3 propose une étude comportementale du codec de référence de la norme, dont les conclusions sont à l'origine de nos contributions. Nous reportons spécialement les proportions des différentes composantes du débit ainsi que la sélection spatiale des modes de codage. Cette étude nous a permis de déterminer nos axes de recherche, détaillés dans les parties 2 et 3.

### **Partie 2 :**

- Le chapitre 4 décrit notre module de codage par compétition de prédicteurs de vecteurs mouvement. Les prédicteurs mis en compétition et la sélection de ces prédicteurs sont étudiés avant de donner les résultats complets de cette méthode.
- Le chapitre 5 présente une adaptation d'ensembles de prédicteurs pour le mode Skip. Cette adaptation est une extension de la méthode développée dans le chapitre 4. La sélection des ensembles de prédicteurs est effectuée image par image, à l'aide d'un critère a priori basé sur une classification des macroblocs en fonction du mouvement.
- Le chapitre 6 expose notre nouveau mode de codage prédictif Intra. Ce mode est composé d'une étape de partitionnement d'un macrobloc en partition 1D permettant ainsi de réduire la distance spatiale entre les pixels à prédire et le signal de référence. Pour améliorer l'efficacité de codage de ce mode, nous proposons deux ordres de parcours des partitions, qui ont déjà prouvé leur efficacité pour l'exploitation des redondances temporelles entre les images, en plus de l'ordre de parcours lexicographique. Des prédicteurs adaptés à ce nouveau partitionnement sont également étudiés.

### **Partie 3 :**

- Le chapitre 7 décrit la suppression des indices des prédicteurs des vecteurs induit par le critère débit-distorsion. La méthode relatée dans ce chapitre est le fondement de nos travaux sur la suppression de l'information de compétition grâce à la connaissance du décodeur, des critères de choix de codage.

- Le chapitre 8 développe une technique d'élimination de prédicteurs Intra précédant le processus de sélection afin de réduire le débit des modes Intra. Cette élimination est conjointe à l'encodeur et au décodeur. Le processus d'élimination utilise la distance dans le domaine transformée entre les prédicteurs prédéfinis. Le critère d'élimination utilisé dépend des paramètres de codage. Par conséquent la méthode est adaptée à la fois au contenu des images et aux qualités visées par le codage.
- Le chapitre 9 propose finalement de réduire le débit de l'information de mouvement, en considérant cette information comme une information de compétition. Dans ce cadre, une méthode d'estimation de mouvement conjointe à l'encodeur et au décodeur est présentée, permettant de ne pas transmettre les résiduels de vecteurs mouvement. De plus un partitionnement 1D du macrobloc est mis en place afin d'améliorer cette estimation de mouvement.

#### **Partie 4 :**

- Enfin, dans le chapitre 10, nous proposons de mesurer l'efficacité de codage de l'utilisation conjointe des méthodes développées dans ce manuscrit qui ont à la fois atteint un stade d'étude suffisant et qui nous paraissent être efficaces pour être intégrées dans un éventuel remplaçant du standard de référence.

Une conclusion et des perspectives pour de futurs travaux sont présentées en fin de manuscrit.

Première partie

Codage vidéo : un état de l'art



# Chapitre 1

## Le codage vidéo et la norme H.264/AVC

*L'objectif de la thèse est de proposer des améliorations du schéma de codage du standard H.264/AVC afin de contribuer à son successeur. Ce standard est donc la base de nos travaux. Son efficacité de codage est la borne que nous devons dépasser. Dans ce chapitre, nous décrirons donc l'algorithme de compression de la norme H.264/AVC en expliquant chacune des étapes de codage à l'aide du schéma global. Pour permettre une bonne compréhension de nos contributions exposées dans les parties II et III, nous étudierons tout particulièrement le codage prédictif Intra de la norme H.264/AVC et les différentes propositions de la littérature pour améliorer ce codage, sans oublier de décrire la référence du codage d'image : JPEG 2000. De même, nous décrirons le codage de l'information de mouvement en fonction des stratégies utilisées pour la compensation de mouvement. Enfin, nous exposerons brièvement les méthodes proposées en normalisation, qui ont déjà été jugées efficaces par le Video Coding Expert Group (VCEG).*

### 1.1 Description du schéma global de codage de la norme H.264/AVC

La norme H.264/AVC [H.264/AVC] est un codeur vidéo en boucle fermée, tout comme l'ensemble de ses prédécesseurs. En effet, les informations déjà décodées sont utilisées pour le codage de la partie courante. Le schéma global de codage de cette norme est représenté dans la figure 1.1. La séquence en entrée (en haut à gauche dans le schéma), est une succession d'images (matrices de pixels). Chaque image est découpée en "slices". Un slice est une partie de l'image ou l'image entière en fonction des paramètres d'entrée. Ces slices sont découpés en macroblocs (bloc de taille  $16 \times 16$ ), le macrobloc étant l'unité de codage dans la norme. Chaque macrobloc est codé soit en Intra (boîte 1), soit en Inter (boîte 12). Chacun de ces codages engendre plusieurs résiduels de texture, qui sont comparés avec un module de décision (boîte 2). Le résiduel qui donne la meilleure possibilité de codage, en terme d'un critère débit-distorsion dans le logiciel de référence [JM] de la norme, est alors sélectionné.

Ce résiduel est décorrélé avec la transformée en cosinus discrète (DCT) (boîte 3). Puis, ce résiduel transformé est quantifié (boîte 4) et les coefficients engendrés sont envoyés dans le codeur entropique sans perte (boîte 10) qui produit le train binaire (le flux).

Le standard H.264/AVC étant un codeur en boucle fermée, il est nécessaire d'introduire une boucle de décodage à l'intérieur de l'encodeur. Par conséquent, les résiduels transformés et quantifiés sont déquantifiés (boîte 5) et détransformés (boîte 6) à l'intérieur du codeur. Aux blocs ainsi générés, on applique la prédiction inverse. Cette opération consiste à ajouter le prédicteur sélectionné dans le module de décision (boîte 2) (le meilleur prédicteur Intra ou Inter). Ensuite, un filtre de deblocking est appliqué sur l'image reconstruite, permettant d'éliminer certaines dégradations produites par le module de quantification (boîte 4) qui engendre des pertes d'information. Ce filtre lisse les images de référence en bordure des blocs. Enfin, les macroblocs et slices décodés sont stockés en mémoire (boîte 9). Les blocs décodés de l'image courante, stockés dans ce module, sont utilisés pour le calcul des prédicteurs Intra. De même, les images précédemment décodées et débloquentées (lissées) sont utilisées pour le codage Inter. En effet, le mouvement entre le bloc courant et ces images est estimé (boîte 11), le prédicteur engendré par cette estimation de mouvement servira pour la prédiction Inter que l'on appelle généralement compensation de mouvement (boîte 12). Afin que le décodeur soit capable de retrouver ce prédicteur Inter, un vecteur correspondant au mouvement entre le bloc courant et le prédicteur est transmis au décodeur. Les vecteurs issus du codage Inter sont prédits (boîte 13) puis envoyés dans le codeur entropique. Ces vecteurs prédits puis codés sont ensuite insérés dans le train binaire.

Les différentes étapes de ce schéma de codage peuvent être segmentées en deux parties : le calcul du résiduel et le codage du résiduel. Nous décrivons, alors, brièvement ces deux parties dans la suite de ce chapitre.

### 1.1.1 Calcul du résiduel

Le résiduel du bloc courant, appelé aussi erreur de prédiction, est la différence entre un prédicteur et ce bloc courant. L'expression du résiduel  $e$  est donnée dans l'équation (1.1) où  $p(x, y)$  est le pixel du bloc courant à la position  $(x, y)$  et  $\hat{p}$  est le prédicteur :

$$e(x, y) = p(x, y) - \hat{p}(x, y) \quad (1.1)$$

L'opération de prédiction inverse est donnée dans l'équation (1.2). Le prédicteur  $\hat{p}$  est ajouté au résiduel pour retrouver les pixels  $p(x, y)$  du bloc courant. Le décodeur calcule le prédicteur  $\hat{p}$  à partir des informations déjà décodées et extraites du train binaire :

$$p(x, y) = \hat{p}(x, y) + e(x, y) \quad (1.2)$$

Les prédictions Intra et Inter permettent d'exploiter respectivement les corrélations spatiales et temporelles des séquences vidéo. Les prédicteurs Intra sont calculés à partir des informations déjà décodées de l'image courante et plus précisément des pixels voisins du bloc courant. Nous décrivons plus en détail le codage Intra de la norme H.264/AVC dans la section 1.2.

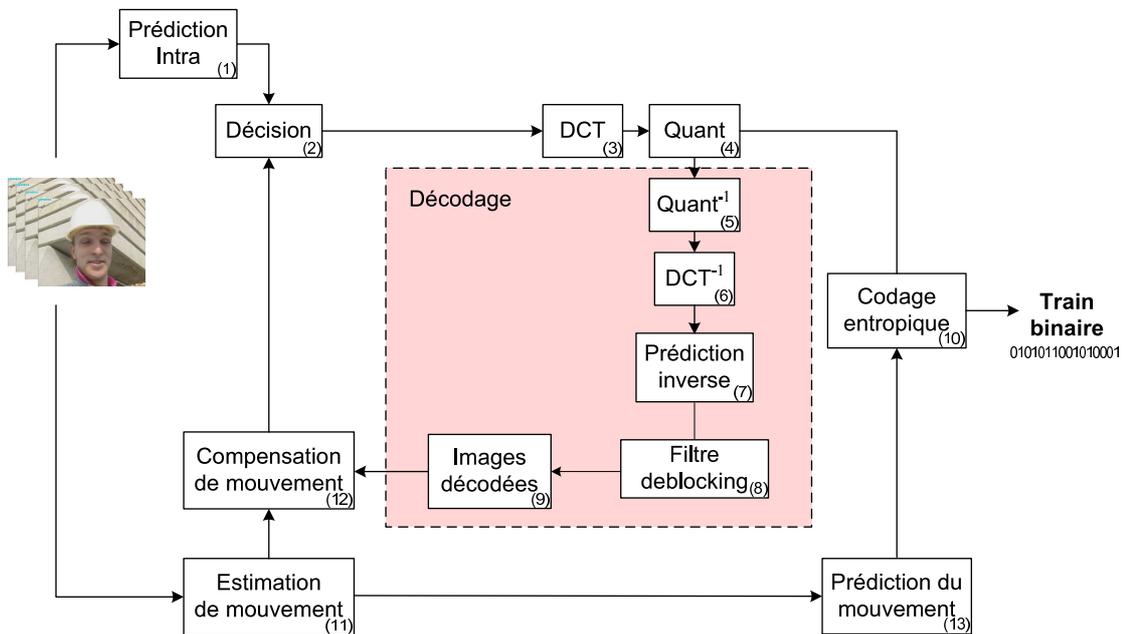


FIGURE 1.1 – Schéma global d'un codeur H.264/AVC.

Dans le standard, on distingue trois types de slices : I, P, B. Les slices I dits Intra utilisent uniquement le codage Intra. Ils sont utilisés pour le codage de la première image d'une séquence, car aucune redondance temporelle ne peut être exploitée dans ce cas. Les images codées uniquement en Intra sont indépendantes des autres images du flux. Par conséquent, on les utilise aussi pour permettre un accès aléatoire à la séquence codée, ou pour réinitialiser la prédiction temporelle, en cas de pertes de paquets par exemple. Les slices P, dits prédictifs, utilisent à la fois le codage Inter et le codage Intra, car, pour chaque macrobloc, ces deux types de codage peuvent être en compétition. Les images P utilisent des images dans le passé pour la compensation de mouvement. Finalement les slices B, dits bidirectionnels, utilisent aussi ces deux modes de codages (Intra et Inter) mais la compensation de mouvement peut être effectuée à la fois sur des images dans le passé et des images dans le futur. De plus, pour ces slices B, la prédiction temporelle d'un bloc peut être la différence entre le bloc courant et un prédicteur résultant d'une moyenne entre deux blocs. Ces deux blocs peuvent provenir d'une image dans le passé et d'une image dans le futur, ou de deux images dans le passé ou de deux images dans le futur.

L'estimation de mouvement est le processus qui permet de réaliser une mise en correspondance entre des éléments de deux trames différentes, basée sur le critère de fidélité. Pour le bloc courant, ce processus a pour but de rechercher dans les images précédemment encodées un bloc de même taille représentant le même objet ou une même partie de l'objet. Ce processus de recherche du meilleur prédicteur temporel est effectué uniquement à l'encodeur, c'est donc un processus non-normatif. L'estimation de mouvement correspond à un algorithme de mise en correspondance de blocs ou "Block Matching" (BMA), avec comme critère de sélection la minimisation du critère débit-distorsion. Cependant, l'estimation de mouvement n'étant pas normative, le choix du critère

de sélection ou de l'algorithme dépend de chaque implémentation. Le codage Inter de la norme H.264/AVC offre un partitionnement variable du macrobloc. Les tailles de bloc possibles sont  $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ ,  $8 \times 8$ , et chaque bloc  $8 \times 8$  peut avoir un partitionnement  $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$  ou  $4 \times 4$ . Ce partitionnement variable du macrobloc permet de segmenter le mouvement lorsque c'est nécessaire, notamment pour les bords des objets. De plus, l'estimation de mouvement et la compensation de mouvement sous-pixellique sont utilisées dans le standard jusqu'au  $\frac{1}{4}$  de pixel, ceci afin de palier au caractère discret des images numériques alors que la capture se fait dans le domaine continu. Cette méthode a un impact sur le coût des vecteurs mouvement par rapport à une estimation de mouvement au pixel entier. En effet, si un vecteur indique une translation au pixel entier de  $(1, 1)$ , étant donné que les vecteurs parcourent le pixel entier, le  $\frac{1}{2}$  pixel et le  $\frac{1}{4}$  de pixel, le vecteur a pour valeur  $(4, 4)$ . Enfin, l'option des images de référence multiple [WZG99] a été intégrée au standard H.264/AVC. Dans cette méthode, deux blocs voisins peuvent avoir deux prédicteurs temporels dans deux images différentes. Une partie de nos contributions est dédiée à l'amélioration du codage des vecteurs mouvement. Nous proposons donc, dans la section 1.3, une description détaillée du codage de cette information.

Le filtre de deblocking (lissage des effets de bloc) [LJL<sup>+</sup>03] est un outil d'amélioration de qualité visuelle des images décodées, dégradées par les effets de bloc engendrés par la quantification des coefficients des résiduels transformés. Cependant, ce filtre est intégré dans la boucle de décodage à l'intérieur du codeur pour améliorer la qualité des images de référence qui deviennent plus "lisses" (moins d'effets de bloc) et permettent ainsi d'améliorer l'estimation de mouvement.

### 1.1.2 Codage du résiduel

Le codage des résiduels permet à la fois d'exploiter les redondances spatiales de notre signal et d'éliminer une partie de l'information, afin de réduire la quantité d'information à transmettre. Ces résiduels issus du codage Inter et Intra sont transformés, puis les coefficients générés sont quantifiés avant d'être codés avec un codage entropique. Dans cette section nous présentons ces trois étapes.

#### 1.1.2.1 Transformation

L'étape de transformation a pour objectif de décorréler le signal et de réduire le nombre de coefficients significatifs dans le bloc. Pour cela, on utilise une transformée par bloc permettant de mettre en évidence les fréquences spatiales du signal 2D. En effet, l'énergie des résiduels engendrés par le codage Inter et le codage Intra est concentrée dans les basses fréquences. Notons que dans la norme H.264/AVC on exploite, avec la transformée, des redondances spatiales contenues dans des résiduels de bloc et les redondances spatiales des blocs de l'image, comme dans le standard JPEG [JPEG].

Dans la majorité des standards de codage vidéo et d'image, la transformée utilisée pour décorréler le signal est la transformée DCT par bloc [ANR74]. Cette transformée est la transformée de Fourier sur la partie paire de l'image. La formule de la DCT pour un bloc de taille  $N \times N$  est définie par :

$$F(u, v) = \frac{2}{N} c(u)c(v) \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} x_{ij} \cos\left[\frac{\pi}{N} u\left(x + \frac{1}{2}\right)\right] \cos\left[\frac{\pi}{N} v\left(y + \frac{1}{2}\right)\right] \quad (1.3)$$

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

où  $x_{ij}$  est le pixel du bloc courant à la position  $(i, j)$  et  $F(u, v)$  le coefficient de la matrice transformée à la position  $(u, v)$ . Le standard H.264/AVC utilise une transformée entière. Cette transformée a les mêmes propriétés que la DCT classique. L'avantage est qu'elle permet de transmettre des coefficients entiers et non des coefficients en virgule flottante, ce qui permet à la fois de réduire le débit et d'éviter les problèmes d'implémentation liés à la précision, comme dans les standards précédents [H.263], [MPEG2]. Cette transformée, pour des blocs de taille  $4 \times 4$  est définie par :

$$F = C.X.C^T \quad \text{avec} \quad C = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \quad (1.4)$$

Comme on peut le voir, l'implémentation de cette transformée ne nécessite que des décalages et des additions, ce qui engendre une faible complexité [HK01] par rapport à la DCT classique de l'équation (1.3). Notons qu'une partie de la transformée est effectuée dans l'étape de quantification, ce qui permet d'obtenir l'équation (1.4) contenant uniquement des entiers compris entre -2 et 2. Enfin la norme H.264/AVC permet d'utiliser une transformée DCT de taille  $8 \times 8$  définie par :

$$F = C.X.C^T \quad \text{avec} \quad C = \begin{bmatrix} 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \\ 12 & 10 & 6 & 3 & -3 & -6 & -10 & -12 \\ 8 & 4 & -4 & -8 & -8 & -4 & 4 & 8 \\ 10 & -3 & -12 & -6 & 6 & 12 & 3 & -10 \\ 8 & -8 & -8 & 8 & 8 & -8 & -8 & 8 \\ 6 & -12 & 3 & 10 & 10 & -3 & 12 & -6 \\ 4 & -8 & 8 & -4 & -4 & 8 & -8 & 4 \\ 3 & -6 & 10 & -12 & 12 & -10 & 6 & -3 \end{bmatrix} \quad (1.5)$$

Cette transformée est plus complexe que la DCT  $4 \times 4$ , néanmoins sa complexité reste inférieure à celle de la DCT classique de l'équation (1.3).

### 1.1.2.2 Quantification

La quantification scalaire a pour but de réduire l'espace des valeurs des coefficients des résiduels transformés pour réduire l'entropie du signal. Cette opération  $Q$  consiste à diviser chaque coefficient du résiduel du bloc transformé par son coefficient de quantification provenant d'une matrice de quantification et à ne garder que la partie entière. Les matrices de quantification dépendent en général d'un pas de quantification (QP) relatif à la dégradation que l'on souhaite introduire sur les résiduels de bloc. Cette opération introduit naturellement des pertes d'information dans le signal. En

effet, l'opération de déquantification  $Q^{-1}$  ne permettra pas de retrouver le signal original. L'opération conjointe de quantification et de déquantification est donnée par la formule suivante :

$$Q^{-1}Q(x) \mapsto mq \quad \text{si } x \in [mq - \frac{q}{2}; mq + \frac{q}{2}[ \quad (1.6)$$

où  $x$  est un coefficient du résiduel transformé et  $q$  est le coefficient ayant la même position dans la matrice de quantification. Pour que la reconstruction du signal soit optimale, un "offset" de quantification est ajouté à chaque coefficient. Dans le logiciel de référence de la norme H.264/AVC, cet offset de quantification est fixé de manière adaptative en fonction de statistiques obtenues sur les blocs précédemment quantifiés.

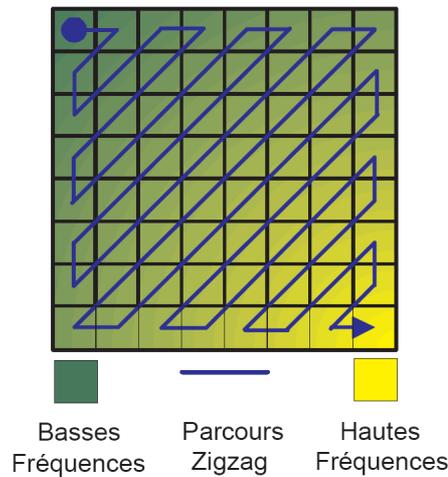
### 1.1.2.3 Codage entropique et arithmétique

A la suite de la quantification, la matrice du résiduel transformé contient beaucoup de valeurs égales à zéro. La matrice quantifiée va être parcourue par ordre croissant des fréquences. Notons que les basses fréquences se situent en haut à gauche et les hautes fréquences se situent en bas à droite du bloc des coefficients, comme représenté dans la figure 1.2. A la suite de ce parcours dit en zigzag on obtient un vecteur unidimensionnel qui va être codé à l'aide de l'algorithme de Run Length Coding (RLC). Cet algorithme remplace les coefficients égaux à zéro par la distance entre deux valeurs différentes de zéro. Le nouveau vecteur généré par le RLC est codé soit à l'aide du CAVLC (Context Adaptive Variable Length Coding), soit à l'aide du CABAC (Context Adaptive Binary Arithmetic Coding) en fonction du profil de la norme utilisé. Ces deux codages sont adaptés au contexte, ce qui signifie que pour chaque composante (résiduel, mode, vecteur, etc.) le contexte utilisé dépend des informations précédemment décodées. Par exemple, pour le codage arithmétique du module CABAC [MSW03], les probabilités utilisées pour le codage de la valeur courante dépendent de la composante courante et des valeurs voisines de cette même composante déjà décodée. Les probabilités de chaque contexte sont fixes dans la norme. Le nombre et les valeurs de ces probabilités possibles dépendent des composantes.

Dans cette section nous avons volontairement étudié uniquement le schéma du standard H.264/-AVC. Toutefois, précisons qu'il existe plusieurs codeurs normalisés : les prédécesseurs de cette norme, mais aussi les concurrents comme VC1 [VC106] ou AVS [FMW]. De même, le domaine de la recherche en compression vidéo étant vaste, de nombreux codeurs vidéo ont été développés. Ces codeurs ont parfois des schémas de codage différents de celui de H.264/AVC, néanmoins l'exploitation des redondances spatiales et/ou temporelles reste l'objectif recherché.

## 1.2 Normes de compression en codage d'images fixes

Le codage Intra dans les normes vidéo est similaire au problème du codage d'images fixes. D'ailleurs, certains standards vidéo utilisent un codage équivalent à l'incontournable norme JPEG [JPEG] pour la compression des images Intra. Dans cet état de l'art nous n'allons pas détailler l'ensemble des techniques de codage d'images fixes. Nous nous attarderons sur le codage Intra de la norme H.264/AVC et sur les propositions d'amélioration de ce codage par prédiction, avant de donner un bref aperçu de la norme de référence en codage d'image, JPEG 2000 [JPEG2000].

FIGURE 1.2 – Parcours zigzag par ordre de fréquences d'un bloc de coefficients DCT  $8 \times 8$ .

### 1.2.1 Codage par prédiction

Le codage prédictif Intra exploite les redondances spatiales conjointement avec une prédiction et une transformation du résiduel engendré. Il peut être défini par quatre attributs principaux : le partitionnement du macrobloc (forme et taille des partitions), l'ordre de parcours des partitions, les prédicteurs utilisés pour calculer le résiduel et naturellement la transformation de ce résiduel. Nous étudierons donc dans cette section les différentes possibilités proposées dans la littérature pour ces quatre attributs.

#### 1.2.1.1 Partitionnement

Le codage Intra de la norme H.264/AVC utilise trois partitionnements du macrobloc :  $16 \times 16$ ,  $8 \times 8$  et  $4 \times 4$  comme montré dans la figure 1.3. Ces trois partitionnements sont mis en compétition au niveau macrobloc. Dans [Wie01] les partitionnements utilisés pour le codage Inter de la norme H.264/AVC ( $16 \times 16$ ,  $16 \times 8$ ,  $8 \times 16$ ,  $8 \times 8$ ,  $8 \times 4$ ,  $4 \times 8$ ,  $4 \times 4$ ) ont été proposés pour le codage Intra. Dans cette proposition, le codage Intra dispose du même partitionnement que le codage Inter, améliorant ainsi l'efficacité de ce codage grâce à la multiplication des possibilités de codage. Dans [ZZM<sup>+</sup>04] les auteurs proposent de découper le macrobloc en deux parties égales, en fonction des contours traversant ce macrobloc. Ce découpage peut être vertical ou horizontal. Finalement dans [DEY<sup>+</sup>07] deux nouveaux modes, de taille  $16 \times 16$  et  $8 \times 8$ , sont ajoutés aux modes Intra de la référence. Ces nouveaux modes contiennent des partitions géométriques découpées à l'aide de lignes paramétriques, le codage de cette segmentation pour les modes Inter ayant été proposé dans [BDB05]. De plus, les auteurs utilisent une interpolation linéaire des pixels adaptée à ces partitions géométriques. Ce travail est intéressant puisqu'il n'est pas basé sur un partitionnement carré ou rectangulaire du macrobloc comme dans les standards et la majorité des propositions. Ces partitionnements améliorent le schéma de référence.

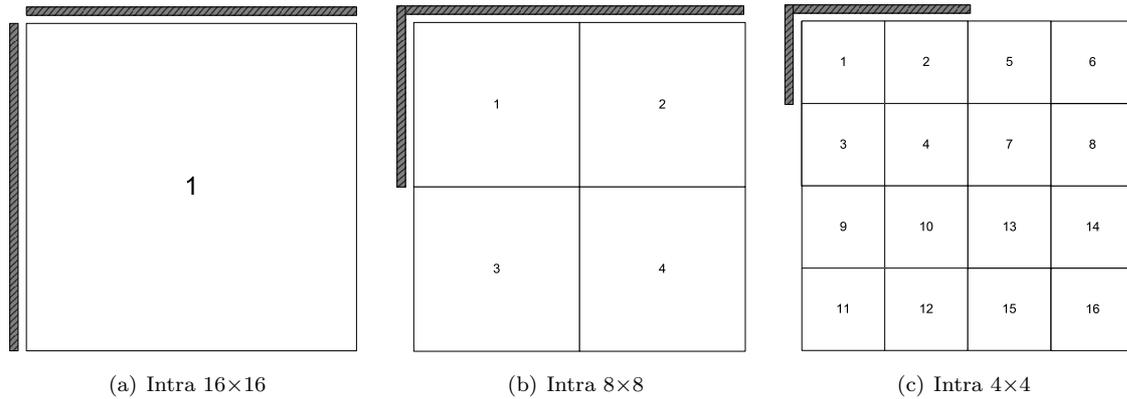


FIGURE 1.3 – Partitionnements de l’Intra de la norme H.264/AVC. Le signal de référence pour le premier bloc est représenté en gris. Les numéros à l’intérieur des blocs correspondent à l’ordre de parcours des partitionnements.

### 1.2.1.2 Ordre de parcours des partitions

La norme H.264/AVC utilise un ordre lexicographique dit “raster” pour le parcours des blocs  $8 \times 8$  et  $4 \times 4$  d’un macrobloc, comme nous pouvons le voir dans la figure 1.3. Cet ordre permet uniquement d’avoir un signal de référence en haut et à gauche. La méthode BIP (Bidirectional Intra Prediction) [STC07] utilise un ordre de parcours qui commence par le bloc le plus en bas à droite d’un groupe de quatre blocs, ce qui permet d’utiliser des prédictions bidirectionnelles. Cet ordre proposé pour les blocs  $4 \times 4$  et  $8 \times 8$  est représenté dans la figure 1.4. Dans cette méthode, la distance spatiale entre les pixels du bloc 1 et les pixels de référence est plus élevée que dans la norme H.264/AVC. Cependant les pixels de référence du bloc 2 proviennent de 3 directions (en haut, à gauche et en bas), de même pour le bloc 3 (en haut, à gauche et à droite). Pour le bloc 4, les pixels de référence entourent le bloc, permettant ainsi d’augmenter l’efficacité de la prédiction Intra. Ce concept d’ordres de parcours multiples pour une partition a été décrit dans [GLH04]. Les auteurs proposent d’utiliser cette méthode pour un partitionnement composé de quatre blocs, ce qui représente 24 ordres de parcours possibles. Cependant seulement deux ou quatre ordres de parcours sont utilisés dans les résultats expérimentaux. Notons que cette description inclut le cas de la méthode BIP. Dans [LHS06] les auteurs proposent d’utiliser la méthode DPCM (Differential Pulse Code Modulation) pour chaque bloc  $4 \times 4$ . Pour cette méthode, chaque pixel est prédit à l’aide d’un pixel précédemment décodé, horizontal ou vertical en fonction de la prédiction choisie. Cette technique peut seulement être appliquée au codage sans perte. Cette méthode est judicieuse car elle utilise les pixels voisins du pixel à prédire, mais elle ne peut pas être utilisée pour un codage avec pertes. Dans [Tom07] l’image est “retournée” verticalement et/ou horizontalement engendrant ainsi quatre images possibles. L’image donnant le meilleur coût débit-distorsion est alors sélectionnée. L’efficacité de la méthode repose sur les caractéristiques asymétriques de la prédiction Intra. Le meilleur gain est obtenu sur une image contenant uniquement des rayures que les huit directions des prédicteurs des modes Intra  $4 \times 4$  et  $8 \times 8$  ne peuvent prédire efficacement. Il serait intéressant de

tester cette méthode au niveau bloc.

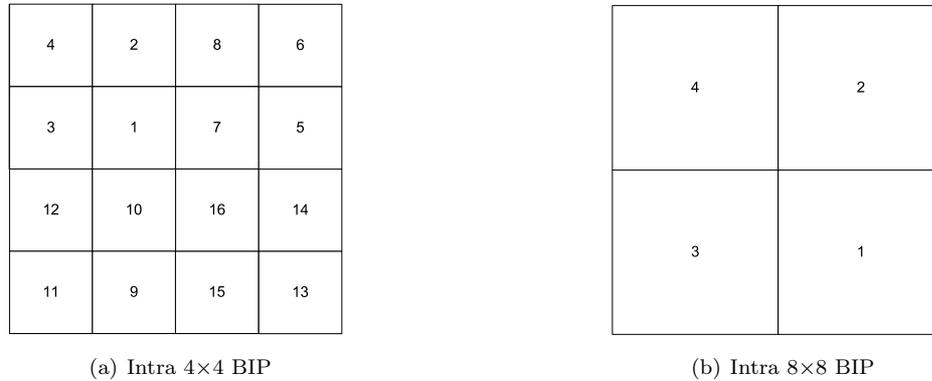


FIGURE 1.4 – Ordre de parcours de l'Intra 8×8 et l'Intra 4×4 dans la méthode BIP.

### 1.2.1.3 Prédicteurs Intra

Le calcul des prédicteurs est l'attribut le plus étudié dans la littérature. Rappelons que la norme H.264/AVC utilise quatre prédicteurs pour le partitionnement 16×16 et neuf prédicteurs pour les partitionnements 4×4 et 8×8. Ces prédicteurs sont calculés avec la ligne en haut et la colonne de gauche du bloc courant qui appartiennent à des blocs déjà décodés. Pour le mode 16×16, le prédicteur vertical est une copie de la ligne au-dessus du macrobloc pour chacune des lignes du prédicteur. De même le prédicteur horizontal est une copie de la colonne de gauche dans chacune des colonnes du prédicteur. Le troisième prédicteur (DC) est la moyenne de la ligne du haut et de la colonne de gauche. Ce prédicteur correspond à une prédiction, dans le domaine transformé, du premier coefficient du bloc. Enfin le prédicteur plane est la moyenne, pixel à pixel (dans le sens diagonal), de la ligne du haut et la colonne de gauche. Les neufs prédicteurs Intra, représentés dans la figure 1.5, sont le prédicteur DC et huit autres prédicteurs correspondant à huit directions, incluant les directions verticale et horizontale. Les prédicteurs pour les blocs de taille 8×8 sont construits de la même manière. Les indices des prédicteurs Intra 4×4 et 8×8 subissent eux aussi une forme de prédiction. En effet, si le prédicteur est égal au prédicteur le plus probable, un seul bit sera transmis au décodeur dans le cas du CAVLC ou un seul bit sera utilisé pour la binarisation dans le cas du CABAC. Si le prédicteur n'est pas égal à ce prédicteur le plus probable, la valeur du prédicteur sera codée ou binarisée avec 4 bits. Le prédicteur le plus probable est calculé en fonction des deux blocs voisins en haut et à gauche.

En considérant que l'outil de compensation de mouvement sous-pixelique est efficace pour l'exploitation des redondances temporelles, dans [PPJ07] de nouveaux prédicteurs sont générés à l'aide des filtres utilisés pour la création du demi pixel. Cependant cette méthode engendre des gains faibles sans même que l'information liée à ces nouveaux modes de codage n'ait été codée. Dans [MTKY07a] l'analogie avec les multiples images de référence pour la compensation de mouvement est faite pour

les prédicteurs spatiaux. Les prédicteurs sont calculés avec plusieurs lignes ou colonnes de référence au lieu d'une seule ligne et une seule colonne voisines dans la référence. Dans [MTKY07b] les auteurs proposent des prédictions Intra pondérées, générées avec un gradient. Dans [TYTA07] différents ensembles de directions pour les prédicteurs Intra sont utilisés. Ces directions sont ajoutées entre les huit directions des modes Intra  $4 \times 4$  et  $8 \times 8$ . Cette méthode augmente le nombre de prédicteurs pour la compétition et les gains obtenus sont plus encourageants (près de 3% en profil Baseline Intra) que ceux obtenus avec les méthodes [PPJ07] et [MTKY07a]. L'augmentation du nombre de prédicteurs pour le codage Intra a aussi été étudié dans [ZZM<sup>+</sup>04] pour le standard vidéo AVS. Ce standard utilise, comme H.264/AVC, une prédiction spatiale pour le codage des images Intra. Les prédicteurs ajoutés proviennent de trois techniques : 32 directions contre 8 pour la référence, des résolutions multiples dans le même esprit que l'article [PPJ07], des lignes de références multiples comme dans [MTKY07a].

A l'opposé de ces travaux qui multiplient le nombre de prédictions possibles, les auteurs du papier [MT05] proposent d'ajouter un mode de taille  $16 \times 16$  qui prend en considération pour chaque bloc  $4 \times 4$  les blocs  $4 \times 4$  précédemment décodés, ce qui correspond au mode Intra  $4 \times 4$  avec un seul prédicteur pour tout le macrobloc au lieu des 16 nécessaires pour la référence. Lorsque l'on étudie la sélection des modes Intra, ce nouveau mode remplace à la fois le mode  $16 \times 16$  classique mais aussi le mode  $4 \times 4$ . Cela signifie que le débit des prédicteurs de l'Intra  $4 \times 4$  est parfois trop élevé par rapport aux possibilités de codage offertes.

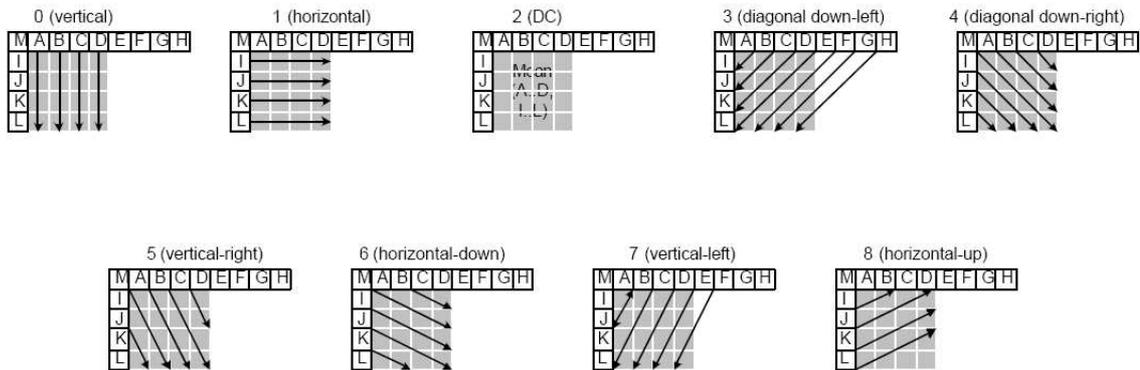
D'autres méthodes proposent d'améliorer le schéma de prédiction à l'aide d'un raffinement de l'erreur de prédiction. Dans [HNEY<sup>+</sup>07], le choix d'utiliser ou non le raffinement basé sur des caractéristiques des blocs voisins est fait à l'aide du critère débit-distorsion bloc par bloc ou macrobloc par macrobloc. Dans [YNK07] une DCT 1D est utilisée pour calculer l'énergie des blocs voisins. Cette énergie est ensuite ajoutée aux prédicteurs Intra utilisés. L'idée est de garder l'énergie des blocs voisins qui est perdue avec le processus de construction des prédicteurs Intra.

Dans la recherche de prédicteurs toujours plus élaborés, les auteurs de [YC02] et [KSK04] proposent, en analogie avec la compensation de mouvement du codage Inter, l'utilisation de prédicteurs provenant d'une recherche équivalente à celle de l'estimation de mouvement. Les vecteurs provenant de cette recherche sont codés de la même manière que ceux des modes Inter. Dans [KTCS06] un "template matching" est utilisé pour la création de nouveaux prédicteurs pour le codage Intra.

Finalement les partitionnements et les ordres de parcours de partitions proposés dans [DEY<sup>+</sup>07], [GLH04] et [STC07] requièrent la création de prédicteurs adaptés à ces modifications.

#### 1.2.1.4 Transformée

Les résiduels Intra de la norme H.264/AVC peuvent être transformés, comme les résiduels Inter, soit avec une DCT  $4 \times 4$  soit avec une DCT  $8 \times 8$ . Dans [Wie01] la taille de la transformée DCT est adaptée à chaque taille des nouvelles partitions proposées pour le codage Intra. Le même auteur a aussi proposé [Wie03] d'utiliser une taille de transformée adaptée au partitionnement du macrobloc pour le codage Inter avec succès. Enfin dans [Kar07], un ensemble de transformées KLT est associé à chaque direction de prédiction (à chaque prédicteur de l'Intra). Dans ce cas, la transformée est adaptée à la direction exploitée par le prédicteur sélectionné.

FIGURE 1.5 – Les neuf prédicteurs des blocs Intra  $4 \times 4$  de la norme H.264/AVC.

La transformée est une étape incontournable en codage d'images. Cet outil est intégré dans les normes de compression d'image sans même utiliser un module de prédiction. Comme nous allons le voir dans la section suivante, ces codages peuvent parfois être plus efficaces que le codage Intra d'H.264/AVC.

## 1.2.2 Les standards de codage d'images

Le groupe JPEG (Joint Picture Expert Group) JTC1/SC29/WG1 a été formé par l'ITU et l'ISO dans le but de collaborer à la normalisation d'un standard de codage d'images. Ce groupe a donné naissance à la norme très utilisée du codage d'images fixes JPEG [JPEG] en 1992. Cette norme définit une segmentation de l'image en blocs de taille  $8 \times 8$ . Ces blocs sont ensuite transformés avec la DCT, puis les coefficients sont quantifiés et codés à l'aide d'un codeur entropique. On retrouve une partie du schéma global de cette norme dans le schéma de la norme H.264/AVC. En effet, pour les slices I, si l'on enlève le module de codage prédictif Intra, on obtient le même schéma (découpage en bloc, transformé, quantification, codage entropique).

Le groupe JPEG a normalisé le standard JPEG 2000 en 2001 [JPEG2000]. Pour cette norme, le groupe s'est fixé, en plus de l'objectif de proposer un standard plus efficace que JPEG, des contraintes sur le format des images décodables, et sur le type de contenu de ces images. JPEG 2000 a été conçu pour pouvoir coder des images naturelles, médicales, des images de texte, des graphiques etc. L'idée est d'avoir une seule norme pour toutes les applications. Ce standard comporte un mode de compression sans perte en plus du mode de compression avec pertes. Il est possible de définir des régions d'intérêt (ROI). Son principal atout est sa scalabilité en résolution. C'est-à-dire que d'un même flux JPEG 2000 on peut extraire plusieurs résolutions d'une même image sans décoder l'image dans sa plus haute résolution.

Le codage de JPEG 2000 s'effectue en cinq étapes. La première est une étape de prétraitement consistant à découper l'image en "tuiles" puis à décorréliser les composantes de couleur de ces tuiles à l'aide d'une transformée. Le découpage en tuiles permet de réduire l'espace mémoire durant l'application de l'algorithme, en segmentant l'exécution tuile par tuile. L'autre utilité est de permettre de

décoder uniquement une partie de l'image. Le découpage en tuiles engendre une baisse de l'efficacité globale de codage. La norme JPEG 2000 permet donc d'utiliser une seule tuile par image.

La deuxième étape transforme les pixels des tuiles avec la transformée en ondelettes discrète (DWT pour Discrete Wavelet Transform). Cette transformée permet une représentation multirésolution. La transformée intégrée dans JPEG 2000 est la transformée en ondelettes de Daubechies avec des filtres 9/7 [ABMD92] pour le codage avec pertes et la transformée en ondelettes réversible avec des filtres 5/3 pour le codage sans perte. Dans cette transformée, chaque tuile est d'abord filtrée avec un filtre passe-bas et un filtre passe-haut sur les lignes, suivi d'un sous-échantillonnage d'un facteur deux. A cette étape nous obtenons deux signaux qui sont chacun filtrés et sous-échantillonnés passe-haut et passe-bas sur les colonnes. Nous obtenons alors quatre sous-bandes, dont les coefficients représentent les fréquences verticales et horizontales du signal original. La sous-bande filtrée passe-bas en lignes puis passe-bas en colonnes donne une imagerie d'approximation. Les trois autres sous-bandes contiennent les détails horizontaux, verticaux et diagonaux, comme illustré dans la figure 1.6. Pour la décomposition en niveaux supérieurs, on appliquera le même procédé sur les coefficients d'approximation.

Comme les autres standards de codage d'image et de vidéo avec pertes, la norme JPEG 2000 utilise une étape de quantification. Dans cette étape, les coefficients des sous-bandes sont quantifiés uniformément avec une zone morte centrale. Chaque sous-bande peut avoir son propre coefficient de quantification adapté au contenu [Use01]. Les coefficients quantifiés de chaque sous-bande sont découpés en blocs et les coefficients de ces blocs sont ensuite codés avec un codage arithmétique. Enfin, les bitstreams de chacun de ces blocs sont rangés dans le bitstream final.

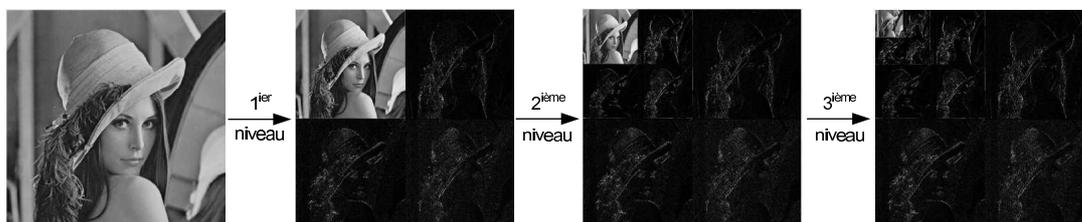


FIGURE 1.6 – Décomposition en ondelettes de l'image *Lena* sur trois niveaux avec un banc de filtres 9/7.

Le standard JPEG 2000 offre des gains significatifs par rapport à son prédécesseur JPEG et propose de nombreuses fonctionnalités nouvelles. Les gains obtenus par rapport à JPEG sont de environ 10% à haut débit et ils atteignent 50% à bas débit [CC00]. Ce codeur d'image a modifié l'étape fondamentale de JPEG : la transformée. Comme on l'a vu précédemment, le codage de la norme H.264/AVC propose un codage prédictif des blocs dans le domaine pixel. Cependant les autres étapes de codage sont sensiblement équivalentes au codage de la norme JPEG. Il est alors intéressant de comparer ces deux normes afin d'évaluer ces deux différentes stratégies de codage.

Les études comparatives entre JPEG 2000 et le codage Intra H.264/AVC sont nombreuses et obtiennent généralement les mêmes conclusions. Dans [MWG05] le profil High de la norme H.264/AVC

est comparé au standard JPEG 2000 sur des images monochromes. Les performances des deux codeurs sont identiques pour cette configuration. Cependant, si la transformée DCT  $8 \times 8$  est désactivée, le standard JPEG 2000 obtient de meilleures performances. Des tests identiques ont été présentés dans [GMW04] et [Top05] sur un ensemble de séquences HD et rapportent une supériorité du codage Intra du profil High par rapport à JPEG 2000 aux alentours de 0.2-1.0 dB. Dans [ODE06] et [ODE07] JPEG 2000 est comparé au profil Main et au profil High de H.264/AVC, utilisant uniquement des slices Intra (nous noterons dans la suite ces configurations Main III et High III). JPEG 2000 obtient de meilleurs résultats que le Main III pour les séquences de haute résolution. Cependant, pour les séquences de basse résolution, le Main III et le High III sont systématiquement plus efficaces que JPEG 2000. Notons que dans [ODE07] les outils utilisés du profil High III et de JPEG 2000 ont été choisis pour obtenir la meilleure efficacité de codage pour chacune des deux normes. Le profil High III est meilleur en moyenne que JPEG 2000 pour toutes les résolutions. Néanmoins, pour certaines séquences JPEG 2000 est plus efficace. Le standard MJPEG 2000 [MJPEG2000], qui est l'extension de JPEG 2000 pour les séquences vidéo, est comparé au profil Main III de la norme H.264/AVC et au codeur PACC dans [MGC03]. Le codeur PACC [MC97] est un codeur d'image et de vidéo basé sur les ondelettes. Notons que MJPEG 2000, comme son prédécesseur MJPEG [MJPEG], exploite uniquement les redondances spatiales dans la vidéo. Des résultats similaires à l'état de l'art sont obtenus. Le codeur PACC obtient une meilleure efficacité de codage que le profil Main III pour ces basses résolutions. Il aurait été intéressant de le comparer au profil High III.

Pour que l'étude soit complète, il faut aussi comparer l'efficacité en termes de complexité de calcul. Dans [ARK<sup>+</sup>04] les deux normes sont comparées en termes de débit-distorsion avec des mesures objectives (PSNR) et subjectives (MOS pour "Mean Opinion Score") de la qualité. Les auteurs proposent aussi une étude de complexité au niveau du décodeur pour le JM 7.3 [JM] de la norme H.264/AVC en donnant le nombre d'opérations pour chacune des étapes du décodage. L'étude de complexité de JPEG 2000 est faite sur l'implémentation Jasper [Jasper]. La conclusion obtenue par les auteurs est que l'Intra d'H.264/AVC est plus efficace que JPEG 2000 en termes de qualité objective et subjective, néanmoins JPEG 2000 a une complexité de calcul moindre. Cette conclusion est un peu hâtive puisque le JM n'est pas un code optimisé. Le but de ce logiciel est seulement informatif, il ne reflète absolument pas la complexité réelle de la norme. De même, l'implémentation Jasper ne reflète pas la complexité de calcul de JPEG 2000 puisqu'il est, lui aussi, un logiciel à but informatif.

Dans le chapitre 3 nous allons, pour confirmer les résultats de l'état de l'art, comparer l'efficacité du codage de la norme H.264/AVC pour le profil Baseline III et le profil High III avec la norme JPEG 2000.

## 1.3 Etat de l'art sur le codage de l'information de mouvement

### 1.3.1 Représentation du mouvement

En compression vidéo, afin d'exploiter au mieux les redondances temporelles entre les images successives, on introduit une notion de mouvement pour caractériser les différences entre ces images. Il existe différentes manières de coder cette information de mouvement. La manière la plus intuitive est sans doute d'affecter un vecteur déplacement (vecteur mouvement en coordonnées cartésiennes) à chaque objet d'une séquence vidéo. Ces méthodes ont été largement étudiées et intégrées dans le standard MPEG-4 [MPEG4]. Dans ce type de codage, chaque élément d'une scène est codé indépendamment : on sépare le fond des objets en mouvement. Dans chaque image, un objet est représenté par la déformation de son contour et par un vecteur mouvement traduisant son déplacement. Cependant ces méthodes n'ont pas été utilisées par les industriels malgré leur intégration dans le standard MPEG-4 car la segmentation de la scène en objets est difficile à obtenir avec des algorithmes rapides. De plus, le fond d'une séquence est rarement statique à cause des défauts des capteurs, et le débit du codage des contours et de leurs déformations n'est pas négligeable.

Une stratégie opposée au codage par objet est le codage par champs denses qui consiste à utiliser un vecteur mouvement par pixel. Bien entendu, dans cette méthode, le coût de l'information de mouvement est élevé. Les algorithmes de compression utilisant les champs denses regroupent les vecteurs mouvement afin de limiter ce coût. Dans [HP01] le champ dense est calculé à l'aide d'une estimation de mouvement basée sur les champs de Markov (MRF) [KD92]. Les vecteurs lisses sont regroupés (en fonction de leur direction et de leur sens) puis un codage prédictif (cf. 1.3.3) exploitant les redondances spatiales du champ de vecteurs est appliqué. Dans [DH98] les auteurs proposent d'utiliser l'algorithme de Horn et Shunck [HS81] pour calculer les champs denses de vecteurs mouvement. Cet algorithme hiérarchique produit un champ de vecteurs pour chaque "niveau" de résolution ; les niveaux s'échelonnant de 1 vecteur pour toute l'image à 1 vecteur par pixel. Pour chaque niveau de résolution, les frontières dans le champ de vecteurs sont détectées afin de segmenter le mouvement. Les différentes zones sont ensuite lissées à l'aide d'une opération morphologique.

L'information de mouvement peut aussi être représentée par maillage. Pour la méthode de Triangle Motion Compensation (TMC) [NH91], [Bru90] l'image est découpée en triangles. Le mouvement est alors représenté par la déformation de ces triangles d'une image à l'autre. Cette déformation correspond à une transformation 2D affine. Dans [Cam04] une représentation du mouvement par maillage est proposée conjointement avec une transformée en ondelettes  $t+2D$ .

Enfin la représentation par bloc du mouvement est la plus utilisée dans les standards vidéo. Ce partitionnement des images implique le codage d'un vecteur mouvement en coordonnées cartésiennes pour chaque bloc. Ces algorithmes, à taille de bloc fixe, ne nécessitent pas de codage de la segmentation du mouvement et sont adaptés à la transformée carrée utilisée dans la quasi-totalité des standards d'image fixe et vidéo : la DCT 2D. Le défaut de cette méthode est qu'elle n'exploite pas les corrélations spatiales du champ de vecteurs à l'intérieur des objets. Dans [CYC90] les vecteurs issus d'un BMA sont regroupés à l'aide d'un algorithme de Split and Merge. La segmentation

engendrée est ensuite codée comme un arbre pour lequel chaque noeud contient un vecteur mouvement. Enfin la norme H.264/AVC [H.264/AVC] applique une segmentation du macrobloc allant du  $16 \times 16$  jusqu'au  $4 \times 4$ . Cependant le codage du mouvement n'est pas hiérarchique comme dans [DH98], [CYC90] mais chaque bloc a un vecteur.

### 1.3.2 Codage des vecteurs mouvement avec pertes

Dans les standards de codage d'image et de vidéo avec pertes, on quantifie uniquement l'information de texture afin de réduire l'entropie de cette information. Le codage des vecteurs mouvement avec pertes [JFB95], [YYW95], [DSCW00] a pour philosophie de traiter le mouvement en tant qu'information quantifiable, comme l'information de texture. La proportion de l'information de mouvement peut être plus élevée que celle liée à la texture, notamment à bas et très bas débit. Dans [AAAB05] une quantification de vecteurs mouvement de très grande précision pixellique est introduite dans un codeur ondelettes. Le vecteur mouvement quantifié, transmis au décodeur, est  $\tilde{mv} = \text{round}(mv/Q_{mv})$  où  $Q_{mv}$  est le pas de quantification attribué au vecteur courant  $mv$ . Le codage est fait en boucle ouverte ce qui signifie que la quantification des vecteurs mouvement est faite après la compensation de mouvement. Par conséquent la reconstruction exacte par le décodeur n'est pas possible. Cependant les applications visées sont les bas débits, pour lesquels la reconstruction exacte n'est pas essentielle [ABMD92]. Afin d'améliorer le schéma, dans [AAB06] une allocation de débit optimum entre le débit alloué au codage de la texture et celui des vecteurs mouvement quantifiés est proposée. Ce schéma de compression des vecteurs a été testé dans H.264/AVC [CAA<sup>+</sup>09] en boucle fermée. Pour chaque bloc, un pas de quantification du vecteur ( $Q_{mv}$ ) est sélectionné à l'aide du critère débit-distorsion. L'équation prend en compte le codage prédictif des vecteurs utilisés dans la norme H.264/AVC. Par conséquent, le prédicteur du vecteur est lui aussi quantifié. Cette méthode nécessite la transmission d'un  $Q_{mv}$  pour chaque vecteur, ce qui engendre une nouvelle information de codage très coûteuse. En pratique, la sélection du  $Q_{mv}$  est effectuée image par image. Les résultats préliminaires sont encourageants.

### 1.3.3 Codage prédictif des vecteurs mouvement

Le coût de l'information de mouvement, pour les algorithmes de compression utilisant une compensation de mouvement par bloc, dépend de trois paramètres :

- la taille des blocs utilisés (plus les blocs sont de petite taille, plus le nombre de vecteurs à coder est élevé).
- la résolution sous-pixellique utilisée pour la compensation de mouvement (la valeur d'un vecteur au  $\frac{1}{4}$  de pixel est multipliée par quatre).
- l'entropie de l'information.

Pour réduire l'entropie de l'information, on utilise généralement un codage prédictif. Dans ce codage on considère non plus l'entropie de la source mais celle des résiduels de cette source. Pour le codage des vecteurs mouvement le résiduel  $\varepsilon_{mv}$  va être transmis à la place du vecteur  $mv$ . Le résiduel  $\varepsilon_{mv}$  est donné par :

$$\varepsilon_{mv} = mv - p \quad (1.7)$$

où  $p$  est le prédicteur à transmettre. L'efficacité de la méthode dépend de la pertinence du prédicteur utilisé. Dans le cas des algorithmes hiérarchiques [DH98], [CYC90] chaque vecteur mouvement peut être prédit par la valeur du vecteur parent. Dans les standards vidéo, afin d'exploiter les redondances spatiales des champs de vecteurs mouvement, la valeur du prédicteur dépend de la valeur des vecteurs voisins déjà encodés/décodés. Ce prédicteur correspond à un médian (cf. section 1.3.4). Dans [Ebr94] le prédicteur est le module de ces vecteurs. Dans [Kri97] le prédicteur est le vecteur qui a le plus d'occurrences dans une fenêtre autour du vecteur à prédire. Les redondances temporelles entre les champs de vecteurs mouvement de deux images ont aussi été exploitées [ZZ91], [YVK95]. Dans [YVK95] les résultats obtenus pour des séquences contenant des champs de vecteurs mouvement complexes sont meilleurs que pour les séquences contenant des mouvements simples. Pour ces séquences, l'utilisation d'un prédicteur spatial est plus efficace. Pour réduire la quantité d'information, le codeur VC1 [VC106], transmet en plus de l'erreur de prédiction du vecteur, un drapeau indiquant quelle résolution pixellique a été utilisée.

D'autres méthodes plus évoluées choisissent parmi plusieurs prédicteurs. La méthode définie dans [HP01] propose un médian des vecteurs voisins appartenant à une même classe, ce qui requiert un stockage de la classe utilisée pour la prédiction de chaque vecteur. La segmentation des classes de prédicteurs dépend des différents mouvements. L'hypothèse est que les prédicteurs spatiaux sont inefficaces dans le cas d'une rupture du mouvement dans le champ de vecteurs. Cette hypothèse est aussi vérifiée dans [YVK95], avec l'utilisation de prédicteurs temporels. Pour réduire la complexité de l'estimation de mouvement dans [DNG01], une compétition entre un nombre élevé de prédicteurs est effectuée et l'indice du prédicteur est transmis au décodeur en plus de l'erreur de résiduel. Dans [CW97], les redondances spatiales et temporelles sont exploitées pour améliorer le codage des vecteurs. Le choix entre deux types de prédicteurs dépend des statistiques locales, ce qui ne permet pas un choix optimal entre les deux corrélations. Une compétition exhaustive entre un prédicteur spatial et un prédicteur temporel est utilisée dans [TWL05] mais uniquement pour le mode Direct des images B. Finalement une compétition exhaustive entre trois vecteurs voisins spatiaux est mise en oeuvre dans [DKBR99]. Dans cette méthode, seule la redondance spatiale est exploitée, ce qui peut être inefficace lors d'une grande disparité dans le champ de vecteurs mouvement. Dans ce cas, une prédiction temporelle devrait être plus efficace.

### 1.3.4 Codage de l'information de mouvement dans H.264/AVC

#### 1.3.4.1 Le prédicteur médian

La norme H.264/AVC utilise un codage prédictif des vecteurs mouvement. Le prédicteur  $p$  de l'équation (1.7) est un médian spatial pour chacune des composantes (horizontale et verticale). Nous noterons ce prédicteur  $mv_{H.264}$ . Les trois vecteurs voisins du vecteur courant  $mv$ , utilisés pour le calcul du médian sont  $mv_a$ ,  $mv_b$ ,  $mv_c$ , représentés dans la figure 1.7. En fonction de la taille des blocs voisins et du bloc courant, le vecteur  $mv_c$  peut-être remplacé par le vecteur  $mv_d$ . Pour des cas particuliers, dépendant des caractéristiques des blocs voisins, le prédicteur  $p$  peut être égal à  $mv_a$

$mv_d$	$mv_b$	$mv_c$
$mv_a$	$mv$	

FIGURE 1.7 – Localisation des vecteurs spatiaux et temporels utilisés pour la prédiction du vecteur courant  $mv$ .

ou  $mv_b$  ou  $mv_c$  ou 0. Ces caractéristiques sont l'appartenance de ces blocs à l'image, la taille du bloc courant et des blocs voisins et les images de référence utilisées pour le codage des blocs voisins. Par exemple, si un seul des vecteurs voisins a la même image de référence que le vecteur courant, la valeur de  $p$  est égale à ce vecteur voisin. De même, si l'un des blocs  $a$ ,  $b$ ,  $c$ , ou  $d$  est codé en Intra (modes pour lesquels il n'y a pas de vecteur mouvement) le vecteur mouvement pour ce bloc est égal à 0.

Les slices B (cf. section 1.1) utilisent des images dans le futur et le passé pour la compensation de mouvement [FG03]. De plus, il est possible d'utiliser des prédictions bidirectionnelles qui sont des combinaisons linéaires de deux compensations de mouvement impliquant l'utilisation de deux vecteurs mouvement (un vecteur par prédicteur de bloc). Les images de référence pour une prédiction bidirectionnelle peuvent se trouver dans le futur et le passé ou toutes dans le futur ou toutes dans le passé. Pour le médian, ceci ajoute une contrainte supplémentaire de direction (future ou passée ou les deux) pour les vecteurs voisins  $mv_a$ ,  $mv_b$ ,  $mv_c$  et  $mv_d$ .

#### 1.3.4.2 L'information de mouvement pour le mode Skip

Le mode Skip est un mode particulier du codage Inter pour les slices P (cf. section 1.1). Un macrobloc codé avec ce mode n'a ni résiduel de texture ni résiduel de vecteur mouvement, la seule information transmise est le mode (mode Skip). Cependant le Skip a un vecteur mouvement. Ce vecteur mouvement est le vecteur  $mv_{H.264}$  pour un bloc Inter  $16 \times 16$  défini précédemment, excepté les cas où les blocs  $a$  et  $b$  n'appartiennent pas à l'image. En effet, le mode Skip est censé servir pour les zones d'une séquence ne contenant pas ou peu de mouvement (généralement le mouvement a une probabilité d'apparition plus élevée dans le centre des images que dans les bordures). De même si un des vecteurs  $mv_a$ ,  $mv_b$  est égal à 0, la valeur du vecteur pour le mode Skip est égale à 0. Dans ce cas, on force le vecteur à être égal à 0, afin de favoriser la sélection du mode Skip si le mouvement est nul.

#### 1.3.4.3 L'information de mouvement pour le mode Direct

Le mode Skip des images P a son équivalent pour les images B : le mode Direct [TWL05]. Dans la norme H.264/AVC il existe deux types de mode Direct : le spatial et le temporel. Ce mode a deux vecteurs mouvement :  $mv_1^{L0}$  et  $mv_1^{L1}$  représentés dans la figure 1.8.  $L0$  et  $L1$  font respectivement référence aux "listes" (images de référence P) passée et future de la figure 1.8.

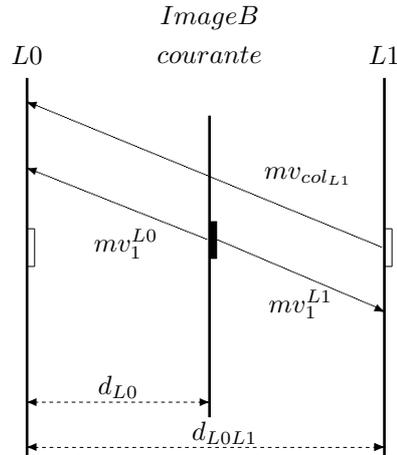


FIGURE 1.8 – Prédicteurs du mode Direct temporel des slices B.

Les vecteurs mouvement pour le Direct spatial sont calculés comme le prédicteur d'un bloc  $16 \times 16$  bidirectionnel. Le mode Direct temporel utilise le vecteur temporel (le vecteur "collocated"  $mv_{col_{L1}}$ ) de l'image passée de référence  $L1$  qui traverse l'image B courante. Ce vecteur est ensuite mis à l'échelle en fonction des distances temporelles entre l'image B courante et les images de référence. Les vecteurs mouvement pour le mode Direct temporel sont donnés par les formules suivantes :

$$mv_1^{L0} = \frac{mv_{col_{L1}}}{d_{L0L1}} \times d_{L0} \quad (1.8)$$

$$mv_1^{L1} = \frac{mv_{col_{L1}}}{d_{L0L1}} \times (d_{L0} - d_{L0L1}) \quad (1.9)$$

où  $d_{L0}$  est la distance temporelle entre l'image courante et l'image passée  $L0$  et  $d_{L0L1}$  est la distance temporelle entre l'image future et l'image de référence  $L0$ .

## 1.4 H.264/AVC et après ?

Depuis la fin de la normalisation d'H.264/AVC, le groupe JVT s'est focalisé sur la normalisation de l'annexe scalable d'H.264 (H.264/SVC [H.264/SVC]) et l'annexe multivue (H.264/MVC [H.264/MVC]). Cependant les activités classiques d'amélioration du codage vidéo ont continué durant cette période. Afin d'encourager de nouvelles contributions, il a été décidé au 26<sup>ième</sup> meeting VCEG la création d'un logiciel, nommé KTA (Key Technical Area) [KTA]. Ce logiciel regroupe l'ensemble des outils proposés depuis la fin de la normalisation d'H.264/AVC et qui apportent des gains jugés significatifs par le groupe VCEG. Le jugement de l'efficacité d'une méthode est un compromis entre l'efficacité de compression, la complexité de calcul et de mémoire mais aussi de son impact sur la résistance aux erreurs. Le JM KTA est basé sur le logiciel de référence de la norme : le JM11.0 [JM].

Le premier outil qui a été intégré est le  $\frac{1}{8}$  de pixel [Wed00]. Cet outil est simplement l'extension du  $\frac{1}{4}$  de pixel de la norme pour la compensation de mouvement. Il a cependant une faible efficacité pour

les hautes résolutions, il n'est donc plus considéré comme un outil intégré. L'Adaptive Interpolation Filter (AIF) [VEW<sup>+</sup>05b], [VO06] est aussi un outil d'amélioration de la compensation de mouvement sous-pixellique. L'idée est de ne pas utiliser un filtre fixe [Wer96] pour le calcul des résolutions sous-pixelliques, mais des filtres 2D adaptés pour chaque image [VEW<sup>+</sup>05a]. Cette méthode étant relativement complexe [VO05] (elle multiplie par trois le temps de calcul au décodeur), une adaptation de filtre séparable a été proposée dans [WW07], qui réduit cette complexité. Cette extension est nommée S-AIF. Les choix entre la méthode de référence et l'outil AIF se fait slice par slice avec le critère débit-distorsion [VO07]. Toujours pour l'amélioration de la compensation de mouvement sous-pixellique, l'outil Directional Adaptive Interpolation Filters (D-AIF) [RUL07] utilise une adaptation d'un filtre unidimensionnel en diagonale et non horizontalement et verticalement comme la norme H.264/AVC. Les auteurs insistent sur le fait que la méthode est moins complexe que l'outil AIF. Pour améliorer encore l'efficacité de codage, dans [RUHL08] les mêmes auteurs proposent de calculer leurs filtres d'interpolation directionnelle sur des entiers de 16 bits au lieu des 8 bits définis dans la norme, permettant ainsi une plus grande précision dans le calcul du sous-pixel. Les gains sont comparables à l'utilisation de flottants sur 32 bits.

Les outils intégrés dans le KTA ne concernent pas tous l'amélioration de la compensation de mouvement sous-pixellique. Une nouvelle compétition pour le codage des résiduels de texture issue de la compensation de mouvement a été proposée dans [NM06] et intégrée dans le KTA. Cet outil nommé Adaptive Prediction Error Coding (APEC) permet la sélection entre les transformées DCT de la norme et un codage sans transformée. Dans [TC06] une sélection entre deux matrices de quantification pour les résiduels de texture engendrés par la compensation de mouvement est proposée et intégrée dans le KTA. Dans cette méthode, nommée Adaptive Quantization Matrix Selection (AQMS), la compétition entre les deux matrices se fait au niveau image. La première est la matrice de quantification de la norme et la seconde est une matrice de quantification uniforme. La méthode Internal Bit Depth Increase (IBDI) [CN07] permet d'augmenter la précision des calculs à l'encodeur et au décodeur. Les calculs sont effectués sur 12 bits au lieu de 8 bits afin d'améliorer l'efficacité de codage. Les résultats sont meilleurs avec 8% de gain sur les séquences 720p en profil Baseline IPP. A ce jour, une seule contribution sur le codage Intra a été acceptée [YK07]. Ce nouveau mode Intra Mode Dependent Directional Transform (MDDT) ajoute de nouvelles prédictions correspondant à celle de la méthode BIP expliquée dans la section 1.2. Les résiduels engendrés sont ensuite transformés à l'aide de la transformée KLT. Enfin dans [JL06b], [JL06a], nous avons proposé une méthode de compétition exhaustive pour le codage des vecteurs mouvement (MVComp). Elle a été jugée efficace et peu complexe par le groupe VCEG. Nous décrivons en détail, dans le chapitre 4, cette méthode qui a été intégrée dans le JM KTA.

Tous les outils cités précédemment impliquent une modification du décodeur. Ces outils seraient donc normatifs s'ils étaient intégrés dans une future norme. Cependant le KTA peut intégrer des outils non normatifs comme l'outil Rate Distortion Optimized Quantization (RDOQ) [KYC08a] qui optimise la sélection du QP au niveau macrobloc et son extension à la quantification de la chrominance [KYC08b]. Une implémentation de l'Adaptive Quantization Parameters (AQP) [SW98] a aussi été ajoutée dans le KTA. Cet outil adapte le QP macrobloc par macrobloc à l'aide du critère débit-distorsion. La différence ( $\Delta_{QP}$ ) de pas de quantification est transmise macrobloc par

macrobloc.

Actuellement, aucun test utilisant l'ensemble des outils intégrés n'a été présenté, car certains de ces outils ont des temps d'exécution élevés à cause de leur implémentation en multi-passe et parce qu'ils ne sont pas tous compatibles, d'un point de vue efficacité de codage et de décodage. D'ailleurs, des contributions concernant la réduction de complexité de certains outils ont été proposées [RU07], [TCY08], [Wed01] et la compatibilité des outils deux à deux devient une priorité [WT08]. Nous avons proposé des résultats de différents outils et de leurs combinaisons deux à deux dans [JL07] avec une moyenne de 21% de gain sur les séquences 720p pour le profil Baseline lorsque l'outil AIF et MVComp sont utilisés et 9% de gain sur les séquences QCIF en profil High IBBP. La complexité de chaque outil présent dans le KTA 1.8 a été évaluée [Hor08] en termes de nombre de cycles machine par pixel, pour le codeur et décodeur. Les résultats reportés en fonction de ce nombre et du pourcentage de réduction de débit montrent que certains outils n'augmentent pas la complexité de manière significative au codeur et au décodeur. Parmi ceux-ci, certains sont même moins complexes que la référence.

L'ensemble des outils intégrés dans le JM KTA constitue des améliorations incrémentales de la norme H.264/AVC. Cette approche est induite par la nature du logiciel de référence. En effet, il engendre une efficacité de codage élevée, grâce à une combinaison optimale d'outils normatifs et non normatifs. Il est par conséquent difficile de proposer une rupture du schéma global avec une efficacité de codage équivalente, en un temps de développement raisonnable. Bien entendu, les approches en rupture avec le schéma classique ont été et sont toujours étudiées. L'utilisation d'une transformée en ondelettes ou d'une transformée en ondelettes hybride de type xlets pourrait constituer cette rupture technologique pour un standard de codage vidéo. Ce type de transformée impliquerait, entre autres, une modification de la compensation de mouvement afin qu'il soit adapté à la taille de cette transformée. L'approche d'analyse-synthèse de texture pour le codage d'image ou de vidéo consiste à créer ou extraire des morceaux d'images déjà encodées, représentant une texture significative, pour les utiliser comme prédicteurs après une éventuelle déformation [Wol94]. Ces méthodes ont été comparées à la norme H.264/AVC dans [NNMB<sup>+</sup>03]. Les auteurs précisent que les résultats sont probants à très haut débit, pour une qualité visuelle équivalente. Remarquons que dans cette étude, l'algorithme d'analyse synthèse est semi-automatique. La rupture technologique pourrait être engendrée par le besoin de nouvelles fonctionnalités. Par exemple, l'utilisation de la compression vidéo dans un environnement subissant des pertes de données provoque de fortes baisses de qualité. La norme actuelle exploite largement les redondances temporelles et spatiales entre les données. Par conséquent, la perte d'un seul bit peut engendrer de fortes dégradations sur plusieurs images. L'élaboration d'un standard de codage vidéo robuste aux pertes de donnée engendrerait cette rupture technologique. Des algorithmes basés sur le codage distribué pourraient constituer cette rupture. Un codage par objets de nouvelle génération pourrait aussi engendrer une rupture du schéma classique ou du schéma de codage orienté objets connus, tel que celui utilisé dans la norme MPEG-4 Part 2. Dans [BB00], la notion d'objet est élargie au concept d'évènement visuel. C'est-à-dire que les objets considérés peuvent se chevaucher et chacun d'eux représente un évènement visuel, comme une texture ou du bruit. Cette approche reste pour le moment théorique.

## 1.5 Résumé du chapitre

La norme H.264/AVC est le point de départ de nos contributions. Dans ce chapitre, nous avons détaillé le schéma de codage de cette norme. On retiendra que le codage Intra est équivalent au problème du codage d'images fixes. On peut décomposer ce codage en quatre phases : le partitionnement du macrobloc, les prédicteurs utilisés, l'ordre de parcours de partitions et la transformation des résiduels. Les redondances temporelles sont exploitées par le codage Inter qui utilise une compensation de mouvement par bloc. Le standard offre une multitude de possibilités de codage pour un macrobloc. Dans le chapitre suivant nous verrons le processus de sélection du meilleur codage.

Enfin, lors des réunions du groupe VCEG, des améliorations du schéma d'H.264/AVC ont été proposées, dont certaines ont déjà été jugées efficaces et intégrées dans le JM KTA. Bien qu'aucun appel à contributions pour une future norme n'ait été lancé, les contributeurs de ce groupe travaillent déjà à son élaboration.



## Chapitre 2

# Méthodes de codage par compétition

*L'objectif de ce chapitre est de définir le codage par compétition qui est le fondement de nos contributions, que ce soit pour la partie II, dans laquelle nous utiliserons ce type de méthode pour améliorer l'efficacité de codage mais aussi pour la partie III, où nous tenterons de réduire l'information liée à ces techniques de codage. Nous détaillerons en particulier les critères de sélection utilisés dans ces méthodes. Puis, nous examinerons les possibilités de codage qui ont été mises en compétition dans l'état de l'art.*

### 2.1 Définition

Une méthode de codage par compétition permet de sélectionner un codage parmi un ensemble de codages possibles. Le dictionnaire (Le Petit Robert) définit le mot compétition par une “*recherche simultanée par plusieurs personnes d'un même avantage, d'un même résultat*”. En considérant que les mots “*plusieurs personnes*” peuvent être remplacés par “*possibilités de codage*”, la définition d'une méthode de codage par compétition est la recherche par plusieurs possibilités de codage d'un même résultat : le meilleur codage. En codage vidéo avec pertes, la notion de meilleur codage est généralement équivalente à celle du meilleur compromis entre la distorsion et le débit. Ce compromis peut être étendu à des contraintes de robustesse [ZM05] et de temps d'exécution de décodage.

Une méthode de codage par compétition est définie par les possibilités de codage mises en compétition et par le processus de sélection de la meilleure possibilité de codage. La sélection peut provenir de comparaisons exhaustives ou a priori, comme on peut le voir dans la figure 2.1 représentant le schéma d'une méthode de codage par compétition de  $N$  possibilités. Un a priori peut être un choix humain, applicatif ou un choix provenant de statistiques. Les tests exhaustifs sont des choix dits objectifs : minimisation du critère débit-distorsion par exemple.

Dans le schéma de la figure 2.1, la sélection parmi les  $N$  possibilités de codage est faite soit par des tests exhaustifs engendrant la transmission de l'indice de la possibilité de codage, soit par

un critère a priori ; le critère a priori permet d'éliminer certaines possibilités de codage. Il reste alors  $n$  possibilités. La recherche du meilleur codage parmi ces  $n$  possibilités peut se faire soit par un autre critère a priori soit par des tests exhaustifs. Si les choix des critères a priori utilisés sont reproductibles au décodeur, la transmission de l'indice correspondant au codage utilisé n'est pas nécessaire. Dans le cadre des codeurs vidéo, une méthode de compétition peut être définie pour chaque couche du codage (applicative, vidéo, image, région, objet, bloc, etc.) mais aussi pour chaque étape du codage (prédiction, transformée, quantification, etc.). Notons qu'une méthode de codage utilisant un ensemble de codages successifs n'est pas une méthode de codage par compétition. Dans [MK02], une transformée fractale et une transformée DCT sont utilisées, mais conjointement, c'est-à-dire que les deux transformées ne sont pas mises en compétition mais constituent une "nouvelle" transformée. En effet, la transformée DCT va être utilisée pour coder les basses fréquences et la transformée fractale va permettre de coder les hautes fréquences. De même, dans [KCL92], les images sont transformées à l'aide de la transformée en ondelettes puis les coefficients résultant de cette opération sont encodés par des DCT orientées en fonction de la sous-bande à laquelle ils appartiennent. Ces méthodes utilisant des transformées successives sont différentes du standard H.264/AVC pour lequel les deux tailles de DCT sont mises en compétition, par exemple.

L'efficacité d'une méthode de codage par compétition dépend des codages mis en compétition, du processus de sélection de ces codages, et du coût de l'information lié à cette compétition, tant au niveau du taux de compression engendré que de celui de la complexité de calcul. Dans ces méthodes, l'efficacité en terme de compression n'est pas liée à l'efficacité indépendante de chacune des possibilités de codages mais à leurs efficacités complémentaires.

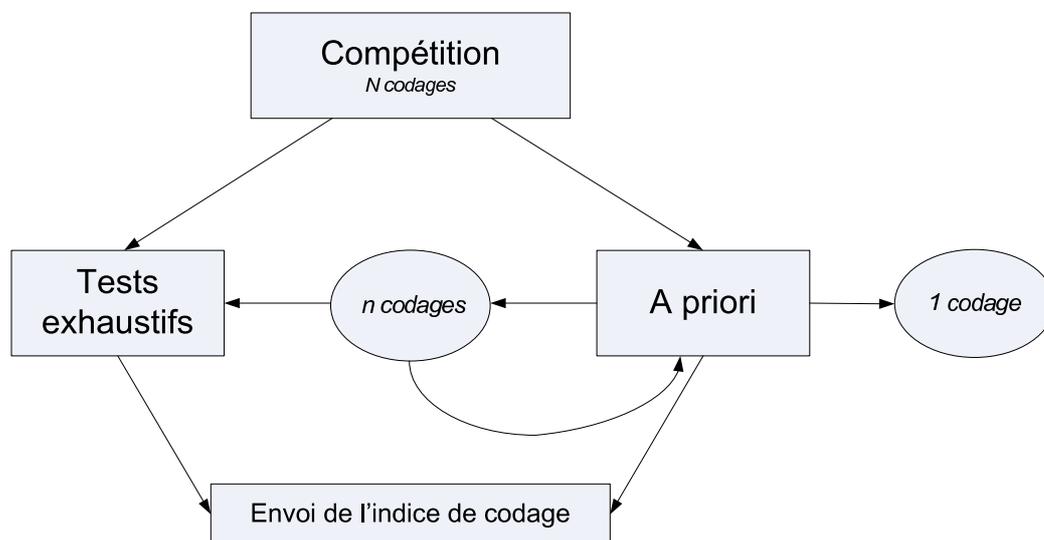


FIGURE 2.1 – Schéma d'une méthode de compétition pour  $N$  codages.

## 2.2 Sélection du meilleur codage

La sélection du meilleur codage parmi l'ensemble des codages possibles est obtenue soit par une sélection basée sur un critère exhaustif, soit par une sélection provenant de critères a priori. Dans cette section, nous définirons les critères exhaustifs utilisés pour comparer les différents codages, notamment le plus utilisé : le critère débit-distorsion. Ensuite, les compétitions basées sur des a priori seront évoquées. Nous définirons particulièrement ce que l'on entend par a priori, en expliquant l'utilité de ce mode de sélection pour limiter les tests exhaustifs impliquant la réduction de complexité de calcul.

### 2.2.1 Critères exhaustifs

Une sélection exhaustive signifie que toutes les possibilités de codage sont évaluées avec la même métrique. La possibilité de codage minimisant (ou maximisant selon la métrique) l'ensemble des évaluations est alors sélectionnée comme meilleur codage. En compression vidéo et d'image avec pertes, la sélection est le processus qui permet de choisir la possibilité de codage qui obtient le meilleur compromis entre la qualité et le débit.

La problématique de débit qualité pour le codage d'un signal avec pertes a été définie en 1948 par Shannon [Sha48]. Le codage avec pertes est basé sur le fait qu'il n'est pas nécessaire d'envoyer l'ensemble de l'information originale pour "comprendre" l'information. La perte introduite dans l'information originale a pour but de réduire le nombre de bits à transmettre ou à stocker. Dans une technique de compression avec pertes, le nombre de bits correspond à la quantité d'information nécessaire pour décrire l'information relative à un certain critère de fidélité du signal de départ. Ce problème peut être résolu par la méthode du "Rate-Constraint" (débit contraint) [HC97], [LJO98], [SG88] qui est défini par l'équation suivante :

$$\min\{D\}, \text{ avec } R \leq R_c \quad (2.1)$$

où  $D$ , la distorsion, est la mesure inverse de la qualité,  $R$  (rate), le débit, et  $R_c$  est le débit maximum que l'on peut utiliser. Cette méthode est donc la recherche de la minimisation de la distorsion  $D$  et d'un débit  $R$  inférieur à la contrainte  $R_c$ . On peut aussi résoudre la problématique débit-distorsion avec l'optimisation Lagrangienne [Eve63]. Cette méthode d'optimisation du débit-distorsion est donnée par la formule suivante :

$$\min\{J\}, \text{ avec } J = D + \lambda R \quad (2.2)$$

où  $\lambda$  est le paramètre de Lagrange, et  $J$  est le critère débit-distorsion que l'on notera RD dans la suite du manuscrit. L'analyse théorique du critère RD a été largement étudiée durant ces 60 dernières années. La difficulté avec ce critère est de déterminer le paramètre de Lagrange  $\lambda$ . En effet, l'optimisation du critère RD est différente selon les caractéristiques du signal d'entrée [OR98]. Les signaux 1D peuvent facilement être associés à un modèle mathématique [Ber71], [Ben48] mais les signaux 2D ou 3D comme les images ou les vidéos ne comportent pas les mêmes caractéristiques. De plus, pour le codage de ces signaux, on utilise des codeurs complexes qui exploitent des redondances spatiales

et temporelles avec des transformées, des prédictions, etc. L'analyse mathématique théorique de ces signaux est très difficile à résoudre [OR98]. Il est nécessaire d'adapter l'optimisation du critère RD en fonction des caractéristiques des codeurs vidéo [Wie02] et d'image. De même, cette sélection doit dépendre de l'utilisation de ces images et vidéos [ZM05]. En effet, certaines applications nécessitent un compromis entre l'efficacité de codage et la robustesse. Enfin, dans une méthode de compression avec pertes, l'étape de quantification, qui altère l'information, va modifier le compromis. Il est alors essentiel que le paramètre de Lagrange dépende des pertes introduites par la quantification [HC97].

Les mesures d'estimation de la distorsion sont aussi un facteur d'efficacité de la sélection avec le critère RD [SW98]. Les mesures les plus utilisées sont : le SAD, la SSD, la MSE, le PSNR etc. Cependant, ces mesures ne reflètent pas forcément la perception visuelle [JJS93]. Dans [CPN04], afin de sélectionner le meilleur codeur vidéo au niveau objet, chaque codeur est estimé à l'aide du critère RD. La mesure de la distorsion est la  $MSE_{Text}$  : la MSE dans le "domaine texture" qui prend en compte la distorsion géométrique engendrée par un codage.

Afin d'adapter la sélection en fonction du signal à coder, dans [WG91] il est proposé que le paramètre  $\lambda$  du critère RD dépende des macroblocs déjà encodés. Cependant, ces algorithmes ne prennent pas en compte les dépendances entre les macroblocs. Par exemple, dans la norme d'H.264/AVC, le codage du macrobloc courant dépend des macroblocs voisins : les vecteurs des modes Inter sont prédits en fonction de la valeur des vecteurs voisins, le signal utilisé pour le calcul des prédicteurs Intra provient des macroblocs voisins, et les contextes adaptatifs du CABAC dépendent du codage de ces derniers. Dans [WLM<sup>+</sup>96], les auteurs proposent de résoudre l'optimisation de la sélection des modes de codage. Ils formulent la problématique du débit-distorsion pour un GOP (c'est-à-dire que le choix de chaque macrobloc dépend de l'ensemble des codages des macroblocs composant le GOP). La formulation de ce problème est théorique et inapplicable en terme de complexité de calcul. Ils reformulent donc le problème en considérant que le codage dépend du codage du macrobloc précédent et du suivant. Cette méthode est implémentée en utilisant la programmation dynamique.

En considérant que le paramètre de Lagrange doit être différent selon l'étape du codage, dans [SW98], deux  $\lambda$  sont définis : un pour l'estimation de mouvement  $\lambda_{Motion}$  et un pour la sélection des modes de codage  $\lambda_{Mode}$ . Ces deux paramètres dépendent du pas de quantification. Cette méthode est implémentée dans le logiciel de référence de la norme H.264/AVC [JM] : le JM. Pour l'estimation de mouvement, le débit  $R$  est égal au débit  $R_{mv}$  [Gir94]. Le critère RD, pour cette étape est donné dans l'équation suivante :

$$J = D + \lambda_{Motion} R_{mv} \quad (2.3)$$

Dans le JM, le débit des vecteurs mouvement,  $R_{mv}$  est estimé pour chaque vecteur lors de l'estimation de mouvement, pour chacun des modes Inter ayant un vecteur. Ce débit correspond au coût du résiduel  $\varepsilon_{mv}$  estimé à l'aide de la table de Golomb. Ce coût est le même quelque soit le type du codage entropique utilisé (CAVLC ou CABAC). Pour la sélection au niveau macrobloc, le critère RD est donné par la formule suivante :

$$J = D + \lambda_{Mode} (R_r + R_m + R_o + R_{mv}) \quad (2.4)$$

où  $R_r$  représente le débit des résiduels de texture,  $R_m$  le débit des modes,  $R_o$  les débits des autres

informations (CBP, Stuffing Bits, reference frames, etc.) et  $R_{mv}$  le débit des vecteurs mouvement. Pour chaque macrobloc, le débit réel au bit près est calculé pour chaque mode de codage Inter et Intra. Dans ce cas,  $R_{mv}$  représente le vrai nombre de bits. Pour chaque mode, le critère RD est calculé avec un  $\lambda_{Mode} = (\lambda_{Motion})^2$  comme paramètre de Lagrange.

Le mode Skip n'ayant pas de résiduel, l'équation du débit-distorsion devient pour ce mode :

$$J_{SKIP} = D_{SKIP} + \lambda_{Mode}R_m \quad (2.5)$$

où  $D_{SKIP}$  est la distorsion engendrée par ce mode et  $R_m$  est le nombre de bits nécessaire au codage de l'indice "Skip" qui est le seul débit engendré par ce mode. Le critère RD pour le mode Direct est équivalent. En pratique, le coût  $\lambda_{Mode}R_m$  est négligeable comparé à la distorsion  $D_{SKIP}$ . Le débit  $R_m$  est souvent inférieur à un bit, que ce soit pour le CABAC ou pour le CAVLC. En effet, les contextes du codage de l'indice pour le mode Skip, représenté par la valeur 0, sont binarisés avec un seul bit. Ce mode a une forte probabilité de sélection. Par conséquent, le codage arithmétique obtient de bonnes performances et permet de coder un macrobloc sauté avec moins d'un bit. Pour le CAVLC, un codage RLC est utilisé pour le codage de ce mode : lorsqu'un premier macrobloc est codé en Skip, on écrit la valeur 0 dans le bitstream puis on compte le nombre de macroblocs successifs codés en Skip (dans l'ordre de parcours de l'image). Si on arrive à la fin de l'image, ou si le nouveau macrobloc courant n'est pas codé en Skip, on écrit ce nombre.

### 2.2.2 Critères "a priori"

Les critères a priori sont des choix humains, applicatifs ou statistiques. Le but de ce type de sélection est en général de réduire la complexité de calcul. Pour cela, les estimations de mouvement rapides éliminent certaines zones de la fenêtre de recherche. Ces estimations peuvent être des approches multi-résolutions [Bie88], [DK92], hiérarchiques [ZZB91], ou basées sur la descente de gradient [LKF96]. Ces méthodes utilisent les propriétés des vidéos pour réduire le nombre de calculs du critère utilisé pour l'algorithme BMA (critère RD dans le cas du codage vidéo). Comme ces estimations ne testent pas l'ensemble des possibilités de codage, la sélection de la meilleure efficacité de codage est (théoriquement) sous-optimale par rapport à une recherche exhaustive. Les recherches rapides du meilleur prédicteur Intra sont basées sur des critères statistiques ou des caractéristiques du signal d'entrée, comme dans [WLN07], où une détection de contour, basée sur la transformée de Haar, est utilisée pour sélectionner ou non un ensemble réduit de prédicteurs. Dans [YPL04], les macroblocs sont classés en fonction de leur contenu. Si ce contenu est lisse, un ensemble restreint de prédicteurs est testé. Dans ces deux cas, l'ensemble des possibilités de codage n'a pas été évalué avec le critère RD mais avec un critère "a priori". Dans [REK97], un codage dynamique est utilisé. Plusieurs codeurs sont alors mis en compétition. Le nombre de ces compétitions étant trop large et inutile en fonction des applications visées, un ensemble de solutions admissibles est déterminé en fonction de l'application visée. Les critères d'élimination sont la complexité et l'efficacité de codage de chacun des codeurs. Le meilleur codeur parmi l'ensemble des codeurs retenus est ensuite sélectionné avec le critère RD. Le critère a priori proposé provient donc d'un choix humain ou d'un choix applicatif.

## 2.3 Possibilités de codage par compétition

Dans cette section, nous passons rapidement en revue différentes possibilités de codage mises en compétition pour le codage vidéo, afin de déterminer nos axes de recherche pour l'introduction de nouveaux modules de codage par compétition.

### 2.3.1 Compétition de codeurs

Dans ce paragraphe, nous présentons la compétition de modes de codages/codeurs. La majorité des références étudie des codages par objets. Les compétitions se situent donc à "haut niveau", dans le sens où ces codages sont parfois des codeurs vidéo ou d'image indépendants.

L'utilisation de plusieurs types de codage est souvent appelée codage dynamique [RELB<sup>+</sup>97]. Son principe réside sur l'observation qu'une seule technique de codage ne peut être efficace pour toutes les applications [Fle99]. Le codage dynamique est une combinaison de multiples techniques de compression activées sur différentes zones de l'image [RCLB<sup>+</sup>96]. Dans ces méthodes, une première phase de segmentation est généralement appliquée. Ce type de méthode est inclus dans les méthodes de codage par objets. Dans [CPN04], un codage par objets utilisant quatre codeurs vidéo mis en compétition est proposé. Les quatre codeurs sont : le "3D model-based coder" (représentation d'un GOP à l'aide d'un modèle 3D), le "Sprite coder" (utilisation d'un mouvement paramétrique et de JPEG 2000 [PMCM01]), "2D+t wavelet coder" (un codeur ondelettes spatio-temporel) et "H.264/AVC object based coder" (H.264/AVC modifié pour être compatible avec le codage par objets). Pour chaque objet, la texture et l'information de mouvement sont codées à l'aide d'un de ces codeurs. Dans [REK97], une compétition entre différents codages est aussi proposée. Chaque image est découpée en régions (union de blocs de tailles variables). Les modèles (codages) utilisés sont : le modèle de compensation de mouvement (précision du mouvement au  $\frac{1}{2}$  pixel), le modèle du fond (pour les parties sans mouvement), modèle de textes et de graphiques (la région est représentée en binaire) et le modèle fractal (compression avec une transformée fractale). Les auteurs précisent que le modèle de fond est un cas particulier du modèle de compensation de mouvement. Notons que la segmentation en régions est faite à l'aide d'une structure par arbre.

Dans [BB00], six codeurs considérés comme indépendants sont mis en compétition. Chaque codeur est chargé de coder un "événement visuel". Cet article est en relation avec le codage dynamique. Les six codeurs sont : "Gain Expert" (compense les différences au modèle), "Remplacement Expert" (retire ou ajoute les objets contenus ou non dans le modèle), "Biais Expert" (code les pixels différents entre le modèle et l'image), "Motion Expert" (code les déplacements spatiaux des objets), "Lighting Expert" (code les variations globales dans la luminosité) et "Additive expert" (code le bruit de l'image). Cet article est différent de ceux de l'état de l'art car les "Experts" (les codeurs) ne sont pas chargés de coder uniquement des objets. L'idée est que les codeurs proposés sont chargés de coder un événement visuel et non une zone de l'image. Les codeurs présentés sont complémentaires au sens où ils peuvent être une suite de codage au sens "hybride", ou complémentaires au sens du codage par objets. Par conséquent, il est difficile de déterminer le niveau où a lieu la compétition entre les codeurs (est relativement flou). Cette description reste théorique et est présentée comme un appel à contributions.

### 2.3.2 Compétition de prédictions

Dans la norme H.264/AVC [H.264/AVC], le codage de la texture se fait par prédiction, que ce soit pour exploiter les redondances spatiales ou temporelles. La recherche du meilleur résiduel de texture est, dans le schéma des standards vidéo actuels, la partie du codage qui engendre le plus grand nombre de tests exhaustifs.

Pour exploiter les corrélations spatiales, le codage prédictif Intra [FPR95] utilise plusieurs partitionnements du macrobloc :  $16 \times 16$ ,  $8 \times 8$  et  $4 \times 4$ . Chacun de ces partitionnements a plusieurs prédicteurs et toutes ces prédictions sont mises en compétition. L'amélioration du codage prédictif Intra est largement étudiée dans l'état de l'art, comme nous avons pu le voir dans le chapitre 1 de cette même partie. Toutes ces nouvelles méthodes sont systématiquement mises en compétition avec les modes existants.

Pour exploiter les redondances temporelles, la norme utilise la compensation de mouvement avec plusieurs tailles de blocs [YC88] ce qui offre une meilleure segmentation du mouvement lorsque cela est nécessaire [YCC90]. Pour chaque taille de bloc, une compétition de prédiction de bloc est donc effectuée. Les meilleures prédictions pour chacun des partitionnements du macrobloc sont aussi mises en compétition. Les prédictions peuvent provenir du passé et du futur pour les slices B, et de plusieurs images de référence [WZG99]. Dans le cas de l'exploitation des redondances temporelles, le partitionnement d'une image peut descendre au niveau pixel [CYC90].

### 2.3.3 Compétition de transformées

L'étape de transformation est cruciale en compression d'image et de vidéo. Elle permet de décorréler le signal. Plusieurs transformées ont été utilisées dans le domaine de la compression et les deux plus connues sont la DCT et la transformée en ondelettes [ABMD92]. Ces transformées ont des caractéristiques différentes et une efficacité variable selon le contenu. Il est alors intéressant de les mettre en compétition pour exploiter une efficacité de codage accrue. Dans [WK97], une compétition exhaustive entre la transformée DCT et la transformée en ondelettes au niveau bloc est implémentée. Les gains sont cependant faibles.

La norme H.264/AVC utilise deux tailles de transformée DCT,  $4 \times 4$  et  $8 \times 8$ . L'utilisation de taille multiple de la transformée DCT est assez fréquente dans l'état de l'art [MTI00]. Les premiers articles sur le sujet sont apparus dans les années 90 [Che89]. Dans [TG96], une compétition entre une DCT 1D et une DCT 2D est proposée pour coder les résiduels de texture engendrés par la compensation de mouvement. Dans [Wie03], 4 transformations sont proposées :  $4 \times 4$ ,  $4 \times 8$ ,  $8 \times 4$ ,  $8 \times 8$ . L'information concernant la taille de la transformée utilisée n'est pas envoyée au décodeur car elle est adaptée aux partitionnements du macrobloc. Dans [Val02] des tailles de DCT  $4 \times 4$ ,  $8 \times 8$ ,  $16 \times 16$  sont proposées avec une quantification adaptée à chacune d'elles, afin d'obtenir une distorsion équivalente. Notons que l'auteur utilise des tailles de blocs (carrés) allant du  $2 \times 2$  au  $16 \times 16$ . Comme nous l'avons vu dans le chapitre précédent, l'outil APEC [NM06] met en compétition les transformées de la norme avec un mode de codage sans transformée. De plus, le nouveau mode de codage Intra intégré dans le KTA propose aussi une transformée KLT en plus des DCT classiques.

Les méthodes basées sur le codage dynamique offrent, par leurs compétitions de codage, des

compétitions de transformées. Dans [REK97], un codage utilisant la transformée DCT et un codage utilisant une transformée fractale [Jac92], sont utilisés, cette dernière étant visiblement plus efficace pour le codage des contours [EWK95]. Dans [CPN04], plusieurs codeurs sont mis en compétition. Ils utilisent soit une transformée DCT soit une transformée en ondelettes.

### 2.3.4 Autres compétitions

Dans un codeur vidéo, la compétition peut se situer à d'autres étapes ou d'autres niveaux que ceux dont on a déjà parlé. Une analyse du mouvement entre deux images pour adapter le type des images (I, P, B) dans un GOP a été réalisée dans [LH97]. La mesure MSPE (Mean Square Prediction Error) est calculée entre deux images successives pour déterminer les changements. En fonction de la valeur de cette mesure, l'image est codée en I, P ou B. Le même but est recherché dans [YNYS99] où une adaptation de la structure du GOP est proposée. La taille du GOP et l'intervalle entre les images P sont déterminés en fonction des caractéristiques des images originales. Ils utilisent notamment l'activité et la variance induite par une simple compensation de mouvement pour chaque macrobloc, mais aussi la moyenne sur la luminance totale de chaque image.

Une compétition de prédicteurs de vecteurs mouvement est testée dans [DKBR99] entre trois prédicteurs spatiaux. On notera que la compétition de QP est autorisée dans la norme et a été intégrée dans le KTA avec l'outil RDOQ [KYC08a]. De même, la matrice de quantification des résiduels de texture peut être l'objet d'une compétition comme dans l'outil AQMS [TC06], lui aussi intégré dans le KTA. Enfin, une compétition d'orientations de blocs, pour améliorer l'efficacité de codage des résiduels du codage Intra et Inter, a été présentée dans [Rob08].

## 2.4 Ajout d'une nouvelle possibilité de codage

Le codage par compétition est largement représenté dans la norme H.264/AVC et se révèle très efficace. L'ajout d'une nouvelle possibilité de codage a un coût en terme de débit. En effet, pour que le décodeur puisse distinguer la possibilité de codage choisie par le codeur, il est nécessaire d'indiquer cette possibilité dans le bitstream. Le débit de l'information de compétition ne doit pas dépasser celui du gain moyen en terme de débit-distorsion apporté par le nouveau module de codage. Notons  $J_{old}$  le coût RD moyen de la référence,  $J_{new}$  le coût RD moyen engendré par l'ajout du nouveau module de codage à la référence et  $R_{comp}$  le débit lié à la compétition entre le nouveau module et la référence. Pour que l'ajout de ce module soit efficace il faut que :

$$J_{old} > J_{new} + R_{comp} \quad (2.6)$$

Si l'on considère que l'information de cette nouvelle compétition est un indice, le débit  $R_{comp}$  dépendra du niveau où se fait la compétition et de la probabilité de sélection du nouveau module de codage lorsqu'un codage arithmétique est utilisé pour le codage de l'indice. Par exemple, si la compétition se fait au niveau image, le débit  $R_{comp}$  a un coût négligeable dans le débit total. A l'opposé, si cette compétition se fait au niveau bloc, ce débit devient considérable et peut engendrer

des pertes d'efficacité. Notons, enfin, que le coût RD  $J_{new}$  peut aussi contenir le débit d'informations additionnelles liées à la nouvelle possibilité de codage.

## 2.5 Conclusion

La compétition entre plusieurs possibilités de codage a été étudiée. Le principe du codage dynamique réside sur le fait qu'une seule technique de codage ne peut pas être efficace pour toutes les zones de l'image. Il est donc essentiel d'étudier plusieurs possibilités de codage pour sélectionner celle qui obtient la meilleure efficacité. La compétition se fait à plusieurs niveaux : au niveau séquence pour les choix applicatifs, au niveau image pour sélectionner le type d'image, au niveau objet pour le codage dynamique et au niveau bloc pour le codage de même nom. La sélection parmi l'ensemble des possibilités de codage se fait avec le critère débit-distorsion. Ce critère pondère le débit en fonction de la distorsion avec le paramètre de Lagrange. Ce paramètre doit dépendre de la quantification, du type de codage mis en compétition et des applications visées. La sélection des possibilités de codage peut se faire par des choix sous-optimaux basés sur des a priori qui sont généralement utilisés pour réduire la complexité de calcul.

L'efficacité de la norme est sans préalable, provenant d'une multitude de compétitions de prédictions. Peut-on encore ajouter des prédictions à celles qui existent déjà dans le standard? Nous tenterons de répondre à cette question dans la partie II, de ce manuscrit en proposant des modules de codage basés sur ce principe. Les méthodes de codage par compétition ont cependant un coût élevé lié à la transmission de l'indice représentant la meilleure possibilité de codage. Cette information est non négligeable comme nous allons le voir dans le chapitre suivant. Dans la partie III, nous proposerons des méthodes permettant de réduire cette information en utilisant notamment les critères de choix de codage.



## Chapitre 3

# Etude comportementale du codec de référence de la norme H.264/AVC

*Dans ce chapitre, nous présentons une étude du codeur de référence de la norme H.264/AVC (JM) [JM], que nous avons effectuée avant de fixer nos axes de recherche. Nous étudions la proportion de chacune des composantes du débit (résiduels de texture, vecteurs mouvement, modes, etc.). Nous détaillons aussi la répartition spatiale des coûts en débit, en distorsion et en débit-distorsion, pour comprendre quelles parties des images nécessitent une amélioration, afin de proposer de nouvelles possibilités de codages efficaces. De même, nous examinons la répartition spatiale des modes de codage. Enfin, nous proposons d'analyser le débit de chacun des types d'images dans une configuration réaliste d'utilisation.*

### 3.1 Analyse des slices I, P et B

Comme nous l'avons vu dans le chapitre 1, les slices I utilisent des modes de codage Intra qui exploitent uniquement les redondances spatiales du slice courant. Les slices P utilisent des modes Intra et Inter et exploitent par conséquent les redondances spatiales et temporelles des images passées. Finalement, les slices B utilisent en plus des modes Intra et des modes Inter des slices P, des modes Inter exploitant à la fois les redondances dans le passé, le futur ou le passé et le futur dans le cas des prédictions bidirectionnelles. Ces trois types d'images ne sont pas égaux devant le nombre de possibilités de codage et le nombre de redondances exploitées. Cette inégalité se traduit par une différence en terme d'efficacité de codage et par conséquent de représentation dans le flux.

### 3.1.1 Efficacité de codage

La figure 3.1 représente les courbes débit-distorsion de la deuxième image de la séquence *Foreman* CIF lorsque celle-ci est encodée en I, P ou B. L'efficacité de codage du slice Intra est nettement inférieure à celui d'un slice P qui a lui-même une efficacité de codage inférieure aux slice B. Dans la configuration étudiée, provenant des recommandations VCEG [TSW05], les slices P sont codés avec un QP inférieur à 1 de celui des slices I et que les slices B sont codés avec un QP inférieur à 1 du QP des slices P ( $QP_B = QP_P - 1 = QP_I - 2$ ).

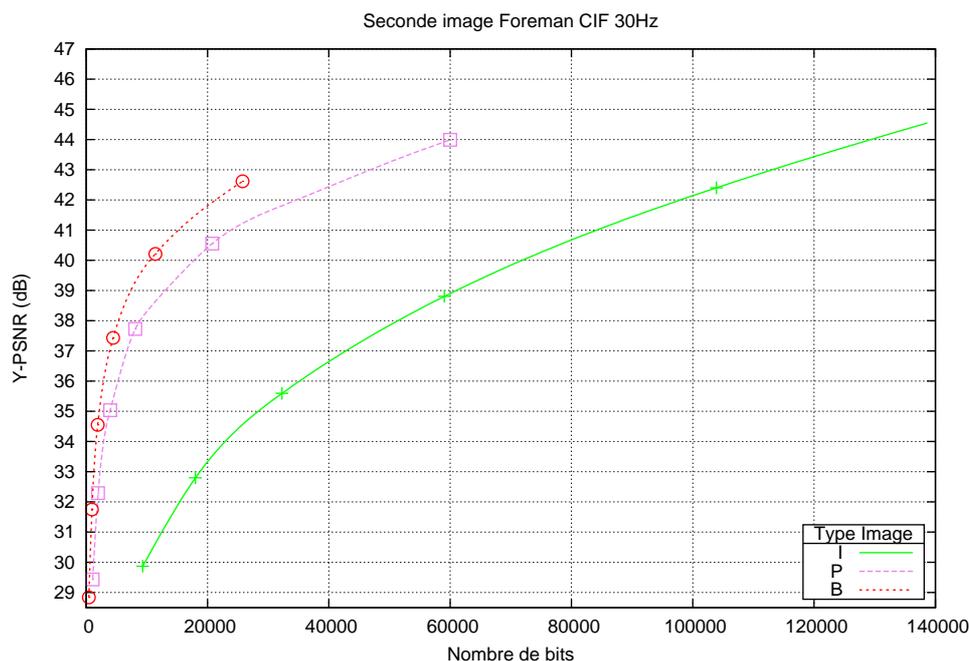


FIGURE 3.1 – Courbes débit-distorsion de la deuxième image de la séquence *Foreman* CIF 30Hz codée en I, P et B. Ces trois types d'images proviennent respectivement de la configuration High III, High IPP et High IBBP.

Le tableau 3.1 donne le pourcentage de réduction de débit calculé avec la métrique de Bjøntegaard [JP07] sur l'ensemble des séquences VCEG (cf. annexe A), pour trois ensembles de QP, d'une image P par rapport à une image I, d'une image B par rapport à une image I et d'une image B par rapport à une image P. Pour l'ensemble de QP 22-37 (recommandé par le groupe VCEG), les slices P sont en moyenne 68% plus efficaces que les slices I, les slices B sont 92% plus efficaces que les slices I et les slices B sont plus efficaces en moyenne que les slices P de 74%. Ces résultats sont obtenus avec une configuration High IBBP GOP 12 qui augmente la distance temporelle entre deux images P, engendrant une baisse d'efficacité. Dans une configuration IPP, l'efficacité de codage des slices P par rapport au slices I est plus élevée que dans cette configuration. On notera que l'écart d'efficacité est plus faible à haut débit qu'à bas débit. D'ailleurs dans [SV04] deux configurations de H.264/AVC

	P/I			B/I			B/P		
QP	17-32	22-37	27-42	17-32	22-37	27-42	17-32	22-37	27-42
Pourcentage de réduction de débit	57	68	74	84	92	94	65	74	76

TABLE 3.1 – Pourcentage de réduction de débit moyen d’une image P par rapport à une image I, d’une image B par rapport à une image I et d’une image B par rapport à une image P. Les débits et PSNR moyen ont été calculés sur l’ensemble des séquences VCEG pour le profil High IBBP GOP 12.

sont comparées : High III et High IBBP. Les résultats obtenus montrent qu’à très haut débit l’écart entre les deux configurations est fortement réduit.

La différence d’efficacité de codage des différents types de slice provient d’une part des redondances temporelles exploitées, mais aussi du nombre de possibilités de prédiction par pixel. En effet, avec l’algorithme du JM, un pixel peut être prédit par trois modes Intra : Intra  $16 \times 16$ , Intra  $8 \times 8$  et Intra  $4 \times 4$ . Le mode Intra  $16 \times 16$  a 4 prédicteurs, l’Intra  $8 \times 8$  a 4 blocs qui ont chacun 9 prédicteurs, soit 9 possibilités pour un pixel. De même pour l’Intra  $4 \times 4$ , soit 22 prédicteurs pour un pixel codé en Intra. Pour les images P, lors de l’estimation mouvement avec une fenêtre de recherche  $32 \times 32$ , un pixel sera comparé pour chacun des modes à 1024 autres pixels dans le cas de la recherche au pixel entier. L’augmentation du nombre de prédicteurs pour l’Intra est un axe de recherche intéressant qui a déjà largement été étudié, comme on a pu le voir dans le chapitre 1.

### 3.1.2 Proportion des images I P B dans un flux

L’efficacité de codage engendrée par chacun des types de slices est donc différente. Dans un flux le nombre de slices encodés en I, P ou B n’est pas le même. Dans les recommandations VCEG les configurations contiennent une seule image I pour l’ensemble de la séquence, excepté dans le cas des configurations III qui ne comportent que des slices Intra (cf. annexe A). Ces configurations ne sont pas réalistes pour toutes les applications. Celles qui ne comportent qu’une seule image I pour une séquence de 300 images sont dédiées aux applications de stockage. En revanche pour les applications nécessitant la transmission du flux sur un réseau, telles que les applications de télévision, de streaming etc. le nombre d’images Intra est plus élevé. Dans le cas de la transmission de flux pour la télévision (24 images/seconde), il est nécessaire de transmettre une image Intra toutes les demi-secondes pour que le temps d’attente durant un changement de chaîne (de flux) n’excède pas cette demi-seconde. De plus cette configuration offre davantage de robustesse aux pertes de paquets. Nous avons donc défini une configuration avec des GOP composés de 12 images, une image Intra pour 11 images Inter. Pour le profil Baseline IPP ces GOP sont donc composés d’une image Intra et 11 images P et pour le profil High IBBP d’une image Intra, de 3 images P et de 8 images B (cf.annexe A).

La figure 3.2 représente l’histogramme cumulatif du nombre de bits nécessaires au codage de chaque type de slice I, P, B en fonction des pas de quantification pour la configuration High IBBP

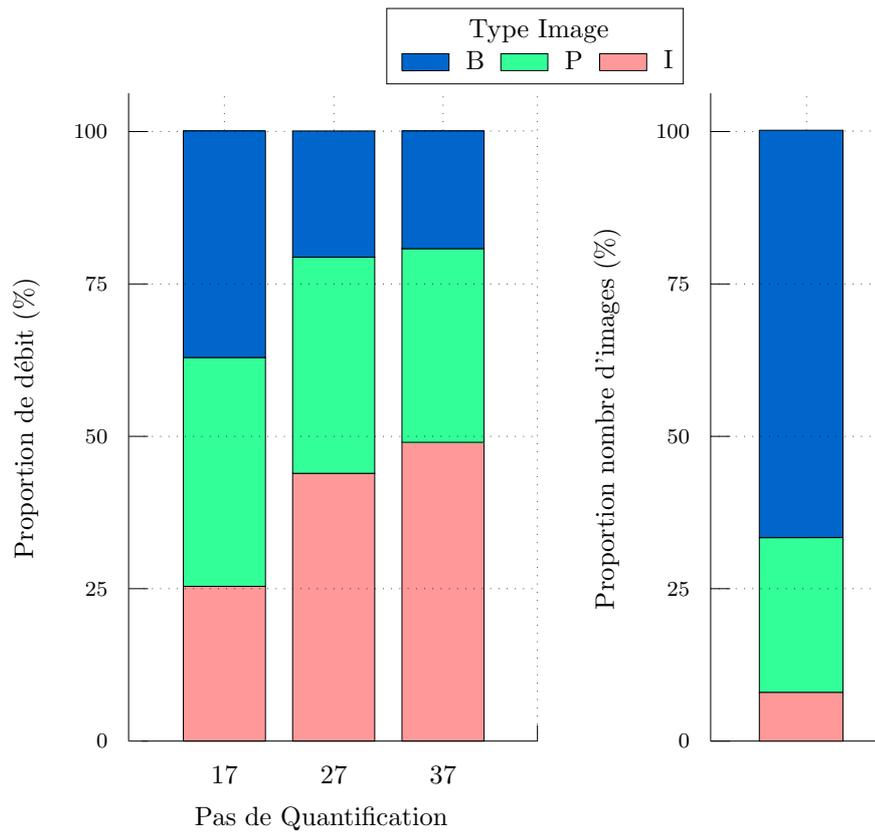


FIGURE 3.2 – Histogramme cumulatif de la proportion de débit de chaque type d'image I, P et B pour la séquence *Foreman* CIF 30Hz codée à l'aide de la configuration High IBBP GOP 12, et de la proportion du nombre d'images.

GOP 12. On voit nettement que pour l'ensemble des QP la proportion de débit est très différente de celle de la proportion en nombre de slices. On remarque aussi que cette proportion évolue en fonction des QP. La proportion des slices B diminue plus fortement que celle des images P. A bas débit, la proportion dans le flux des slices I est très élevée, avoisinant les 50% pour le QP 37, alors que les slices I représentent seulement  $\frac{1}{12}$  de la surface codée. Il sera donc nécessaire pour une éventuelle future norme de codage vidéo d'améliorer le codage des slices Intra.

### 3.1.3 Représentation visuelle des coûts

La figure 3.3 représente visuellement les coûts de débit ( $R$ ), distorsion ( $D$ ) et de débit-distorsion ( $J$ ) pour les images I, P et B de la séquence *Foreman* CIF. La répartition des coûts de distorsion pour les trois types d'images est équivalente, même si l'image B a une distorsion moyenne plus élevée que l'image I et l'image P. Ceci est lié aux QP utilisés pour le codage et donc aux  $\lambda_{Mode}$  et  $\lambda_{Motion}$  des équations (2.4) et (2.3) du chapitre 2. On remarque que la distorsion est plus forte dans les lignes du mur et les contours du visage. Il faut rappeler que la distorsion de la première image I se propage au reste de la séquence. En effet, dans la séquence originale, si une partie à un instant  $t$  est totalement identique à l'instant  $t - 1$ , c'est-à-dire qu'elle ne contient aucun mouvement et qu'il n'existe aucun défaut de capteur, alors la distorsion engendrée dans la première image se propagera dans les suivantes. Dans ce cas précis, on ne pourra compter sur une meilleure reconstruction du résiduel dans les autres images, car chaque coefficient de l'erreur transformée sera toujours inférieur aux coefficients de quantification utilisés.

Dans la représentation visuelle du débit, on retrouve les conclusions des sections précédentes, à savoir que les slices I ont un débit plus élevé que les slices P et que les slices P ont un débit plus élevé que les slices B. On note que la répartition des coûts du débit est proche de celle de la distorsion. C'est-à-dire que les parties de l'image qui coûtent cher en terme de débit, coûtent cher en terme de distorsion. Pour l'image B, la quasi-totalité des blocs de l'image ont un très faible débit. Ces parties correspondent au codage en mode Direct (cf. figure 3.8). Les parties qui ont un débit plus important sont des parties qui subissent des déformations : les yeux qui se ferment, la bouche en mouvement, le pivotement de la tête vers la droite engendrant une déformation du contour du visage, l'ombre dans le cou qui change d'une image à l'autre.

Pour les trois images, l'intensité du coût  $J$  est localisée dans les mêmes zones : des zones contenant des contours. En terme de débit-distorsion, pour les trois types de slices, le codeur de la norme H.264/AVC a une efficacité de codage plus importante pour les zones planes que pour les zones contenant des contours. Cependant, il faut noter que ce ne sont pas forcément les modes Inter qui sont moins efficaces pour le codage des contours car en théorie, s'il n'y a pas de déformation, ce codage devrait produire un résiduel de bloc nul. Pour le codage Intra, il faut améliorer le codage des contours pour engendrer une moindre distorsion afin de limiter la propagation.

Pour l'ajout de nouveaux modules de codage, mis en compétition avec les modules existants, il faut proposer des méthodes de codage plus efficaces pour les contours que pour les zones planes, notamment pour les slices Intra, afin de limiter la propagation de la distorsion dans ces zones. Le codage des contours coûte cher en débit et en distorsion.

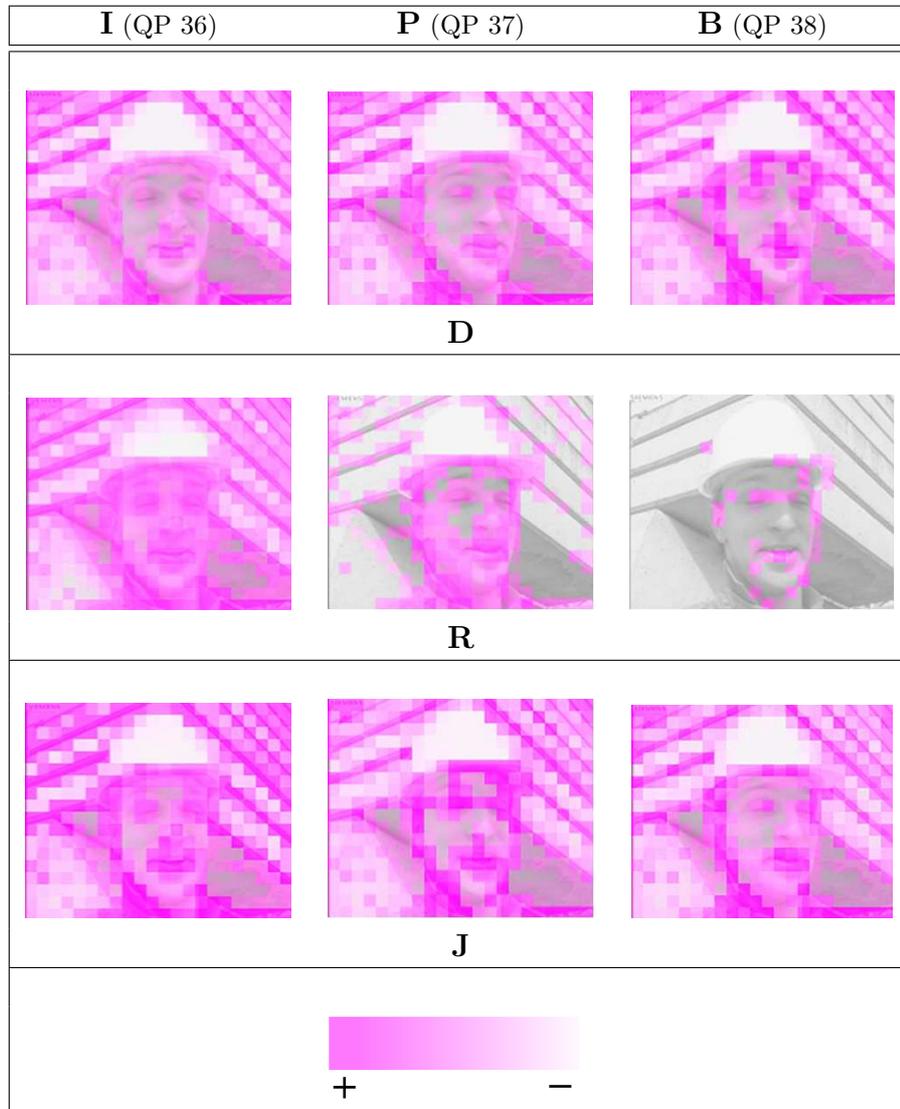


FIGURE 3.3 – Représentation visuelle des coûts J, D et R pour 3 images successives de la séquence *Foreman* CIF 30Hz codées respectivement en I, B, P avec les QP respectivement de 36, 37, 38.

## 3.2 Répartition des informations de codage dans le flux

La figure 3.4 représente l'évolution de la proportion du débit des composantes du flux, en fonction du QP, pour les slices I, en moyenne pour l'ensemble des séquences VCEG. Les débits représentés sont les débits des résiduels de texture, des modes de codage, des prédicteurs Intra  $4 \times 4$  et  $8 \times 8$  et des autres informations. Les autres informations sont les bits de CBP, les Stuffing bits et les  $\Delta_{QP}$ . Le débit des modes représente le débit des modes Intra  $16 \times 16$ ,  $8 \times 8$ ,  $4 \times 4$  et le débit des prédicteurs Intra  $16 \times 16$  qui sont binarisés chacun comme un mode de codage. On voit nettement que le débit des résiduels de texture représente la plus grande proportion dans le flux. L'amélioration du codage de la texture pour le codage Intra est une voie de recherche nécessaire pour l'amélioration du codage des slices I. Néanmoins cette proportion diminue à bas débit, en atteignant 66% du débit total pour le QP 42. Pour ce même QP, le débit des modes Intra atteint presque 20% du débit total, ce qui devient non négligeable. De plus, la figure représente une configuration du profil High, or pour une configuration Baseline la proportion du débit des prédicteurs Intra  $4 \times 4$  atteint 26% du débit total pour le même ensemble de séquences. En effet, le profil Baseline n'autorise pas l'Intra  $8 \times 8$ , qui a un débit de codage des prédicteurs inférieur à celui de l'Intra  $4 \times 4$ , car son nombre de prédicteurs par macrobloc est moins élevé.

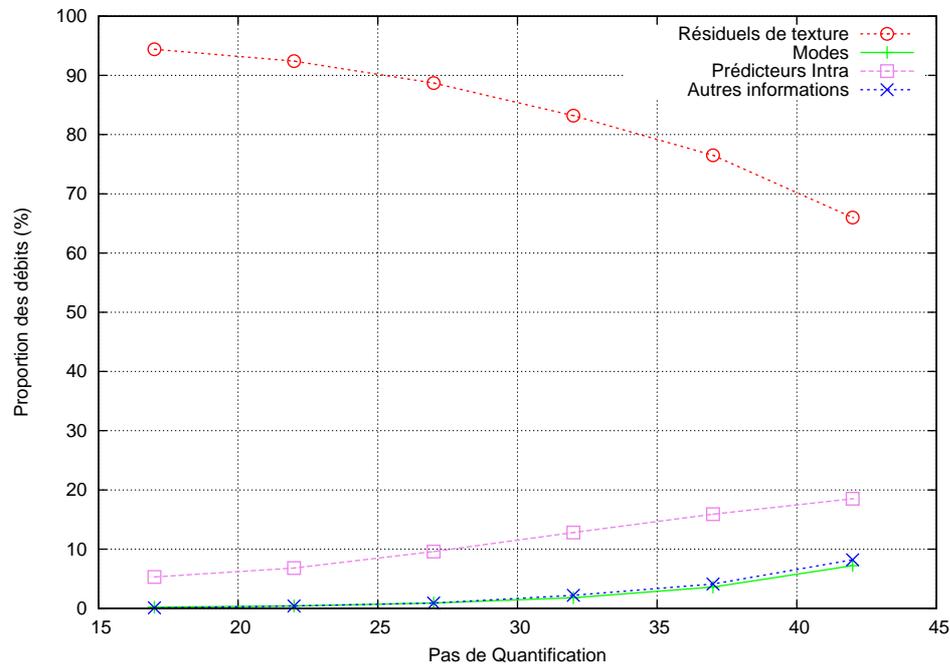


FIGURE 3.4 – Evolution de la proportion des débits des composantes du flux pour les slices I en fonction des QP de la configuration High III calculée en moyenne sur l'ensemble des séquences VCEG.

La figure 3.5 représente l'évolution de la proportion du débit de chaque composante du flux, en fonction du QP, pour les slices P, en moyenne pour l'ensemble des séquences VCEG. La proportion de l'information de mouvement dépasse celle de la texture à bas débit atteignant presque 40% du débit total. La figure 3.6 nous montre la même information que la figure 3.5 pour les slices B. Pour la moitié des QP utilisés, l'information de mouvement est plus élevée que l'information de texture. La réduction du débit de l'information de mouvement est une voie de recherche intéressante et motivante par les gains élevés que ces recherches pourraient engendrer. Une information importante donnée par ces figures est la proportion des débits liés au codage des informations de compétition. En effet, la somme de la proportion des modes de codage et celle des vecteurs mouvement est plus élevée que la proportion du codage de la texture pour les trois derniers QP pour les slices P et des quatre derniers pour les images B. Il paraît alors nécessaire de travailler le codage de ces informations pour les bas débits.

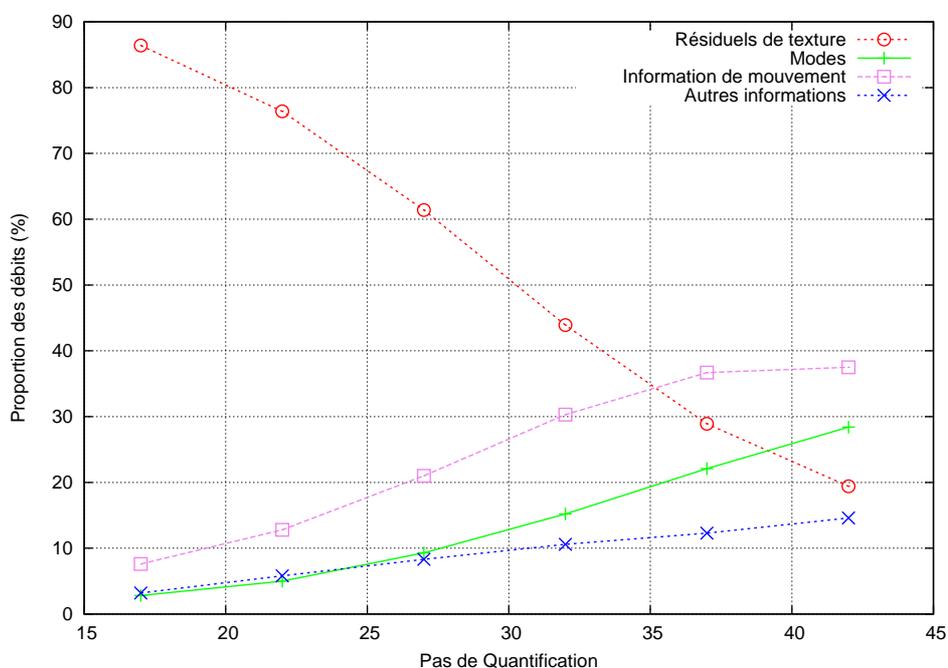


FIGURE 3.5 – Evolution de la proportion des débits des composantes du flux pour les slices P en fonction des QP de la configuration Baseline IPP calculée en moyenne sur l'ensemble des séquences VCEG.

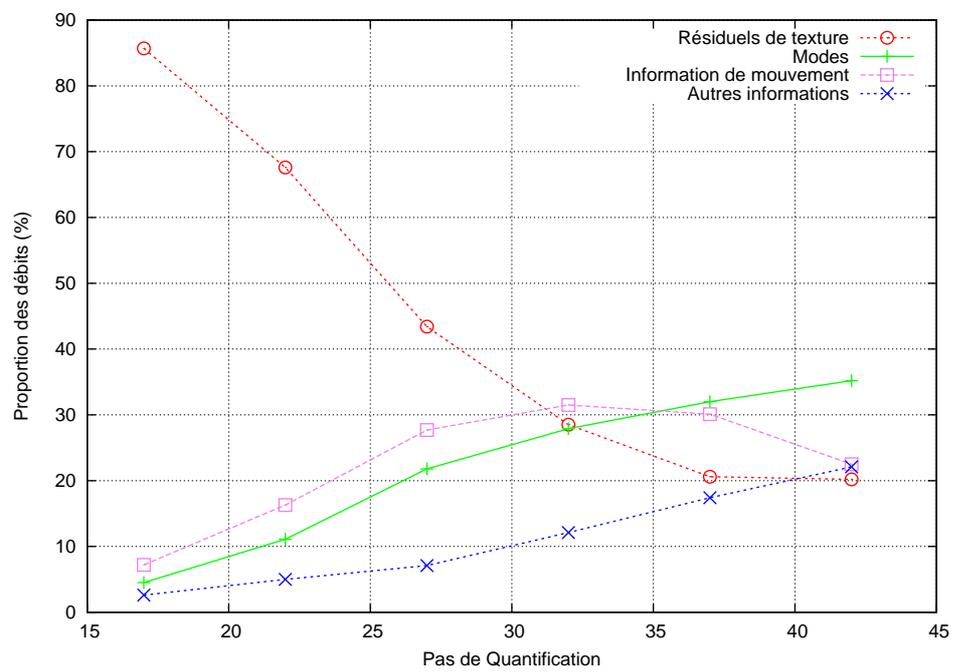


FIGURE 3.6 – Evolution de la proportion des débits des composantes du flux pour les slices B en fonction des QP de la configuration High IBBP calculée en moyenne sur l'ensemble des séquences VCEG.

### 3.3 Sélection des modes de codage

#### 3.3.1 Slices Inter

Le tableau 3.2 donne la proportion de sélection des modes de codage pour les slices P et B pour le profil High IBBP qui est calculée en moyenne sur l'ensemble des séquences VCEG. On remarque tout d'abord que le nombre d'Intra est plus élevé pour les slices P que pour les slices B. Ceci est expliqué par la distance temporelle entre deux images P qui engendrent des prédictions temporelles moins efficaces, et qui se traduit par une augmentation du nombre d'Intra. En effet, le codage Intra, par définition, ne perd pas en efficacité avec l'éloignement des images. Lorsque l'on utilise une configuration IPP, le pourcentage de sélections des modes Intra est plus faible (4,1% en moyenne pour le Baseline IPP). Tous les modes Intra et Inter ont une sélection qui décroît avec l'augmentation des valeurs des QP, seul le mode Skip ou le mode Direct voient leur sélection augmenter. La sélection atteint 87% pour le mode Direct spatial pour les images B. Dans une configuration IPP le nombre de sélections du mode Skip est plus élevé car la distance temporelle entre les images P est plus faible. Les modes Inter de petite taille (8×8, 8×4, 4×8 et 4×4) sont sélectionnés à très haut débit, cette sélection est infime à bas débit. Ces écarts de sélection sont liés au compromis débit-distorsion. En effet, à haut débit, le  $\lambda_{Mode}$  de l'équation (2.4) du chapitre 2 donne un poids plus fort à la distorsion dans le critère RD ( $J$ ), par conséquent les modes de petites partitions sont plus souvent sélectionnés puisqu'ils permettent de meilleures prédictions liées à la décomposition plus fine du mouvement. Pour les QP élevés, le débit a le poids le plus fort dans cette équation (2.4), donc les modes à très faible coût de codage tel que le Skip ou les grandes partitions Inter, qui utilisent moins de vecteurs, sont beaucoup plus souvent sélectionnés. Notons enfin que la sélection des différentes tailles de partitions et du type de prédiction dépend, bien entendu, du contenu des séquences, mais aussi de la résolution de ces dernières.

On remarquera finalement que le nombre de vecteurs codés est moins élevé à bas débit. Par conséquent, le nombre moyen de bits dédiés à cette information est plus faible, mais la proportion de cette information dans le débit total augmente comme on l'a précisé dans la section 3.2. Ceci est expliqué par la quantification des résiduels de texture qui diminue fortement la proportion de cette information dans le débit total et, par conséquent, augmente la proportion des autres informations. Cependant, un autre phénomène participe à l'augmentation de la proportion de l'information de mouvement : la variation d'efficacité de la prédiction des vecteurs. En effet à haut débit les partitions de petites tailles sont plus nombreuses, le mode Skip (ou Direct) est moins fréquent, par conséquent le champ de vecteurs est plus dense. Les vecteurs sont plus proches spatialement, ils sont donc mieux prédits. A bas débit, le mode Skip (ou Direct) et les modes de grandes tailles de partition ont une sélection élevée, les vecteurs sont plus éloignés spatialement et donc moins bien prédits qu'à haut débit. De plus, ces vecteurs sont moins précis à cause de la non transmission d'erreurs de prédictions induite par le mode Skip (ou Direct) et par l'augmentation du poids du débit dans le critère débit-distorsion. Lorsque l'on calcule l'entropie des résiduels de vecteurs mouvement sur la séquence *Bigships* 720p 60 Hz on obtient une entropie égale à 2,37 pour le QP 17 et à 2,83 pour le QP 42. Alors qu'au contraire l'entropie des résiduels de texture quantifiés (et transmis) diminue avec l'augmentation du QP. Une solution pour réduire partiellement l'entropie des vecteurs mouvement à

QP	P Slices				B Slices			
	Inter			Intra	Inter			Intra
	Skip	16×16	8×8 8×4	16×16	Direct	16×16	8×8 8×4	16×16
		16×8 8×16	4×8 4×4	8×8 4×4		16×8 8×16	4×8 4×4	8×8 4×4
17	5,8	43,8	33,3	17,1	38,9	41,7	16,9	2,5
22	12,8	48,9	26,2	12,1	53,4	33,8	11,6	1,2
27	25,1	44,1	19,4	11,4	66,8	26,4	5,9	0,9
32	39,0	38,5	11,6	10,9	75,6	21,5	2,2	0,7
37	52,3	32,2	5,1	10,4	82,3	16,6	0,6	0,5
42	64,9	24,4	1,4	9,3	87,0	12,6	0,1	0,3
<b>Moyenne</b>	33,3	38,6	16,2	11,9	67,3	25,5	6,2	1,0

TABLE 3.2 – Pourcentage de sélection des modes de codage pour les slices P et B de la configuration High IBBP calculé en moyenne sur l'ensemble des séquences VCEG.

bas débit, serait d'utiliser des prédicteurs temporels des vecteurs à la place du prédicteur spatial, car la distance temporelle entre les images (par conséquent, la distance entre le vecteur courant et ses prédicteurs), n'évolue pas en fonction du QP. De plus, les corrélations temporelles entre les champs de vecteurs mouvement sont importantes, comme le montre la figure 3.7 représentant la composante horizontale des vecteurs mouvement extraite d'un codage IPP pour deux images successives de la séquence *Foreman* CIF 30Hz.

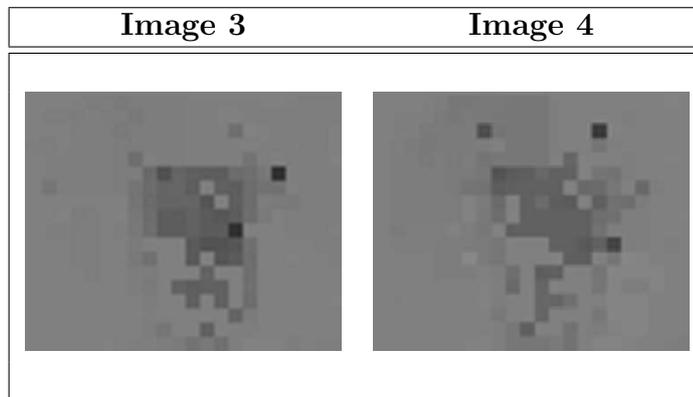


FIGURE 3.7 – Représentation visuelle des valeurs de la composante  $X$  des vecteurs mouvement des images 3 et 4 de la séquence *Foreman* CIF 30Hz codée avec le profil Baseline IPP. Pour la représentation visuelle les valeurs de la composante horizontale sont centrées en 128.

La représentation de la sélection spatiale des modes est montrée dans la figure 3.8 pour la séquence *Foreman*. Ces trois images sont celles utilisées dans la figure 3.3, pour les mêmes QP. Pour l'image P, le mode Skip est sélectionné dans les zones ayant peu de mouvement et les zones planes.

En effet, les autres modes Inter sont sélectionnés dans les parties qui ont plus de mouvement (le visage) mais aussi pour des blocs contenant des hautes fréquences. Par exemple, certains macroblocs correspondant aux rayures du mur sont codés en Inter non Skip alors que des macroblocs voisins qui ont le même mouvement sont codés en Skip. Le meilleur exemple est le casque qui a le même mouvement que le reste du visage mais dont l'intérieur est complètement skippé. Toujours dans cette image P on note que les modes de codage Intra apparaissent dans les zones qui subissent des déformations ou des zones qui ne faisaient pas partie de la scène dans les images précédentes. Par exemple la bouche qui subit une déformation d'une image à l'autre en plus du mouvement du visage, nous noterons aussi le bas du cou qui est modifié par le changement d'éclairage lié au mouvement de la tête.

Dans l'image B de la figure 3.8 très peu de modes autres que le mode Direct ont été sélectionnés. On remarquera que les modes Inter hors Direct ont été sélectionnés dans des parties en déformation par rapport aux images précédentes. Notons que c'est spécialement ces parties qui ont un coût de codage plus élevé en débit que le reste de l'image, comme on a pu le voir dans la figure 3.3.

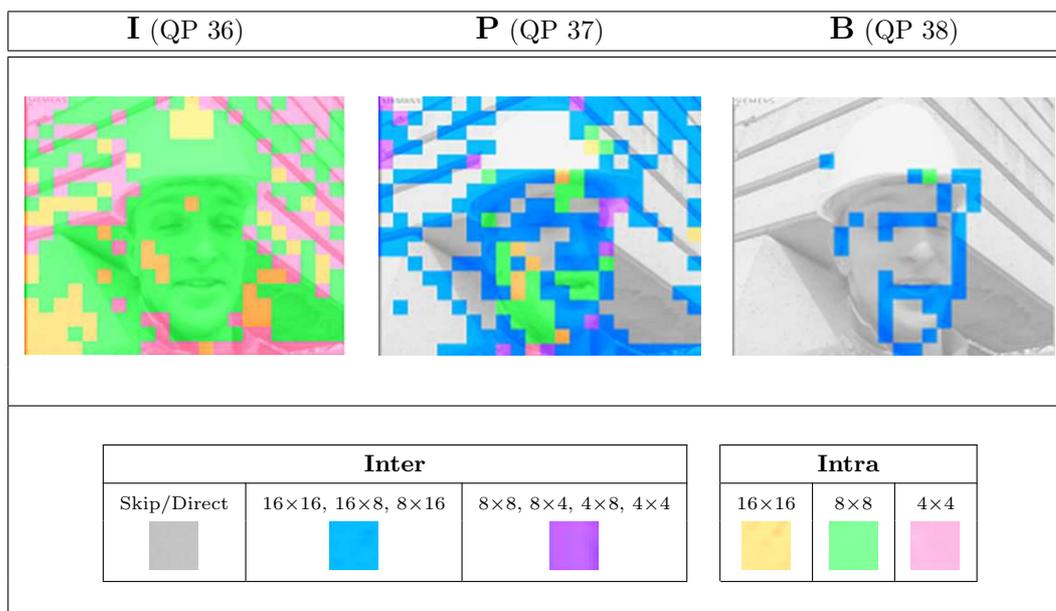


FIGURE 3.8 – Représentation spatiale de la sélection de 3 images de la séquence *Foreman* CIF 30Hz, codées respectivement en I, P, B avec les QP respectivement de 36, 37, 38.

### 3.3.2 Slices Intra

Le tableau 3.3 donne le pourcentage de sélection des modes Intra en fonction des résolutions pour l'ensemble des séquences VCEG. On remarque que le mode Intra 4×4 est le plus sélectionné pour les résolutions QCIF et CIF. Mais pour la résolution 720p, l'Intra 8×8 devient le mode le plus sélectionné avec 63,3% de sélection contre 15,3% pour l'Intra 4×4. Cette augmentation de sélection de

Résolution	Intra 16×16	Intra 8×8	Intra 4×4
<b>QCIF</b>	10,2	41,1	48,7
<b>CIF</b>	8,6	34,5	56,9
<b>720p</b>	21,4	63,3	15,3
<b>Moyenne</b>	14,4	48,1	37,5

TABLE 3.3 – Pourcentage de sélection des modes Intra 16×16, 8×8 et 4×4 en fonction des résolutions pour la configuration High III, en moyenne pour l'ensemble des séquences VCEG pour six QP (17-42).

CrowdRun	Intra 16×16	Intra 8×8	Intra 4×4
<b>480×270</b>	9,9	23,3	66,8
<b>960×540</b>	11,7	39,4	48,9
<b>1920×1080</b>	16,4	61,9	21,7
<b>3840×2160</b>	18,7	66,2	15,1

TABLE 3.4 – Pourcentage de sélection des modes Intra 16×16, 8×8 et 4×4 pour la séquence *CrowdRun* 50Hz à différentes résolutions spatiales (même nombre d'images) pour la configuration High III, en moyenne pour six QP (17-42).

l'Intra 8×8 par rapport à l'Intra 4×4 est due au fait que les séquences à basse résolution contiennent proportionnellement plus de hautes fréquences que les séquences à haute résolution, sachant que les hautes fréquences sont mieux prédites avec un partitionnement plus petit des macroblocs. Il faut aussi ajouter que ceci est vrai si l'on considère que les types de contenus filmés ne sont pas différents d'une résolution à l'autre. C'est-à-dire que l'on ne change pas les usages de capture des images (on filme les mêmes scènes) mais de visionnage (on utilise des résolutions et des écrans plus grands). Pour se convaincre de cette remarque, regardons la sélection des modes Intra pour la séquence *CrowdRun* pour différentes résolutions, reportées dans le tableau 3.4. Les résolutions testées vont du 3840×2160 (2160p) à une résolution 480×270. Pour les besoins de ce test, nous avons réduit la séquence originale sans enlever les bords, au risque de ne pas obtenir des résolutions standard. Cette réduction a été effectuée à l'aide du filtre utilisé dans le logiciel de référence de l'annexe scalable H.264/SVC, ceci permettant de comparer le même contenu à des tailles différentes. Dans ce tableau 3.4 on remarque que la proportion de la sélection des Intra 4×4 baisse lorsque la résolution augmente passant de 66,8% pour la résolution 480×270 à 15,1% pour la résolution 3840×2160 à l'opposé de la sélection de l'Intra 8×8 qui passe elle de 23,3% à basse résolution à 66,2% pour la résolution maximale. Notons que la sélection de l'Intra 16×16 augmente elle aussi avec la résolution, cependant sa croissance est moins forte (à peine 2 fois plus élevée contre 3 fois plus élevée pour l'Intra 8×8).

Résolution	Baseline III			High III		
	17-32	22-37	27-42	17-32	22-37	27-42
<b>QCIF</b>	13,0	17,9	23,9	17,0	23,7	32,9
<b>CIF</b>	6,6	8,2	9,7	9,3	13,2	18,1
<b>720p</b>	-4,4	-5,5	-8,1	10,2	12,9	15,8
<b>Moyenne</b>	3,6	4,9	5,9	11,6	15,7	20,8

TABLE 3.5 – Pourcentage de réduction de débit des configurations Baseline III et High III par rapport à JPEG 2000 en fonction des résolutions, pour trois ensembles de QP calculés sur l’ensemble des séquences VCEG.

### 3.4 H.264/AVC et JPEG 2000

Une partie de nos contributions concerne l’amélioration du codage Intra. Afin de comparer ultérieurement nos résultats à la norme de référence du codage d’image, nous présentons, dans cette section, une étude d’efficacité de codage de la norme H.264/AVC et de la norme JPEG 2000. Les résultats proposés sont donnés pour le profil Baseline III et le profil High III générés avec le logiciel de référence de la norme H.264/AVC. Pour obtenir les résultats de la norme JPEG 2000 nous avons utilisé le logiciel KAKADU [KAKADU] avec les options suivantes :

- taille des Codeblocks égale à  $64 \times 64$
- 1 tuile par image.
- respectivement 3, 4, et 6 niveaux de décomposition pour les résolutions QCIF, CIF et 720p.
- pour chacun des tests le débit est fixé à partir du débit obtenu avec la configuration H.264/AVC profil Baseline III.

Pour chaque séquence, le même nombre d’images est codé pour toutes les configurations. Pour JPEG 2000 nous comparons donc la somme des débits et la moyenne des PSNR. Les séquences utilisées sont celles de l’ensemble VCEG. Les résultats en pourcentage de gains des configurations H.264/AVC par rapport à JPEG 2000 sont présentés par résolution dans le tableau 3.5. Nous remarquons que le profil High est meilleur que JPEG 2000 pour toutes les résolutions et il est meilleur pour toutes les séquences. Le gain de ce profil par rapport à JPEG 2000 atteint 20,8% en moyenne à bas débit. Pour les deux profils les gains sont en moyenne plus élevés à bas débit qu’à haut débit. Pour les résolutions QCIF et CIF les deux profils d’H.264/AVC sont plus efficaces que JPEG 2000. Pour la résolution 720p le profil Baseline III est moins efficace que JPEG 2000, mais en moyenne pour toutes les séquences le Baseline reste plus performant. Nous retrouvons donc les résultats de l’état de l’art présenté dans le chapitre 1. La supériorité de JPEG 2000 par rapport au Baseline pour les séquences 720p provient de l’inefficacité du codage Intra  $4 \times 4$  pour cette résolution, comme nous avons pu le constater dans la section précédente. Enfin, il faut remarquer que nous avons présenté uniquement une étude d’efficacité de codage. Les deux configurations de la norme H.264/AVC et la norme JPEG 2000 ont cependant des complexités de calculs différentes. Nous n’avons pas mesuré ces différences délibérément, car les logiciels utilisés ne sont pas optimisés et ne reflètent absolument pas la complexité réelle des deux normes étudiées.

### 3.5 Conclusion et résumé du chapitre

L'étude comparative de l'efficacité de codage des trois types de slice de la norme H.264/AVC a montré que les slices I ont une proportion élevée dans le débit total d'un flux vidéo. L'amélioration du codage des slices Intra paraît être une voie de recherche essentielle pour améliorer l'efficacité de codage global du codeur. Nous proposons dans ce sens deux méthodes pour l'amélioration de ce codage.

La représentation spatiale du débit, de la distorsion et du critère RD, nous a révélé que les zones contenant des hautes fréquences ont un coût RD élevé par rapport aux zones contenant des surfaces planes. Des nouveaux modes de codage améliorant ces zones coûteuses devraient augmenter les performances. Parmi nos contributions, nous avons développé des nouveaux modes de codage adaptés à certaines de ces hautes fréquences (cf. chapitre 6).

L'évolution des modes de codage en fonction des QP montre que les partitionnements de petite taille sont plus souvent sélectionnés à haut débit. Les modes à très faible débit tel que le mode Skip et le mode Direct représentent, eux, la majorité des sélections à bas débit. Les modes contenant des résiduels de texture ou de vecteurs mouvement sont sélectionnés dans les parties des séquences contenant des mouvements linéaires ou des déformations. Nous avons aussi remarqué dans ce chapitre que la proportion élevée de mouvement à bas débit était liée à une augmentation de l'entropie de cette information engendrée par l'éloignement spatial des vecteurs utilisés pour la prédiction. L'utilisation de prédicteurs temporels pour les vecteurs mouvement devrait améliorer l'efficacité de codage. Une amélioration du codage de l'information de mouvement est une voie de recherche prioritaire. Pour répondre à ce besoin, quatre chapitres concernent directement ou indirectement l'amélioration du codage de cette information (cf. chapitres 4, 5, 7, 9).

La répartition des informations de codage dans le flux pour chacun des types de slice évolue en fonction du QP utilisé pour le codage lié notamment aux  $\lambda_{Mode}$  et  $\lambda_{Motion}$  de l'optimisation du critère débit-distorsion. La réduction de l'information de texture est une voie de recherche essentielle pour les hauts débits, mais à bas débit l'amélioration du codage des informations liées à la compétition devient nécessaire. La troisième partie de ce manuscrit est dédiée à l'amélioration du codage des ces informations (cf. chapitres 7, 8, 9).



Deuxième partie

Modules de codage par  
compétition



## Chapitre 4

# Codage de l'information de mouvement par compétition de prédicteurs

*L'efficacité de codage de la norme H.264/AVC réside, entre autres, sur une exploitation très efficace des redondances temporelles entre les images. L'intégration d'outils tels que les multiples tailles de blocs, les multiples images de références, le  $\frac{1}{2}$  et le  $\frac{1}{4}$  de pixel pour la compensation de mouvement permettent à cette norme de dépasser l'efficacité de ses prédécesseurs. Ces outils entraînent un accroissement de la proportion de l'information de mouvement, engendré à la fois par l'amélioration du codage de la texture et par l'augmentation qualitative et quantitative de cette information de mouvement. Cette information atteint 40% du débit total pour un QP égal à 42. Par conséquent, l'amélioration du codage de l'information de mouvement est une voie de recherche essentielle pour l'augmentation de l'efficacité de codage.*

*Dans le but d'améliorer le codage de cette information de mouvement, nous proposons, dans ce chapitre, une méthode de compétition exhaustive de prédicteurs de vecteurs mouvement pour le mode Skip et pour la prédiction des vecteurs. La méthode permet d'exploiter à la fois les redondances spatiales et temporelles des champs de vecteurs mouvement. La sélection entre les prédicteurs est basée sur le critère débit-distorsion.*

### 4.1 Prédicteurs

L'efficacité d'une méthode de codage par compétition réside dans l'efficacité conjointe des codages mis en compétition. Afin de déterminer cette efficacité conjointe, nous avons défini plusieurs prédicteurs spatiaux, temporels et spatio-temporels. Le but n'est pas d'utiliser l'ensemble des prédicteurs définis ci-dessous pour la compétition, mais de déterminer les meilleurs, c'est à dire

---

1. Publications de l'auteur associées : [Revue2], [Conf5], [Conf6], [ConfFR3], [Contrib1], [Contrib2], [Contrib3], [Brev1], [Brev2], [Brev10], [Expo2] et [Expo3].

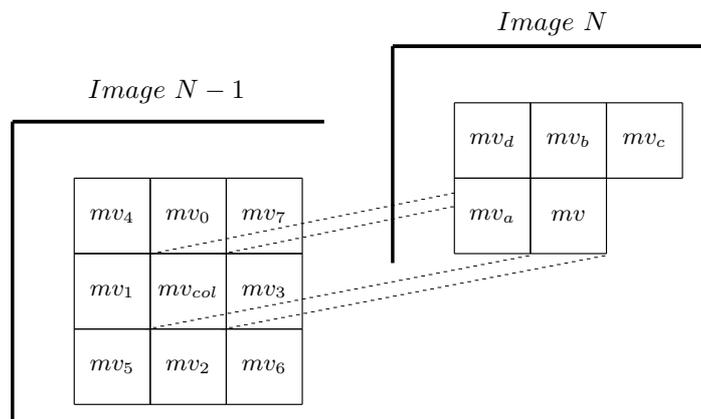


FIGURE 4.1 – Localisation des vecteurs spatiaux et temporels utilisés.

ceux qui exploitent le mieux les corrélations spatiales ou temporelles et qui permettent de réduire l'entropie des résiduels de vecteurs mouvement. Afin d'augmenter l'efficacité des prédictions nous proposons aussi des adaptations de certains prédicteurs aux options des multiples images de référence, des images B et des images B hiérarchiques. Les prédicteurs définis ci-dessous sont basés sur les champs de vecteurs mouvement de l'image courante ou ceux des images précédemment encodées, afin d'exploiter à la fois les redondances spatiales et temporelles de ces champs de vecteurs.

#### 4.1.1 Prédicteurs Spatiaux

Le vecteur prédicteur de la norme H.264/AVC  $mv_{H.264}$ , comme nous l'avons vu dans l'état de l'art, est un prédicteur spatial que nous utiliserons dans notre méthode. A ce prédicteur, nous avons ajouté dans l'ensemble des prédicteurs, les vecteurs mouvement voisins du vecteur courant  $mv$  :  $mv_a$ ,  $mv_b$ ,  $mv_c$ ,  $mv_d$  représentés dans la figure 4.1. Ces vecteurs sont ceux utilisés pour le calcul de  $mv_{H.264}$ . Le calcul du prédicteur  $mv_{H.264}$  dépend de la taille du bloc courant, de la disponibilité des vecteurs voisins ainsi que de l'image pointée par le vecteur courant et celles pointées par les vecteurs voisins. Afin d'avoir dans notre ensemble un prédicteur moins restrictif, nous avons défini un médian spatial dit "étendu"  $mv_{spaEXT}$ . Ce prédicteur est le médian de  $mv_a$ ,  $mv_b$ ,  $mv_c$ , pour chacune des composantes, si les blocs  $a$ ,  $b$  et  $c$  appartiennent à l'image. Sinon, le prédicteur est égal soit à  $mv_a$ , soit à  $mv_b$ , soit à  $mv_c$ , en fonction des disponibilités des blocs  $a$ ,  $b$  et  $c$ . Si aucun des ces blocs n'est disponible, le prédicteur est égal à la valeur 0.

#### 4.1.2 Prédicteurs Temporels

Les prédicteurs temporels que nous avons insérés dans notre ensemble de prédicteurs sont des vecteurs ou des combinaisons de vecteurs se trouvant dans les images précédemment encodées (passées ou passées et futures ou futures pour les images B). Le vecteur temporel "collocated"  $mv_{col}$  est le vecteur utilisé pour le codage du bloc ayant la même position que le bloc courant dans l'image précédente (cf. figure 4.1). Il faut noter que le bloc collocated est le bloc voisin temporel le plus proche du bloc courant.

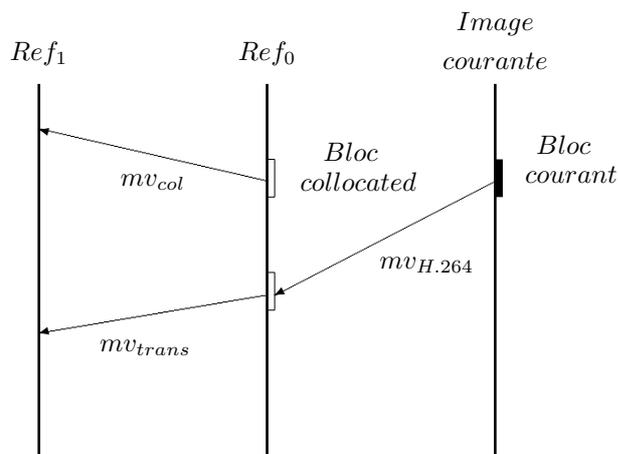


FIGURE 4.2 – Exemple de localisation des vecteurs  $mv_{H.264}$ ,  $mv_{col}$  et  $mv_{trans}$  pour une image de référence.

Afin de suivre au mieux le mouvement d'un objet, nous avons défini le vecteur temporel translaté  $mv_{trans}$  illustré dans la figure 4.2. Ce vecteur est le vecteur mouvement utilisé pour le codage du bloc pointé par le vecteur  $mv_{H.264}$  dans l'image précédente. Ce prédicteur peut fournir une prédiction temporelle plus efficace que le vecteur  $mv_{col}$ , si le vecteur  $mv_{H.264}$  pointe sur un bloc de l'image précédente qui appartient au même objet que celui contenu dans le bloc courant (ou plus exactement à la projection temporelle de cet objet) et que le bloc collocated n'appartient pas à cet objet.

Nous avons ajouté à l'ensemble des prédicteurs temporels deux médians de vecteurs temporels  $mv_{tm5}$  et  $mv_{tm9}$  donnés par les équations suivantes :

$$mv_{tm5} = \text{median}\{mv_{col}, \{mv_i, 0 \leq i \leq 3\}\} \quad (4.1)$$

$$mv_{tm9} = \text{median}\{mv_{col}, \{mv_i, 0 \leq i \leq 7\}\}, \quad (4.2)$$

où les vecteurs  $mv_i, 0 \leq i < 8$  sont les vecteurs voisins du vecteur  $mv_{col}$  représentés dans la figure 4.1.

### 4.1.3 Prédicteur spatio-temporel

La compétition de prédicteurs permet d'exploiter à la fois les redondances spatiales et temporelles si au moins un prédicteur spatial et un prédicteur temporel appartiennent à l'ensemble des prédicteurs en compétition. Afin d'exploiter les redondances spatio-temporelles, nous avons défini un prédicteur qui est une combinaison des vecteurs voisins spatiaux et du vecteur collocated donné par la formule suivante :

$$mv_{spt} = \text{median}\{mv_{col}, mv_{col}, mv_a, mv_b, mv_c\} \quad (4.3)$$

Notons que ce médian pondéré donne plus d'importance à la valeur du vecteur  $mv_{col}$  qu'un simple médian.

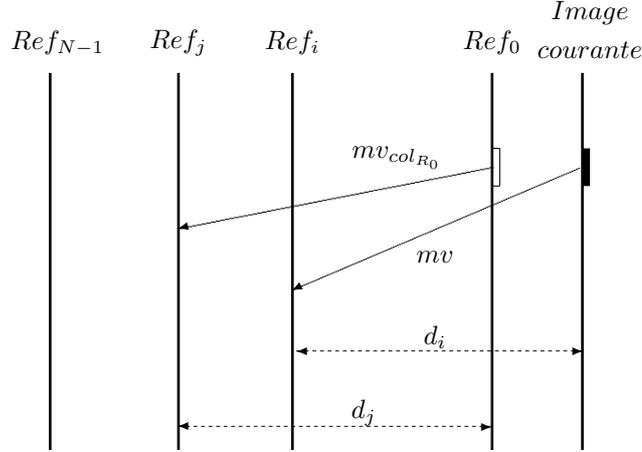


FIGURE 4.3 – Vecteur mouvement collocated de l'image de référence  $Ref_0$  pointant sur l'image  $Ref_j$  et le vecteur courant  $mv$  pointant sur l'image  $Ref_i$ .

#### 4.1.4 Adaptation des prédicteurs pour les images de référence multiple

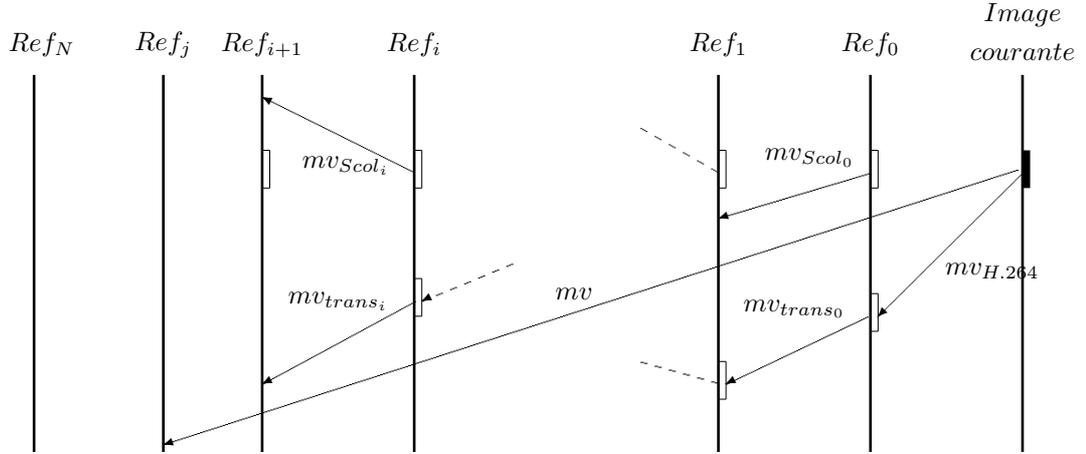
L'option des images de référence multiples pour la compensation de mouvement [WZG99] engendre des disparités dans les champs de vecteurs. En effet, deux vecteurs voisins peuvent pointer sur deux images distinctes. Par conséquent, la distance temporelle parcourue par deux vecteurs voisins peut être différente, modifiant ainsi les corrélations spatio-temporelles entre les champs de vecteurs mouvement. Dans le standard H.264/AVC ce problème est en partie résolu avec le changement de valeurs du prédicteur  $mv_{H.264}$  en fonction de l'image de référence courante et celles des vecteurs voisins (cf. chapitre 1). Les prédicteurs que nous avons proposés dans les paragraphes précédents ont besoin d'être adaptés pour cette option.

En supposant que le mouvement d'un objet est constant, nous proposons une mise à l'échelle des prédicteurs en fonction des distances temporelles parcourues par les vecteurs utilisés pour le calcul des prédicteurs et celle parcourue par le vecteur courant. Dans la figure 4.3, le vecteur collocated de la première image de référence  $mv_{col_{R_0}}$  pointant sur l'image de référence numéro  $j$ ,  $Ref_j$  est représenté ainsi que le vecteur courant  $mv$  pointant sur l'image de référence numéro  $i$ ,  $Ref_i$ . Le vecteur  $mv_{col_{R_0}}$  est mis à l'échelle en fonction de la distance temporelle parcourue par le vecteur courant ( $d_i$ ) et celle parcourue par ce vecteur  $mv_{col_{R_0}}$  ( $d_j$ ). Le prédicteur mis à l'échelle que l'on appellera  $mv_{Scol_{R_0}}$  est donné par la formule suivante :

$$mv_{Scol_{R_0}} = \frac{mv_{col_{R_0}}}{d_j} \times d_i. \quad (4.4)$$

Les autres prédicteurs temporels et spatiaux proposés peuvent être mis à l'échelle de la même manière.

Nous proposons, en plus de cette mise à l'échelle des prédicteurs, un nouveau prédicteur basé sur une somme de vecteurs collocated. Pour la définition de ce prédicteur, considérons que tous les vecteurs mouvement de toutes les images de référence pointent seulement sur l'image précédente. C'est-à-dire que la distance temporelle parcourue par chaque vecteur est égale à  $d_0$ . Ceci peut être

FIGURE 4.4 – Vecteur mouvement collocated  $mv_{Scol_i}$  et temporel traduit  $mv_{trans_i}$ .

obtenu en appliquant une mise à l'échelle de l'ensemble des vecteurs. Dans ce cas on note  $mv_{Scol_i}$  le vecteur collocated mis à l'échelle de l'image  $Ref_i$  pointant sur l'image  $Ref_{i+1}$  représenté dans la figure 4.4. Le prédicteur  $mv_{colsum_j}$  pointant sur l'image de référence courante  $Ref_j$  (image courante jusqu'à l'image  $Ref_j$ ) est la somme de ces vecteurs  $mv_{Scol_i}$  défini par la formule suivante :

$$mv_{colsum_j} = \sum_{i=0}^{i \leq j} mv_{Scol_i}, \quad j < N \quad (4.5)$$

Selon le même principe, nous proposons une somme de vecteurs traduits. En considérant que tous les vecteurs pointent sur l'image qui les précède, le prédicteur  $mv_{transsum_j}$ , pointant sur l'image de référence courante  $Ref_j$ , est la somme des vecteurs  $mv_{trans_i}$  représenté dans la figure 4.4. Le prédicteur  $mv_{transsum_j}$ , est obtenu par la formule suivante :

$$mv_{transsum_j} = \sum_{i=0}^{i \leq j} mv_{trans_i}, \quad j < N \quad (4.6)$$

où  $mv_{trans_i}$  est le vecteur mouvement du bloc à la position donnée par  $mv_{trans_{i-1}}$  dans l'image  $Ref_{i-1}$  pointant sur  $Ref_i$ , à l'exception de  $mv_{trans_0}$  qui est donné par la valeur du prédicteur  $mv_{H.264}$  pointant sur l'image  $Ref_0$ .

#### 4.1.5 Adaptation des prédicteurs pour les slices bidirectionnels

Le codage Inter des slices bidirectionnels (slices B) utilise les corrélations temporelles entre l'image courante et celles dans le futur et le passé. Par conséquent, les caractéristiques des champs de vecteurs sont différentes des slices P. En effet, les vecteurs des images B ne pointent pas tous dans la même direction (passée et future) et ne parcourent pas tous la même distance temporelle. Par conséquent, les seules corrélations spatiales du champ de vecteurs de mouvement courant ne permettent pas un codage optimal de l'information de mouvement. Dans cette section, nous allons définir et redéfinir

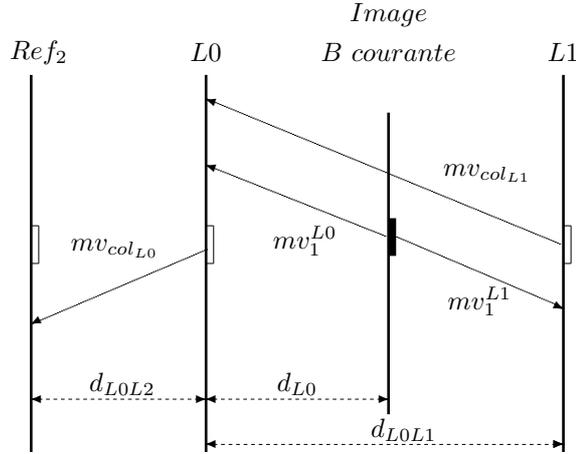


FIGURE 4.5 – Prédicteurs du mode Direct Temporel des slices B, et les distances temporelles associées à la mise à l'échelle.

des prédicteurs temporels basés sur les champs des vecteurs des images P futures ( $L1$ ), des images P passées ( $L0$ ) et des images B précédemment encodées  $B - 1$ .

La prédiction d'un bloc d'une image B peut provenir d'un ou de deux blocs sélectionnés dans une ou deux images P. Si l'on suppose qu'un objet se déplace avec une vitesse constante, le champ de vecteurs mouvement d'une image B est lié au champ de vecteurs mouvement de l'image P dans le futur, car les vecteurs de cette image P traversent l'ensemble des images B. Les prédicteurs définis pour le mode Direct temporel exploitent cette propriété en utilisant le vecteur  $mv_{col_{L1}}$  (provenant de l'image P future  $L1$ ) comme vecteur de base pour le calcul des prédicteurs  $mv_1^{L0}$  et  $mv_1^{L1}$  pointant respectivement dans la direction passée (image  $L0$ ) et la direction future (image  $L1$ ). Ces prédicteurs sont donnés dans les équations (1.8) et (1.9) du chapitre 1. Dans la suite de ce chapitre, nous nommerons ces deux prédicteurs par le même nom  $mv_{Scol_{L1}}$ , afin d'identifier que ces prédicteurs proviennent du vecteur collocated  $mv_{col_{L1}}$  mis à l'échelle.

Selon la même mise à l'échelle, nous utilisons le vecteur collocated  $mv_{col_{L0}}$  de l'image P dans le passé  $L0$  pour le calcul du prédicteur nommé  $mv_{Scol_{L0}}$ . Ce prédicteur nomme indifféremment les deux prédicteurs  $mv_2^{L0}$  et  $mv_2^{L1}$ , pointant respectivement dans l'image passée et l'image future et définis par les équations suivantes :

$$mv_2^{L0} = \frac{mv_{col_{L0}}}{d_{L0L2}} \times d_{L0} \quad (4.7)$$

$$mv_2^{L1} = \frac{mv_{col_{L0}}}{d_{L0L2}} \times (d_{L0} - d_{L0L1}) \quad (4.8)$$

où  $Ref_2$  est l'image pointée par  $mv_{col_{L0}}$  et  $d_{L0L2}$  la distance temporelle entre l'image de référence passée  $L0$  et  $Ref_2$ , comme représenté dans la figure 4.5.

Si l'on fait l'analogie avec les images (matrices de pixels), les corrélations temporelles entre deux champs de vecteurs mouvement devraient augmenter lorsque la distance temporelle entre ces champs de vecteurs diminue. Dans le cas où  $N$  images sont codées entre deux images P, la distance temporelle entre l'image B courante et l'image P dans le futur  $L1$  est toujours plus élevée que la

distance temporelle entre l'image B précédente, excepté dans le cas où l'image B courante est la dernière image B (image B numéro  $N$ ). Dans ce cas la distance temporelle est la même. Cependant les distances parcourues par les vecteurs de l'image B précédente ( $B - 1$ ) sont plus faibles que celles parcourues par les vecteurs de  $L1$ . Par conséquent, les vecteurs de l'image B courante devraient être plus corrélés avec les vecteurs de l'image  $B - 1$  que ceux de l'image  $L1$ . Nous avons donc défini des prédicteurs qui utilisent le champ de vecteurs de l'image B précédente  $B - 1$  (non disponible pour la première image B). Si l'on considère le cas où le vecteur mouvement du bloc collocated de l'image  $B - 1$  pointe uniquement sur l'image passée  $L0$ , le vecteur  $mv_{col_{B-1L0}}$ , représenté dans la figure 4.6, est utilisé pour le calcul des prédicteurs  $mv_3^{L0}$  et  $mv_3^{L1}$  pointant respectivement sur l'image  $L0$  et l'image  $L1$  et défini par les équations suivantes :

$$mv_3^{L0} = \frac{mv_{col_{B-1L0}}}{d_{L0B-1}} \times d_{L0} \quad (4.9)$$

$$mv_3^{L1} = \frac{mv_{col_{B-1L0}}}{d_{L0B-1}} \times (d_{L0} - d_{L0L1}) \quad (4.10)$$

où  $d_{L0B-1}$  est la distance temporelle entre l'image P passée  $L0$  et l'image  $B - 1$ . Nous nommerons  $mv_{Scol_{B-1L0}}$ ; ces deux vecteurs  $mv_3^{L0}$  et  $mv_3^{L1}$ .

Si l'on considère maintenant le cas où le vecteur  $mv_{col_{B-1L1}}$  (cf. figure 4.6) de l'image  $B - 1$ , pointant sur l'image P future  $L1$ , est disponible, nous pouvons définir les vecteurs  $mv_4^{L0}$  et  $mv_4^{L1}$  qui pointent respectivement sur l'image  $L0$  et l'image  $L1$ . Ces vecteurs sont donnés par les équations suivantes :

$$mv_4^{L0} = \frac{mv_{col_{B-1L1}}}{d_{L0B-1} - d_{L0L1}} \times d_{L0} \quad (4.11)$$

$$mv_4^{L1} = \frac{mv_{col_{B-1L1}}}{d_{L0L1} - d_{L0B-1}} \times (d_{L0L1} - d_{L0}) \quad (4.12)$$

Notons que les vecteurs  $mv_4^{L0}$  et  $mv_4^{L1}$  devraient être plus efficaces que  $mv_3^{L0}$  et  $mv_3^{L1}$  car le vecteur  $mv_{col_{B-1L1}}$  traverse le champ de vecteur mouvement de l'image B courante. Nous utiliserons la notation  $mv_{Scol_{B-1L1}}$  pour les deux vecteurs  $mv_4^{L0}$  et  $mv_4^{L1}$ . Enfin, nous avons défini le prédicteur  $mv_{Bcol}$  pour l'image  $B_1$ . Il est égal au vecteur  $mv_{Scol_{L1}}$  si le bloc collocated n'est pas codé en Intra. Sinon ce prédicteur est égal à la valeur  $mv_a$ . Notons que dans le standard H.264/AVC, un slice B peut avoir des images de référence multiples dans le passé et dans le futur. Les mises à l'échelle et les prédicteurs définis dans la section 4.1.4 peuvent aussi être utilisés dans le cas des slices B.

#### 4.1.6 Adaptation des prédicteurs pour les slices B hiérarchiques

Dans le standard H.264/AVC il est possible d'utiliser des images B ordonnées hiérarchiquement [SMW05]. Pour cet ordre de codage, les images B du niveau inférieur de prédiction de la "pyramide" ont pour images de référence les images B du niveau supérieur. Dans [JZG<sup>+</sup>04], un mode Direct temporel pour cet ordre particulier est proposé. Les prédicteurs temporels que nous utiliserons pour les slices B hiérarchiques sont équivalents à ceux définis dans les équations (1.8) et (1.9) du chapitre 1 et les équations (4.7), (4.8), (4.9) et (4.10) de ce chapitre.

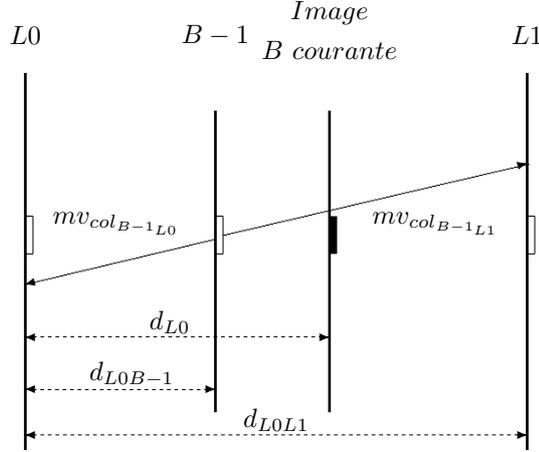


FIGURE 4.6 – Vecteurs mouvement collocated de l'image  $B-1$  et les distances temporelles associées à la mise à l'échelle.

## 4.2 Compétition exhaustive

Dans la méthode de compétition de vecteurs mouvement (MVComp), les prédicteurs définis dans la section précédente vont être mis en compétition. Cette méthode est exhaustive (cf. chapitre 2), dans le sens où toutes les prédictions sont évaluées à l'aide du critère débit-distorsion et que l'indice du meilleur prédicteur est transmis au décodeur. Dans cette méthode de compétition, chaque vecteur mouvement et chaque macrobloc codé en Skip a un prédicteur sélectionné parmi un ensemble de prédicteurs. Les blocs bidirectionnels ont eux deux vecteurs, par conséquent ils ont deux prédicteurs. Le critère débit-distorsion étant différent pour le mode Skip et les autres modes Inter, nous allons expliquer notre méthode pour chacune de ces deux compétitions.

### 4.2.1 Sélection des prédicteurs de vecteurs mouvement des modes Inter

Pour le codage par compétition, le médian de la norme H.264/AVC,  $mv_{H.264}$ , est remplacé par un ensemble  $\mathcal{P}$  composé de  $n$  prédicteurs  $p_i$ . Pour chacun de ces prédicteurs  $p_i$  on calcule le résiduel de vecteur mouvement associé,  $\varepsilon_{mv_i}$ , donné par la formule suivante :

$$\varepsilon_{mv_i} = mv - p_i, \forall i \in \{1, \dots, n\} \quad (4.13)$$

Afin de sélectionner le meilleur prédicteur, le coût de chaque  $\varepsilon_{mv_i}$  et le coût de l'indice  $i$  sont estimés. Le prédicteur qui fournit le coût le plus faible est alors sélectionné. Le débit  $R_{mv}$  de l'équation de débit-distorsion utilisé dans le processus d'estimation de mouvement (équation (2.3) du chapitre 2 ( $J = D + \lambda_{Motion} R_{mv}$ )) ou dans celle de la sélection du meilleur mode de codage (équation (2.4) du chapitre 2), est remplacé par le débit  $R_{mv/mm}$  donné dans la formule suivante :

$$R_{mv/mm} = \min\{\varsigma(\varepsilon_{mv_i}) + \varsigma(i)\}_{i \in \{1, \dots, n\}} \quad (4.14)$$

où  $\varsigma(x)$  est le coût en débit de l'information  $x$ .

Dans le logiciel de référence, ce coût peut être estimé ou être le débit réel, en fonction du mode de codage entropique utilisé (CAVLC ou CABAC) ou de l'étape de codage. Dans notre implémentation, chaque fois que le débit  $R_{mv}$  est estimé dans la référence, celui-ci est remplacé par  $R_{mv/mm}$  pour l'estimation de mouvement et pour la comparaison des modes de codage. Ceci est important car la méthode de codage par compétition des prédicteurs réduit le débit des vecteurs mouvement. Par conséquent l'estimation de mouvement, qui dépend du critère débit-distorsion et donc de l'information de mouvement, ne fournira pas nécessairement le même vecteur mouvement. Notons enfin que si tous les prédicteurs sont égaux, le coût de l'indice  $i$ ,  $\varsigma(i)$ , est nul car le décodeur a les données nécessaires pour tester l'égalité de ces prédicteurs.

### 4.2.2 Sélection des prédicteurs du mode Skip

Comme pour la compétition des prédicteurs de vecteurs mouvement des autres modes Inter, le médian de la norme H.264,  $mv_{H.264}$ , est remplacé par un ensemble  $\mathcal{P}_s$  composé de  $n_s$  prédicteurs  $p_i, \forall i \in \{1, \dots, n_s\}$ . Notons que les ensembles  $\mathcal{P}$  et  $\mathcal{P}_s$  sont distincts. Il n'ont donc pas forcément le même contenu, ni le même nombre de prédicteurs. Pour chacun des prédicteurs de  $\mathcal{P}_s$  on calcule le coût débit-distorsion associé,  $J_{SKIP}^i$ , donné par la formule suivante :

$$J_{SKIP}^i = D_{SKIP}^i + \lambda_{Mode}(\varsigma(i)), \quad \forall i \in \{1, \dots, n_s\} \quad (4.15)$$

où  $D_{SKIP}^i$  est la distorsion liée au prédicteur  $p_i \in \mathcal{P}_s$  et  $\varsigma(i)$  est le coût de codage de l'indice de ce prédicteur. Le prédicteur  $p_i$  qui minimise l'ensemble des coûts  $J_{SKIP}^i$  est sélectionné. Par conséquent, le coût  $J_{SKIP}$  de l'équation (2.5) du chapitre 2 ( $J_{SKIP} = D_{SKIP} + \lambda_{Mode}R_m$ ) est remplacé par l'équation suivante :

$$J_{SKIP} = \min\{J_{SKIP}^i\}_{i \in \{1, \dots, n_s\}} + \lambda_{Mode}R_m \quad (4.16)$$

Comme pour les indices des prédicteurs de vecteurs mouvement des autres modes Inter, le coût de cet indice est nul si l'ensemble des prédicteurs de  $\mathcal{P}_s$  sont égaux. Détaillons maintenant le codage de cette information de compétition.

### 4.2.3 Codage de l'information de compétition

La méthode de codage par compétition proposée nécessite la transmission des indices correspondant aux valeurs des prédicteurs. Pour les modes Inter, excepté pour le mode Skip, ces indices sont écrits juste après le résiduel des vecteurs mouvement. En CAVLC, le coût de l'indice du prédicteur transmis dépend du nombre de valeurs disponibles de l'ensemble  $\mathcal{P}$ . Par exemple, si deux valeurs distinctes existent dans l'ensemble  $\mathcal{P}$ , un seul bit est transmis au décodeur. La même binarisation est utilisée pour le CABAC. Les probabilités des contextes de chaque bit sont les mêmes que celles attribuées aux modes de codage.

Pour le mode Skip un RLC (Run Length Coding) est utilisé en CAVLC (cf. chapitre 1). Les indices des prédicteurs sont donc écrits après l'écriture du nombre de modes Skip successifs. En CABAC, nous utilisons la même binarisation et les mêmes contextes que ceux des indices des prédicteurs de vecteurs mouvement.

Le test d'égalité des prédicteurs permet de supprimer un grand nombre de bits lié à cette information de compétition. Cependant, l'état actuel de l'implémentation est contraignante pour le "parsing" du flux. Cette opération est une lecture du train binaire qui sert à valider le non endommagement du flux. Dans la norme H.264/AVC, cette lecture<sup>1</sup> est opérée slice par slice et dépend uniquement des données du slice courant. Lorsque l'ensemble des prédicteurs des ensembles  $\mathcal{P}$  et  $\mathcal{P}_s$  sont spatiaux, le test d'égalité des prédicteurs augmente la complexité de calcul pour le parsing car il est nécessaire de décoder l'ensemble du champ de vecteurs. Si au moins un prédicteur temporel appartient à l'un des deux ensembles, la valeur du vecteur temporel est nécessaire par conséquent l'indépendance des données du slice courant est incompatible. Ce problème peut être cependant résolu en indiquant le nombre de bits nécessaires au codage des indices de prédiction suivi de l'ensemble de ces indices au début ou à la fin de chaque slice cette méthode nécessite le codage entier du slice avant sa transmission, ce qui n'est pas compatible avec les applications de visiophonie pour lesquelles un codage et une transmission au fil de l'eau sont nécessaires. Cette solution reste cependant viable pour les autres applications.

La réduction du débit de l'information de compétition par égalité des prédicteurs s'inscrit dans une logique d'augmentation des performances de compression basée sur le transfert d'une partie de l'intelligence du codeur vers le décodeur. Ce transfert a pour but de supprimer l'information de compétition comme nous allons le voir dans la partie III. Notons enfin que l'ordonnancement des données pourrait changer dans l'avenir avec des techniques plus évoluées et plus robustes aux pertes de données dues à l'utilisation des réseaux sans fil.

## 4.2.4 Complexité

### 4.2.4.1 Complexité en mémoire

Les prédicteurs spatiaux que nous avons proposés n'engendrent pas de complexité de mémoire supplémentaire. En revanche, les prédicteurs temporels impliquent la sauvegarde des champs de vecteurs mouvement des images de référence. La taille de ces informations dépend de la taille de la fenêtre de recherche et du nombre d'images de référence. En fonction des prédicteurs temporels utilisés, il n'est pas nécessaire de conserver l'ensemble de cette information. Par exemple, pour les vecteurs  $mv_{Scol_{R_0}}$  et  $mv_{trans}$  seul le champ de vecteur de l'image  $Ref_0$  doit être gardé en mémoire. Les prédicteurs  $mv_{col_{sum_j}}$ ,  $mv_{trans_{sum_j}}$  requièrent, au contraire, le stockage de l'ensemble des champs de vecteurs mouvement de toutes les images de référence. Les indices des images de référence indispensables pour le calcul des mises à l'échelle n'ont pas besoin d'être gardés en mémoire car ces calculs peuvent être effectués image par image. Notons que l'information des champs de vecteurs mouvement est déjà en mémoire pour les slices B dans les décodeurs utilisant les profils Main et High du standard H.264/AVC. En effet, le mode Direct temporel utilise cette donnée. L'impact en mémoire supplémentaire est donc pour le profil Baseline et pour les slices P des profils Main et High.

---

1. Le terme lecture est différent du terme décodage : à la fin de la lecture d'un slice, le flux n'est pas transformé en images.

#### 4.2.4.2 Complexité de calcul

L'impact sur la complexité de calcul dépend du nombre et du type de prédicteurs utilisés dans  $\mathcal{P}$  et  $\mathcal{P}_s$ .

Au codeur, chaque prédicteur doit être calculé à partir des champs de vecteurs précédemment codés. Aucune estimation de mouvement supplémentaire par rapport à la référence n'a besoin d'être effectuée. Pour chaque prédicteur  $p_i$ , le résiduel  $\varepsilon_{mv_i}$  est calculé et estimé. Ceci sont les seuls impacts au niveau du codeur. Les autres impacts en complexité sont non-normatifs. Par conséquent, ces impacts dépendent de chaque implémentation. Dans la nôtre, pour chaque prédicteur  $p_i$ , nous calculons :

- le coût de l'indice  $\zeta(i)$
- le coût du résiduel  $\zeta(\varepsilon_{mv_i})$ , qui dépend d'une table de Golomb dans le logiciel de référence
- la distorsion (SAD) pour chaque prédicteur pour le mode Skip.

Dans une implémentation réaliste d'un codeur H.264/AVC, on utilise généralement une sélection priorisée du mode Skip comme choix de codage [Bye03] qui permet une sélection rapide du mode Skip sans tester l'ensemble des autres modes de codage. Avec notre méthode, le nombre de sélections de modes Skip augmente. On devrait donc obtenir un gain en rapidité pour une implémentation réaliste.

Au décodeur, les prédicteurs doivent être comparés afin de déterminer leur éventuelle égalité. Ce test implique le calcul de tous les prédicteurs des ensembles  $\mathcal{P}$  et  $\mathcal{P}_s$ . Ce sont les seuls impacts sur la complexité. De plus, le décodage d'un macrobloc sauté ne requiert aucun parsing du flux ni de déquantification, ni de dé-transformation. Ce mode de codage a donc une complexité faible. Notre méthode augmente le nombre de sélections de ce mode. Par conséquent la complexité devrait être réduite.

La compétition des vecteurs mouvement a été intégrée dans le JM KTA [KTA]. Ce logiciel est basé sur le logiciel de référence de la norme H.264/AVC [JM]. Ce code C n'est pas optimisé, ni en termes de complexité en mémoire, ni en termes de complexité de calcul. Ce logiciel est capable de coder et de décoder l'ensemble des profils et toutes les options. Il intègre des choix de codage complexes et irréalistes pour un codeur temps réel, comme l'estimation de mouvement entière de la fenêtre de recherche au pixel entier et le calcul du débit au bit près pour l'ensemble des modes de codage. Ces outils non normatifs sont très performants en terme d'efficacité de codage mais très coûteux en terme de complexité de calcul. De plus le JM génère de nombreuses statistiques sur les codages et décodages effectués ce qui augmente encore les temps de calcul. Par conséquent, mesurer une complexité de calcul dans ce logiciel n'a pas de sens. Toutefois une étude d'augmentation du nombre de cycles machine par pixel a été effectuée dans [Hor08] pour chacun des outils intégrés dans le JM KTA. La complexité de notre méthode a été estimée avec deux prédicteurs pour le mode Skip et deux prédicteurs pour les autres modes Inter. Ces résultats indiquent que la méthode MVComp est la moins complexe, au niveau du codeur, des méthodes pouvant engendrer une normalisation. Au niveau du décodeur, le nombre de cycles machine est proche de celui de la référence.

Prédicteur	$mv_{H.264}$	$mv_a$	$mv_b$	$mv_c$	$mv_{Scol_{R_0}}$	$mv_{tm5}$	$mv_{tm9}$	$mv_{trans}$	$mv_{spt}$	$mv_{colsum_j}$	$mv_{transsum_j}$
Sélection (%)	12,6	11,1	9,7	7,9	10,6	8,6	7,6	10,4	8,1	7,7	5,7

TABLE 4.1 – Sélection de chaque prédicteur inclus dans l'ensemble  $\mathcal{P}$  pour neuf séquences CIF pour le profil Baseline IPP.

### 4.3 Résultats expérimentaux

L'efficacité d'une méthode de codage par compétition dépend du nombre et de l'efficacité conjointe des prédicteurs proposés. Nous présenterons dans un premier temps une série d'expériences qui nous a permis de déterminer la meilleure configuration pour chacun des ensembles de prédicteurs et pour chaque type d'image. Dans un second temps, nous donnerons les résultats de notre méthode dans une version récente du JM KTA, avant de les analyser pour différentes configurations de codage. Les résultats sont donnés pour des séquences allant de la résolution QCIF jusqu'à la résolution 1080p. Cet ensemble inclut les séquences recommandées par le groupe VCEG.

#### 4.3.1 Sélection des ensembles de prédicteurs

Avant de sélectionner les meilleurs ensembles  $\mathcal{P}$  et  $\mathcal{P}_s$ , nous proposons d'étudier la sélection et l'efficacité de codage de chaque prédicteur. Ces expériences ont été effectuées dans le JM 11.0 de la norme H.264/AVC et proposées dans [LJPP08]. Les pourcentages de réduction de débit sont calculés avec la métrique de Bjøntegaard [Bjø01a] pour les QP 28, 32, 36, 40. Dans cette section, nous reportons uniquement les expériences qui nous ont permis de conclure ou celles qui illustrent nos principales conclusions. Toutefois, pour déterminer les meilleurs ensembles de prédicteurs, nous avons menés de nombreux tests, sur une multitude de séquences et de débits, que nous n'avons pas mentionnés dans ce manuscrit.

##### 4.3.1.1 Efficacité de codage des prédicteurs proposés

Afin de prouver l'efficacité des prédicteurs proposés dans la section 4.1, nous avons examiné le pourcentage de sélection de chaque prédicteur lorsque les onze prédicteurs définis sont inclus dans l'ensemble  $\mathcal{P}$ . Les pourcentages de sélection obtenus sont reportés dans le tableau 4.1. Notons que l'entropie des résiduels de vecteurs mouvement engendrée par l'utilisation conjointe de tous ces prédicteurs est faible par rapport à l'utilisation d'un ou de deux prédicteurs. Cependant cette compétition entraîne des pertes en terme de débit-distorsion car le débit supplémentaire engendré par la transmission des indices des prédicteurs n'est pas compensé par la réduction de débit des résiduels des vecteurs mouvement. Nous pouvons constater, dans ce tableau, que les valeurs des pourcentages de sélection sont proches. Le prédicteur médian de la norme  $mv_{H.264}$  est sélectionné le plus souvent, suivi du vecteur  $mv_a$  et du vecteur temporel  $mv_{Scol_{R_0}}$ .

Prédicteur associé à $mv_{H.264}$	$mv_a$	$mv_b$	$mv_c$	$mv_{Scol_{R_0}}$	$mv_{tm5}$	$mv_{tm9}$	$mv_{trans}$	$mv_{spt}$	$mv_{col_{sum_j}}$	$mv_{trans_{sum_j}}$
Réduction de débit (%)	2,6	3,0	3,0	5,9	5,2	4,8	5,4	4,4	4,5	5,4

TABLE 4.2 – Pourcentage de réduction de débit de chaque compétition d’un prédicteur et du prédicteur  $mv_{H.264}$  inclus dans l’ensemble  $\mathcal{P}$  pour neuf séquences CIF pour le profil Baseline IPP.

La sélection élevée du prédicteur  $mv_a$  ne permet pas de conclure sur son efficacité lorsque celui-ci est mis en compétition. Dans le tableau 4.2 nous donnons le pourcentage de réduction de débit pour dix prédicteurs en compétition avec le médian de la norme  $mv_{H.264}$  pour l’ensemble  $\mathcal{P}$ . Le meilleur résultat est obtenu par la compétition du prédicteur médian et du prédicteur collocated mis à l’échelle  $mv_{Scol_{R_0}}$ . De plus, nous remarquons que la combinaison du prédicteur  $mv_{H.264}$  et d’un prédicteur temporel est toujours meilleure que la combinaison du prédicteur  $mv_{H.264}$  et d’un prédicteur spatial.

Pour le mode Skip, des expériences similaires ont été menées sur l’ensemble  $\mathcal{P}_s$ . Les résultats en terme de sélection sont équivalents à ceux de l’ensemble  $\mathcal{P}$ . Par contre les réductions des débits, en moyenne, sont plus élevées pour la compétition du prédicteur  $mv_{H.264}$  et d’un prédicteur spatial que pour la combinaison du prédicteur  $mv_{H.264}$  et d’un prédicteur temporel. Cependant pour certaines séquences, notamment celles comportant des mouvements rapides, la combinaison du prédicteur  $mv_{H.264}$  et d’un prédicteur temporel est plus efficace comme nous allons le voir dans le tableau 4.7. Dans le chapitre suivant, nous proposerons une adaptation automatique d’ensembles  $\mathcal{P}_s$  pour maximiser les performances de la méthode.

#### 4.3.1.2 Ensembles $\mathcal{P}$ et $\mathcal{P}_s$ des slices P

Afin de sélectionner le nombre de prédicteurs adéquat pour chacun des ensembles  $\mathcal{P}$  et  $\mathcal{P}_s$ , nous avons testé plusieurs ensembles de un, deux et quatre prédicteurs. La meilleure configuration est obtenue avec deux prédicteurs pour la compétition des vecteurs mouvement et deux prédicteurs pour la compétition du mode Skip. Les pourcentages de réduction de débit sont reportés dans le tableau 4.3 pour différentes combinaisons d’ensembles de prédicteurs testés. Les ensembles de prédicteurs  $\mathcal{P}$  des vecteurs mouvement sont  $\mathcal{P}_1 = \{mv_{H.264}\}$ ,  $\mathcal{P}_2 = \{mv_{H.264}, mv_{Scol_{R_0}}\}$  (la meilleure combinaison donnée dans le tableau 4.2) et  $\mathcal{P}_4 = \{mv_{H.264}, mv_a, mv_{Scol_{R_0}}, mv_{tm9}\}$ . Les ensembles de prédicteurs  $\mathcal{P}_s$  pour le mode Skip sont respectivement  $\mathcal{P}_{s1} = \{mv_{H.264}\}$ ,  $\mathcal{P}_{s2} = \{mv_{spaEXT}, mv_a\}$  et  $\mathcal{P}_{s4} = \{mv_{spaEXT}, mv_a, mv_{Scol_{R_0}}, mv_{H.264}\}$ . Notons que pour la quasi totalité des séquences, la configuration  $\mathcal{P}_2$  et  $\mathcal{P}_{s2}$  donne le meilleur résultat. En effet, avec l’utilisation de quatre prédicteurs, le débit des vecteurs mouvement  $R_{mv}$  est réduit par rapport à la configuration avec deux prédicteurs, mais l’augmentation du débit de  $R_{mv/mm}$  dépasse la réduction de débit de  $R_{mv}$ . Le compromis entre

Ensemble de prédicteurs $\mathcal{P}$	Ensemble de prédicteurs $\mathcal{P}_s$	Réduction de débit
$\mathcal{P}_1$	$\mathcal{P}_{s1}$	0
$\mathcal{P}_1$	$\mathcal{P}_{s2}$	2.6
$\mathcal{P}_2$	$\mathcal{P}_{s1}$	3.5
$\mathcal{P}_2$	$\mathcal{P}_{s2}$	<b>5.9</b>
$\mathcal{P}_4$	$\mathcal{P}_{s2}$	4.1
$\mathcal{P}_2$	$\mathcal{P}_{s4}$	4.9
$\mathcal{P}_4$	$\mathcal{P}_{s4}$	3.4

TABLE 4.3 – Pourcentage de réduction de débit pour la compétition des prédicteurs de vecteurs mouvement et la compétition de prédicteurs de vecteurs mouvement pour le mode Skip pour différents ensembles de prédicteurs incluant 1, 2, ou 4 prédicteurs pour neuf séquences CIF pour le profil Baseline IPP.

ces deux débits est expérimentalement meilleur avec la configuration de deux prédicteurs. La même remarque peut être faite pour le mode Skip : la configuration avec quatre prédicteurs engendre une augmentation du nombre de modes Skip mais les gains engendrés sont neutralisés par le débit des indices des prédicteurs.

Dans la suite des expériences, nous utiliserons donc, pour la configuration Baseline IPP et High IPP, deux prédicteurs pour la compétition des prédictions des vecteurs mouvement, le prédicteur  $mv_{H.264}$  et le prédicteur  $mv_{Scol_{R_0}}$  et deux prédicteurs pour la compétition de prédicteurs pour le mode Skip,  $mv_{spaEXT}$  et  $mv_a$ .

#### 4.3.1.3 Ensemble $\mathcal{P}$ des slices B

La présence des slices B dans la configuration High IBBP complexifie la sélection des meilleurs ensembles de prédicteurs. En effet, les vecteurs mouvement des slices B pointent sur les images passées et les images futures. Par conséquent, les adaptations des prédicteurs sont nombreuses comme nous avons pu le voir dans la section 4.1.5. Il faut donc déterminer, parmi tous ces prédicteurs, le meilleur ensemble pour la prédiction des vecteurs mouvement des images B. De plus, l'éloignement des images P engendre une modification de la sélection des partitionnements du macrobloc et donc des redondances spatiales du champ de vecteurs mouvement. Afin de choisir les meilleurs ensembles  $\mathcal{P}$  et  $\mathcal{P}_s$  tentons de répondre aux questions suivantes :

- Est-ce que les prédicteurs sélectionnés pour les images P du profil Baseline IPP sont adaptés à la configuration High IBBP pour lequel la distance temporelle entre deux images P est plus élevée ?
- Quelle est la combinaison de prédicteurs la plus adaptée pour les slices B ? Pouvons-nous utiliser les mêmes prédicteurs que les slices P et exploiter les mêmes redondances ?

Le tableau 4.4 donne le pourcentage de réduction de débit des images P, pour plusieurs combinaisons d'ensembles  $\mathcal{P}$  et  $\mathcal{P}_s$ . La combinaison obtenant le meilleur résultat est celle qui a été sélectionnée pour le profil Baseline IPP. Notons que le gain est cependant réduit, passant de 5,9%

Prédicteurs de l'ensemble $\mathcal{P}$	Prédicteurs de l'ensemble $\mathcal{P}_s$	Réduction de débit
$mv_{H.264} + mv_a$	$mv_{H.264} + mv_a$	0.9%
$mv_{H.264} + mv_a$	$mv_{spaEXT} + mv_a$	1.3%
$mv_{H.264} + mv_{Scol_{R0}}$	$mv_{H.264} + mv_a$	2.3%
$mv_{H.264} + mv_{Scol_{R0}}$	$mv_{spaEXT} + mv_a$	2.5%

TABLE 4.4 – Pourcentage de réduction de débit pour les images P en fonction de chaque combinaison de prédicteurs pour le codage de vecteurs mouvement et du mode Skip pour neuf séquences CIF pour le profil High IBBP.

pour le profil Baseline IPP à 2,5% pour les images P du profil High IBBP. Ceci est expliqué par l'augmentation des distances temporelles entre les images P, qui entraîne une baisse des corrélations temporelles et par conséquent une chute de l'efficacité des prédicteurs temporels. Nous étudierons plus précisément cette baisse de corrélations temporelles dans la section suivante.

Le tableau 4.5 donne les pourcentages de réduction de débit pour chacune des images B ( $B_1$ ,  $B_2$ ) de la configuration High IBBP en fonction de plusieurs ensembles de prédicteurs de vecteurs mouvement. Ces ensembles contiennent le prédicteur  $mv_{H.264}$  et un second prédicteur temporel. En effet, pour les images B, comme pour les images P, la combinaison d'un prédicteur spatial et d'un prédicteur temporel donne systématiquement les meilleurs résultats. Dans ce tableau le vecteur  $mv_{col_{L1}}$  est le vecteur collocated dans l'image future sans mise à l'échelle.  $mv_{Scol_{L0}}$  est le vecteur collocated de l'image P passée mis à l'échelle et donné par l'équation (4.7) et l'équation (4.8) en fonction de la direction du vecteur à prédire. Le vecteur  $mv_{Scol_{L1}}$  est quant à lui le vecteur collocated mis à l'échelle de l'image de référence  $L1$  donné par l'équation (1.8) et l'équation (1.9).  $mv_{Scol_{B-1}}$  est le vecteur collocated de l'image  $B-1$  provenant des équations (4.9) et (4.12) en fonction des directions et des disponibilités de ces vecteurs collocated. Enfin, le prédicteur  $mv_{Bcol}$  pour l'image  $B_1$  a été défini dans la section 4.1.5.

L'intérêt de la mise à l'échelle des prédicteurs est prouvé par la différence de gains entre le prédicteur  $mv_{col_{L1}}$  et le prédicteur  $mv_{Scol_{L1}}$ . De même, les gains obtenus avec la combinaison du prédicteur  $mv_{H.264}$  et du prédicteur  $mv_{Scol_{L1}}$  par rapport à ceux de  $mv_{H.264}$  et  $mv_{Scol_{L0}}$  confirment que le champ de vecteurs des images B a plus de corrélation avec les champs de vecteurs des images de référence futures ( $L1$ ) que passées ( $L0$ ). Ceci était prévisible, car le champ de vecteurs d'une image P future traverse les images B contrairement aux champs de vecteurs des images P passées. Enfin, le meilleur second prédicteur de l'ensemble  $\mathcal{P}$  de la seconde image B ( $B_2$ ) est le vecteur  $mv_{Scol_{B-1}}$  et non le prédicteur  $mv_{Scol_{L1}}$ , car la distance parcourue par les vecteurs  $mv_{col_{B-1L0}}$  et  $mv_{col_{B-1L1}}$  est plus petite que celle parcourue par le vecteur  $mv_{col_{L1}}$ , ce qui permet d'obtenir une plus grande précision de prédiction. Par conséquent, les gains sont plus importants pour  $B_2$  que pour  $B_1$ .

$B_1$		$B_2$	
Second prédicteur	Réduction de débit	Second prédicteur	Réduction de débit
$mv_{col_{L1}}$	-1.7%	$mv_{col_{L1}}$	-0.7%
$mv_{Scol_{L0}}$	1.8%	$mv_{Scol_{L0}}$	2%
$mv_{Scol_{L1}}$	4.7%	$mv_{Scol_{L1}}$	4.8%
$mv_{Bcol}$	4.9%	$mv_{Scol_{B-1}}$	7.4%

TABLE 4.5 – Pourcentage de réduction de débit pour la première et la seconde image B en fonction de différentes combinaisons de prédicteurs pour le codage de vecteurs mouvement et du mode Skip pour neuf séquences CIF pour le profil High IBBP.

Pour les prochains tests, nous utiliserons donc, les prédicteurs  $mv_{H.264}$  et  $mv_{Scol_{R0}}$  pour l'ensemble  $\mathcal{P}$ , et les prédicteurs  $mv_{spaEXT}$  et  $mv_a$  pour l'ensemble  $\mathcal{P}_s$  pour les images P du profil High IBBP,. Pour l'ensemble  $\mathcal{P}$  de l'image  $B_1$ , nous utiliserons les prédicteurs  $mv_{H.264}$  et  $mv_{Bcol}$ , et pour l'image  $B_2$  les prédicteurs  $mv_{H.264}$  et  $mv_{Scol_{B-1}}$ .

Pour la configuration incluant les slices B hiérarchique du profil High, des expériences similaires ont été menées lors d'un stage de fin d'études. Les résultats, reportés dans [Eno07] ont montré que la compétition du médian de la norme et le prédicteur collocated constitue l'ensemble le plus efficace pour le codage de ce type d'ordonnancement.

### 4.3.2 Réduction de débit et analyses

Dans cette section, nous présenterons les résultats de la méthode pour quatre configurations : Baseline IPP, High IPP, High IBBP et High BH. Pour les configurations proposées, nous avons utilisé les recommandations VCEG [TSW05], excepté l'option "RD Picture Decision". En effet, cette option provoque des bugs dans les statistiques du JM KTA. Par conséquent, l'analyse de l'impact de notre méthode sur le codage ne peut pas être effectuée lorsque cette option est activée. Le fait de ne pas utiliser cette option modifie légèrement les pourcentages de gains sur le débit total. Nous verrons dans la section 4.3.3 que les gains de notre méthode sont en moyenne supérieurs de 0,1%, pour les séquences VCEG lorsque cette option est activée. Les configurations utilisées sont détaillées dans l'annexe A. Les expériences présentées ci-dessous ont été menées dans le JM KTA 1.9r1.

#### 4.3.2.1 Configurations IPP

Le tableau 4.6 donne le pourcentage de réduction de débit pour chacune des séquences testées pour le profil Baseline IPP et le profil High IPP. Les pourcentages sont calculés pour trois ensembles de QP représentant les hauts, moyens et bas débits. Les gains moyens pour ces trois débits sont plus élevés pour le profil Baseline que pour le profil High pour les trois ensembles de QP. En moyenne, le profil Baseline obtient 6,1% pour l'ensemble des QP recommandés par le groupe VCEG et 11% à bas

débit. Le profil High obtient pour cette configuration un gain de 5,3% pour les moyens débits et de 9,5% pour les bas débits. Ces gains plus élevés sont expliqués par la forte proportion d'information de mouvement à ces débits. Le meilleur résultat est atteint à bas débit du profil High pour la séquence *Jets* 720p, avec une réduction de débit de 29,7% en moyenne pour les quatre QP. Le plus mauvais résultat est obtenu à bas débit pour la séquence *Container* QCIF avec 1,1% de perte. Ce résultat est expliqué par la faible quantité de mouvement contenu dans cette séquence. Cette remarque est confirmée par les faibles pourcentages de gains des séquences *Silent* et *Paris* qui contiennent, elles aussi, peu de mouvement. En effet, pour ce type de séquences, la proportion de sélection du mode Skip est déjà élevée. Par conséquent, l'augmentation de l'occurrence de ce mode, engendrée par notre méthode, est faible. A l'opposé, les séquences avec des mouvements rapides et complexes, comme la séquence *Foreman* et plus généralement les séquences avec un fond en mouvement, les gains sont supérieurs.

Le gain apporté indépendamment par chacune des deux méthodes n'est pas mesurable précisément puisque ces méthodes modifient les choix de codage. Cependant, dans le tableau 4.3 nous avons donné le pourcentage de réduction de débit lorsque l'on active uniquement la compétition des vecteurs mouvement, puis uniquement celle des vecteurs du mode Skip. Ces pourcentages sont respectivement de 3,5% et de 2,6% pour l'ensemble des séquences testées. La combinaison des deux compétitions est plus efficace et obtient 5,9% de gains dans ce test.

Le tableau 4.7 donne les pourcentages de réduction de débit d'un ensemble de prédicteurs, pour le mode Skip, composé du médian de la norme et du vecteur collocated. Seuls les résultats plus élevés que ceux présentés dans le tableau 4.6 sont reportés. Les séquences à fond fixe obtiennent de meilleurs résultats pour cette configuration. On note que les pertes obtenues dans la configuration du tableau 4.6 sont comblées avec 3% de gain pour la séquence *Container* QCIF à bas débit. On remarque que cet ensemble a un gain record de 36,1% pour la séquence *Station2* 1080p 25Hz. Si l'on avait choisi, séquence par séquence, le meilleur ensemble de prédicteurs pour le Skip, le gain moyen à bas débit serait de 11,8% au lieu de 11%. Dans le chapitre suivant, nous proposons une adaptation automatique d'ensembles de prédicteurs pour ce mode.

Les pourcentages de gains augmentent avec la résolution passant de 0,6% de réduction moyenne en QCIF à 22,1% en 1080p (cf. tableau 4.6). Cependant, il n'existe pas de corrélation entre le pourcentage de gain et la résolution des séquences. En effet, l'accroissement de gains est lié au nombre d'images par seconde comme nous pouvons le constater dans tableau 4.8. Ce tableau donne les résultats du codage de la séquence *Soccer* pour trois résolutions en fonction de plusieurs nombres d'images par seconde (fréquences). Pour toutes les résolutions, on remarque que le gain augmente lorsque la distance entre les images diminue. Par contre, pour une même fréquence, le gain n'augmente pas forcément en fonction de la résolution. Ceci est expliqué par la distance temporelle entre les images. Lorsque cette distance augmente, les corrélations temporelles des champs de vecteurs mouvement sont moins élevées, ce qui engendre une chute de l'efficacité du prédicteur collocated pour la prédiction des vecteurs mouvement.

Le tableau 4.9 donne le pourcentage de réduction de débit de l'information des vecteurs mouvement avec la méthode proposée pour le profil Baseline IPP. En moyenne, l'information de mouvement a été réduite de 16,5%. Cette réduction est supérieure à bas débit, atteignant 42,2% pour le QP 42.

		Baseline			High		
		17-32	22-37	27-42	17-32	22-37	27-42
Séquence	Fréquence	QCIF					
Container	15Hz	0,1	-0,3	-1,1	0,9	0,4	-0,2
Foreman	15Hz	0,4	1,6	3,8	0,6	1,4	3,0
Silent	15Hz	0,2	-0,2	-1,0	0,3	0,1	-0,7
		CIF					
Foreman	30Hz	1,4	3,7	8,2	1,4	3,1	6,8
Mobile	30Hz	1,6	2,9	5,3	2,3	3,8	5,9
Paris	15Hz	0,4	0,9	1,5	0,4	0,4	0,5
Tempete	30Hz	1,3	2,3	4,5	1,8	2,8	4,5
Carphone	30Hz	0,5	1,8	4,7	0,4	1,2	3,5
Crew	30Hz	1,1	2,4	4,7	1,0	2,1	3,8
Ice	30Hz	3,3	4,6	6,1	2,9	4,0	5,1
Schumacher	30Hz	1,5	2,7	4,9	1,3	2,0	3,3
		4CIF					
City	30Hz	1,3	3,4	8,4	1,3	2,6	6,4
Ice	30Hz	2,0	3,8	6,2	1,0	2,1	3,9
Soccer	60Hz	1,9	4,4	9,5	1,3	3,0	6,8
		720p					
Bigships	60Hz	1,9	5,7	15,0	1,9	5,3	13,5
Shuttlestart	60Hz	2,0	4,1	6,5	1,2	2,7	4,1
City_corr	60Hz	3,0	6,1	13,2	2,9	5,4	11,8
Crew	60Hz	1,5	4,2	7,6	1,2	3,0	6,3
Night	60Hz	0,5	0,9	2,1	0,5	0,8	1,2
Raven	60Hz	4,2	11,5	23,2	3,0	8,9	19,2
Stockholm	60Hz	3,8	7,8	16,0	3,4	7,8	15,6
Cyclists	60Hz	5,5	11,3	21,0	4,7	10,0	18,7
Jets	60Hz	12,7	22,9	29,5	13,1	22,8	29,7
Panslow	60Hz	1,9	7,3	15,3	1,8	6,5	13,7
Sailormen	60Hz	2,3	5,5	12,0	1,7	4,5	10,9
Spincalendar	60Hz	4,0	7,8	13,6	4,8	8,8	14,1
		1080p					
Blue sky	25Hz	5,8	14,1	18,3	5,8	7,5	10,5
Intotree	50Hz	2,7	9,4	23,0	2,0	8,3	20,8
Station2	25Hz	10,0	17,0	26,2	9,2	15,1	23,8
Toys and calendar	60Hz	5,9	13,4	20,8	5,1	11,8	18,9
Moyenne		3,0	6,1	11,0	2,6	5,3	9,5

TABLE 4.6 – Pourcentage de réduction de débit de chaque séquence pour le profil Baseline IPP et le profil High IPP, pour trois ensembles de QP.

Séquence	Fréquence	Résolution	Baseline		
			17-32	22-37	27-42
Container	15Hz	<b>QCIF</b>	1,2	2,3	3,0
Silent	15Hz	<b>QCIF</b>	0,4	0,6	1,2
Mobile	30Hz	<b>CIF</b>	1,8	3,3	5,9
Tempete	30Hz	<b>CIF</b>	1,4	2,5	4,8
Ice	30Hz	<b>CIF</b>	3,6	5,3	7,9
Ice	30Hz	<b>4CIF</b>	2,4	4,7	8,0
Night	60Hz	<b>720p</b>	0,5	1,1	2,7
Spincalendar	60Hz	<b>720p</b>	4,7	10,4	16,4
Station2	25Hz	<b>1080p</b>	14,4	25,5	36,1

TABLE 4.7 – Pourcentage de réduction de débit de chaque séquence pour la configuration Baseline IPP avec un ensemble  $\mathcal{P}_s$  pour le mode Skip composé des prédicteurs  $mv_{Scol_{R_0}}$  et  $mv_{H.264}$  pour trois ensembles de QP.

	<b>QCIF</b>	<b>CIF</b>	<b>4CIF</b>
<b>7,5 Hz</b>	1,5	0,9	2,5
<b>15 Hz</b>	3,4	1,7	3,9
<b>30 Hz</b>		3,9	5,4
<b>60 Hz</b>			9,5

TABLE 4.8 – Pourcentage de réduction de débit de la séquence *Soccer* pour les résolutions QCIF, CIF et 4CIF codée en fonction de différents nombres d'images par seconde. La configuration Baseline IPP avec l'ensemble des QP 27-42 est utilisée.

QP	Réduction de débit
17	2,5
22	4,4
27	8,4
32	15,0
37	26,6
42	42,2
<b>Moyenne</b>	<b>16,5</b>

TABLE 4.9 – Pourcentage de réduction du débit de l'information de mouvement de la référence par rapport à celui de la méthode proposée incluant les indices de prédicteurs, pour la configuration Baseline IPP, en fonction des QP, pour l'ensemble des séquences et des six QP utilisés.

Notons que la compétition entre les prédicteurs pour le mode Skip et pour les vecteurs mouvement est intégrée dans la boucle de codage. Par conséquent, la sélection spatiale et quantitative des modes est différente de celle de la référence. Dans le tableau 4.10, nous proposons de comparer la proportion de l'information des vecteurs mouvement pour la méthode proposée et pour la référence. L'écart de proportion est plus important à bas débit, car l'information de mouvement est plus élevée, mais l'écart moyen est seulement de 3%. De plus, à haut débit, cette différence est très faible (0,1% pour le QP 17). Ce maigre écart de proportion est expliqué par l'optimisation du critère débit-distorsion. En effet, nous n'avons pas modifié le coefficient de Lagrange pour notre méthode, donc, lors de l'estimation de mouvement, le compromis entre la distorsion et le débit du vecteur est le même que celui de la référence. Le débit des vecteurs et la distorsion auront alors tous les deux un coût moins élevé (et pas uniquement le débit des vecteurs). La sélection d'un  $\lambda_{Mode}$  et un  $\lambda_{Motion}$  optimaux pour la méthode est une perspective envisagée par notre équipe. Enfin, nous noterons que le débit de l'indice des prédicteurs des vecteurs mouvement représente en moyenne 2,6% et atteint 3,6% pour les QP 32 et 37.

La figure 4.7 présente le pourcentage d'augmentation de sélections du mode Skip pour les séquences CIF en fonction du QP. En moyenne, pour l'ensemble des séquences, l'augmentation de modes Skip est de 6,4%. Cette augmentation est corrélée avec le débit, puisque pour le QP 17 l'augmentation moyenne est de 2,8%, contre 8,4% pour le QP 42. Dans la figure 4.7, on remarque que cette augmentation dépend des séquences. On observe une légère baisse pour la séquence *Paris* et *Ice* pour le QP 17. Pour ces deux séquences à fond fixe, le pourcentage de Skip est déjà très élevé. L'augmentation est donc naturellement plus faible comparée aux séquences comportant du mouvement. De plus, l'ensemble  $\mathcal{P}_s$  spatial n'est pas le plus efficace pour les séquences à fond fixe. Pour un ensemble spatio-temporel  $\mathcal{P}_s$ , l'augmentation est toujours positive et atteint 10% pour la séquence *Ice*.

La proportion de l'indice du mode Skip représente en moyenne 3,7% du débit total. Ce débit évolue en fonction du QP atteignant 11% pour le QP 42, ce qui est faible comparé au 87,5% de

QP	H.264	Méthode proposée		
	Résiduel MV	Résiduel MV	Prédicteur	Total
17	7,1	6,0	1,0	7,0
22	12,2	10,1	1,7	11,9
27	19,9	15,9	2,8	18,7
32	26,9	20,6	3,6	24,2
37	30,9	22,0	3,6	25,6
42	31,5	20,0	3,0	23,0
<b>Moyenne</b>	<b>21,4</b>	<b>15,8</b>	<b>2,6</b>	<b>18,4</b>

TABLE 4.10 – Proportion du débit par rapport au débit total des résiduels de vecteurs mouvement pour la référence et du débit des résiduels et des indices de prédicteurs de la méthode proposée pour la configuration Baseline IPP, en fonction des QP, pour l’ensemble des séquences et des six QP utilisés.

sélection de mode Skip pour ce QP. La combinaison d’une part des deux prédicteurs exploitant la même redondance et d’autre part de la nature du mode Skip, augmente la probabilité que les deux prédicteurs aient la même valeur. Ce cas représente en moyenne 63,8% des sélections du mode Skip et atteint 74% pour le QP 42.

L’efficacité de la méthode de compétition des prédicteurs de vecteurs mouvement réside en l’exploitation des redondances spatiales et des redondances temporelles. Cette efficacité est confirmée par l’analyse de la sélection des prédicteurs  $mv_{Scol_{R_0}}$  et  $mv_{H.264}$ . Le prédicteur temporel est le plus sélectionné en moyenne, avec 43,7% de sélection contre 41,9% pour le prédicteur de la norme H.264/AVC (cf. tableau 4.12). Ces pourcentages excluent le cas d’égalité des deux prédicteurs, qui s’élève à 14,4% en moyenne. La séquence *Foreman* QCIF 15Hz QP 17 obtient le minimum de sélection du prédicteur temporel avec 23% de sélection. Ce pourcentage est élevé comparé au minimum de sélection du médian qui est de 12,3%. Le maximum de sélection pour le prédicteur temporel est de 76,6% pour la séquence *Station2* 1080p 25Hz QP 42. La sélection de chacun des prédicteurs  $mv_{Scol_{R_0}}$  et  $mv_{H.264}$  est corrélée avec l’augmentation du QP, comme on peut le voir dans la figure 4.8. Le médian de la norme est plus souvent sélectionné que le prédicteur collocated à haut débit, mais à bas débit la sélection de ce dernier maximise celle du médian avec 51,5% de sélection. Pour ce QP, le prédicteur  $mv_{H.264}$  a été utilisé pour moins d’un tiers des vecteurs (31,6% de sélection).

Le choix proportionnellement plus élevé de prédicteurs temporels à bas débit est lié à la sélection des modes. Comme nous l’avons remarqué dans le chapitre 3, la proportion des blocs de grande taille est plus élevée à bas débit, ce qui implique une distance spatiale plus grande entre les vecteurs mouvement. Par conséquent, la prédiction spatiale des vecteurs engendre une entropie plus forte à bas débit. La distance temporelle entre le bloc courant et le bloc collocated n’est pas modifiée par l’augmentation de la proportion des blocs de grande taille. Le tableau 4.11 donne la proportion de

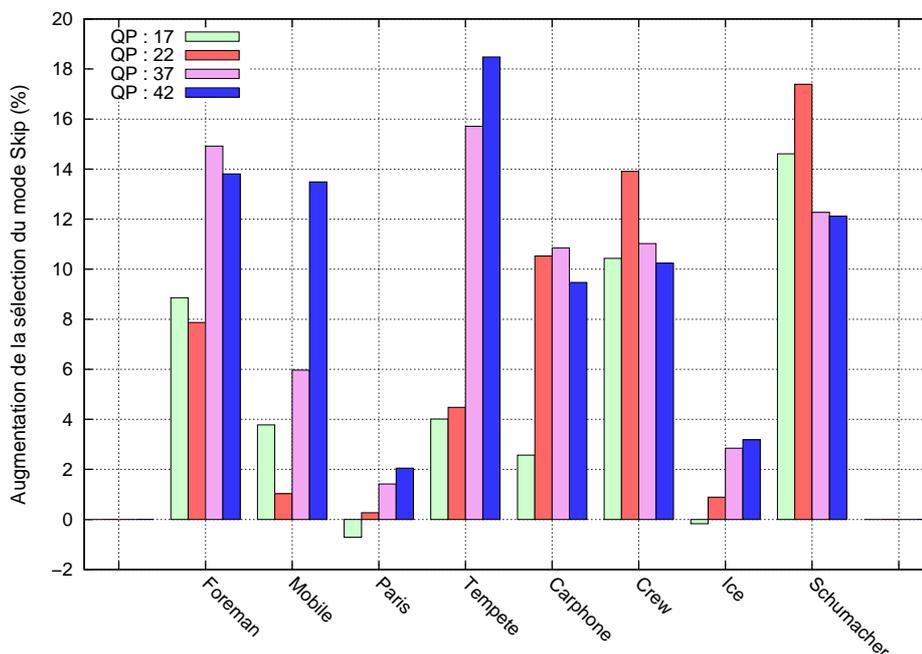


FIGURE 4.7 – Augmentation de sélections du mode Skip avec la méthode proposée, pour les QP 17, 22, 37 et 42 pour huit séquences CIF codées avec le profil Baseline IPP.

sélection des prédicteurs  $mv_{Scol_{R_0}}$  et  $mv_{H.264}$  et de leur égalité, en fonction des tailles de partitions du codage Inter. On remarque que le prédicteur collocated est en moyenne plus souvent sélectionné que le médian pour l'Inter  $16 \times 16$ . En revanche, pour les autres tailles de bloc, le prédicteur médian a un pourcentage de sélection plus élevé. Notons que pour le QP 42, le prédicteur collocated est le plus sélectionné pour toutes les tailles de blocs, excepté pour les blocs  $4 \times 4$ .

Le tableau 4.12 présente le pourcentage de sélection des deux prédicteurs utilisés pour la compétition de prédictions des vecteurs mouvement en fonction de l'indice des images de référence. On remarque que la sélection de  $mv_{Scol_{R_0}}$  est plus élevée que le médian à partir de la seconde image de référence. Cette sélection plus importante provient de l'efficacité de la mise à l'échelle du vecteur collocated comparé au médian de la norme qui n'utilise pas de mise à l'échelle. Avec cette efficacité, le nombre de prédicteurs de bloc sélectionnés dans la troisième image de référence a proportionnellement augmenté de 39%.

Dans les sections suivantes, nous étudierons l'impact de notre méthode lorsque les slices B ou B hiérarchiques sont activés.

#### 4.3.2.2 Configuration High IBBP

Le tableau 4.13 présente le pourcentage de réduction de débit pour la configuration High IBBP et ceux des images P et B de cette configuration. A bas débit, le gain moyen est de 6%. Ce gain est moins élevé que celui du profil High IPP. Cependant, cette configuration a une efficacité de codage

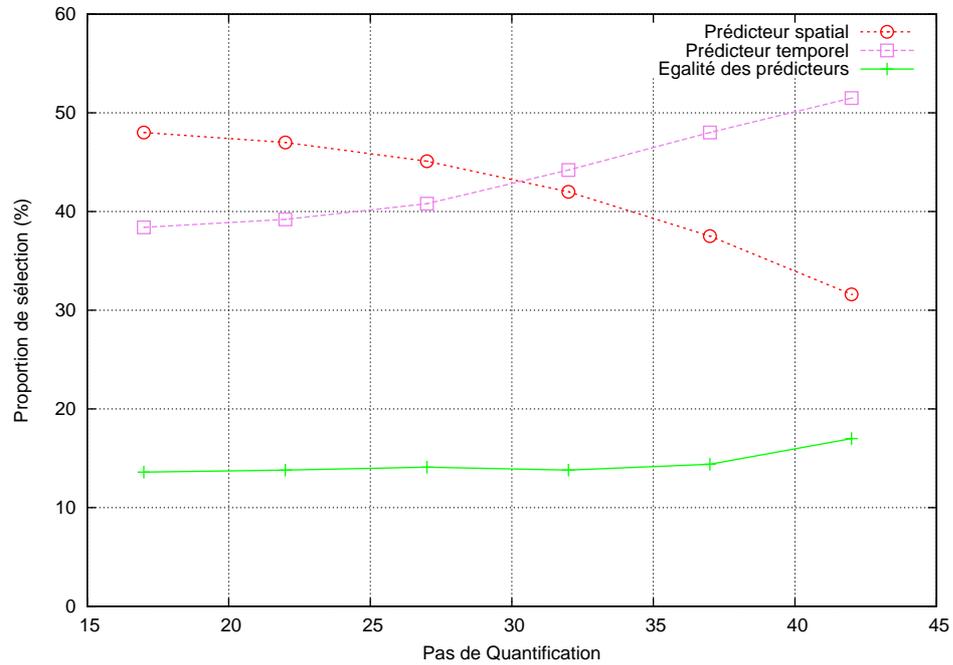


FIGURE 4.8 – Evolution de la proportion de sélection des prédicteurs  $mv_{Scol_{R_0}}$  et  $mv_{H.264}$  et de leur égalité de l'ensemble  $\mathcal{P}$  en fonction des QP de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences.

Tailles de bloc	Prédicteur spatial $mv_{H.264}$	Prédicteur temporel $mv_{Scol_{R_0}}$	Egalité des prédicteurs (même valeur)
16×16	39,7	44,2	16,1
16×8	43,3	42,5	14,2
8×16	45,5	41,3	13,2
8×8	42,9	40,1	17,0
8×4	45,0	39,9	15,1
4×8	44,6	39,3	16,1
4×4	41,1	37,3	21,6

TABLE 4.11 – Proportion de sélection des prédicteurs  $mv_{Scol_{R_0}}$  et  $mv_{H.264}$  et de leur égalité de l'ensemble  $\mathcal{P}$  en fonction de la taille des blocs Inter (sauf Skip) de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences et des six QP utilisés.

Images de référence	Prédicteur spatial $mv_{H.264}$	Prédicteur temporel $mv_{Scol_{R_0}}$	Egalité des prédicteurs (même valeur)
$Ref_0$	43,1	40,7	16,2
$Ref_1$	42,1	46,9	11,0
$Ref_2$	39,0	50,5	10,5
$Ref_3$	33,3	56,5	10,2
<b>Moyenne</b>	<b>41,9</b>	<b>43,7</b>	<b>14,4</b>

TABLE 4.12 – Proportion de sélection des prédicteurs  $mv_{Scol_{R_0}}$  et  $mv_{H.264}$  de l'ensemble  $\mathcal{P}$  et de leur égalité en fonction des images de référence de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences et des six QP utilisés.

beaucoup plus grande. Le gain maximum est obtenu sur la séquence *Jets* 720p 60Hz avec 22,2% de gain. Les remarques de la section précédente sur la relation des gains avec les QP utilisés, ou les gains avec le nombre d'images par seconde, restent vraies.

Les gains sur chacun des types d'image (P et B) sont assez particuliers puisque, pour les moyens et hauts débits, les gains sur les images B sont plus élevés que ceux obtenus pour les images P. Cependant, à bas débit, cette tendance s'inverse et le gain des images P devient légèrement supérieur à celui des images B. Ceci est expliqué par la proportion de l'information de mouvement dans les images B. En effet, à bas débit, cette information est proportionnellement inférieure à celle des indices de modes et des autres informations de codage, comme on peut le voir dans la figure 3.6 du chapitre 3. Par conséquent, l'efficacité de la méthode est plus élevée pour les images P, où l'information de mouvement est la plus représentée dans le débit total à bas débit. (cf. figure 3.5 du chapitre 3). Cette réduction de débit des images P reste cependant inférieure à celle obtenue sur les images P de la configuration High IPP. En effet, l'éloignement des images P engendre une baisse des corrélations temporelles entre les champs de vecteurs mouvement. Par conséquent, l'efficacité de la méthode est réduite. De plus, la proportion de l'information de mouvement dans le débit total diminue avec l'éloignement des images.

#### 4.3.2.3 Configuration High BH

Le tableau 4.14 donne le pourcentage de réduction de la configuration du profil High lorsque sept images B hiérarchiques sont codées entre deux images P. Le même nombre d'images que les autres configurations est utilisé. L'efficacité de la méthode pour cette configuration est inférieure à celle de la configuration avec deux images B car, d'une part, cette configuration est très efficace et, d'autre part, l'impact de la méthode sur les images P est faible alors que ces images ont une forte proportion dans le débit total. En effet, l'éloignement des images limite la sélection du mode Skip et les redondances temporelles entre les vecteurs de deux images P sont faibles. Pour cette

		High IBBP			Images P			Images B		
		17-32	22-37	27-42	17-32	22-37	27-42	17-32	22-37	27-42
Séquence	Fréquence	<b>QCIF</b>								
Container	15Hz	1,8	2,6	2,3	2,0	3,1	2,6	1,2	1,8	2,3
Foreman	15Hz	0,2	0,8	1,9	0,0	0,3	1,0	0,4	1,2	2,8
Silent	15Hz	0,4	0,4	0,1	0,0	-0,2	-0,5	0,9	1,2	0,9
		<b>CIF</b>								
Foreman	30Hz	0,7	2,1	4,1	0,1	1,3	3,0	1,7	3,2	4,7
Mobile	30Hz	1,7	3,4	5,1	1,6	2,8	4,4	2,2	4,6	6,3
Paris	15Hz	0,4	0,5	0,5	0,2	0,1	0,0	0,6	0,8	0,8
Tempete	30Hz	2,2	3,8	5,0	2,2	3,3	4,9	2,4	4,3	4,0
Carphone	30Hz	0,0	0,5	1,7	0,0	0,4	1,7	-0,3	0,3	1,9
Crew	30Hz	0,8	1,6	2,7	0,0	0,2	0,9	2,0	3,7	4,7
Ice	30Hz	2,1	3,1	4,6	0,8	1,1	1,6	4,0	6,2	9,2
Schumacher	30Hz	0,4	0,9	1,5	0,0	0,0	0,3	0,7	1,7	2,9
		<b>4CIF</b>								
City	30Hz	0,3	0,7	3,0	0,1	0,6	2,5	0,0	-0,1	2,9
Ice	30Hz	1,2	2,2	3,4	0,3	0,2	0,4	2,4	4,6	7,2
Soccer	60Hz	0,8	1,8	4,3	0,3	1,0	3,1	1,2	2,9	5,5
		<b>720p</b>								
Bigships	60Hz	2,8	5,6	9,4	2,4	5,6	14,4	5,9	7,3	-2,5
Shuttlestart	60Hz	0,4	1,4	1,9	-0,1	0,9	2,5	1,3	1,2	-0,9
City_corr	60Hz	1,3	3,5	7,4	0,7	1,8	5,4	2,4	7,8	12,3
Crew	60Hz	0,5	1,5	2,6	-0,1	0,1	1,3	1,5	4,2	4,4
Night	60Hz	0,2	0,3	-0,1	-0,5	-0,7	-0,3	1,0	1,7	-0,2
Raven	60Hz	0,4	2,8	7,6	0,2	2,5	8,6	0,2	1,9	4,8
Stockholm	60Hz	2,1	7,2	10,6	2,6	5,6	10,1	-2,8	10,9	11,7
Cyclists	60Hz	2,0	5,5	11,5	1,1	3,5	8,6	3,7	9,6	16,8
Jets	60Hz	13,1	20,5	<b>22,2</b>	12,2	23,9	<b>31,1</b>	16,5	18,5	11,5
Panslow	60Hz	1,9	6,1	9,4	2,0	6,4	14,0	-1,7	7,9	7,1
Sailormen	60Hz	0,7	1,6	3,3	0,4	0,9	3,3	1,7	3,0	1,6
Spincalendar	60Hz	3,9	9,7	11,6	4,9	8,9	11,8	1,8	12,4	13,7
		<b>1080p</b>								
Blue sky	25Hz	3,2	4,8	7,4	1,5	2,8	5,4	5,9	7,1	8,7
Intotree	50Hz	1,6	4,9	12,4	1,2	4,4	14,6	1,7	6,6	10,9
Station2	25Hz	5,3	9,4	13,7	1,9	6,5	13,6	8,8	12,7	14,2
Toys and c.	60Hz	2,2	5,7	8,1	1,7	5,3	10,6	3,0	5,4	5,3
<b>Moyenne</b>		<b>1,8</b>	<b>3,8</b>	<b>6,0</b>	<b>1,3</b>	<b>3,1</b>	<b>6,0</b>	<b>2,3</b>	<b>5,2</b>	<b>5,9</b>

TABLE 4.13 – Pourcentage de réduction de débit de chaque séquence et de leurs images P et B du profil High IBBP, pour trois ensembles de QP.

	High BH	
	22-37	27-42
<b>QCIF</b>	0,6	0,7
<b>CIF</b>	1,1	1,9
<b>4CIF</b>	0,8	2,3
<b>720p</b>	4,0	6,9
<b>1080p</b>	1,1	5,0
<b>Moyenne</b>	<b>2,2</b>	<b>4,2</b>

TABLE 4.14 – Pourcentage de réduction de débit moyen de la configuration du profil High composée de sept images B hiérarchiques, en fonction de chaque résolution, pour deux ensembles de QP.

configuration, le prédicteur collocated est sélectionné dans 20% des cas en moyenne contre 44% pour les images P du profil Baseline IPP. Remarquons que la méthode reste efficace pour les images B avec un gain moyen de 4,5% pour les moyens débits et de 6,5% pour les bas débits.

La méthode MVComp a été proposée en juillet 2006 au groupe VCEG, qui l'a jugée suffisamment efficace, par rapport à la faible complexité engendrée au codeur et au décodeur, pour être intégrée dans le JM KTA. On notera que l'utilisation des vecteurs de l'image  $B - 1$  pour la configuration High IBBP n'a pas été retenue. Les vecteurs temporels utilisés pour les images B sont donc ceux de l'image  $L1$ . Dans la section suivante, nous proposons d'analyser les gains de notre méthode par rapport aux autres méthodes intégrées.

### 4.3.3 Résultats MVComp dans la JM KTA

Les outils présents à ce jour dans le JM KTA ont été décrits dans la section 1.4 du chapitre 1.3.4. Dans le tableau 4.15, sont reportés les gains apportés par chacun de ces outils pour l'ensemble des séquences 720p VCEG. Les tests ont été effectués dans le JM KTA 2.0 avec la configuration High IPP et l'ensemble des options et QP recommandés par le groupe VCEG (cf. annexe A). Seul le mode Intra MDDT, n'a pas été testé car son impact sur cette configuration est naturellement faible. Certains des outils, testés indépendamment de MVComp, engendrent des pertes pour cette configuration et cette résolution. La méthode IBDI maximise les résultats, avec 9,3% de réduction de débit par rapport à la norme H.264/AVC.

Lorsque la méthode MVComp est activée avec chacun des autres outils, les gains par rapport à la norme H.264/AVC sont toujours positifs. Le gain maximum obtenu est de 14,1% avec l'outil AIF.

Les pourcentages de réduction de débit apportée par MVComp par rapport à chaque outil sont aussi reportés dans ce tableau 4.15. Les gains sont supérieurs ou égaux à ceux de MVComp par rapport à la norme H.264/AVC. Cette augmentation atteint 6% de réduction de débit par rapport à la méthode AIF. L'efficacité plus élevée de notre méthode avec l'outil du  $\frac{1}{8}$ -pel est facilement explicable. En effet, le  $\frac{1}{8}$ -pel augmente la proportion de l'information de mouvement car les vecteurs

Référence	H.264	H.264	H.264+Méthode
Configuration	Outil	Outil+MVComp	Outil+MVComp
$\frac{1}{8}$ -pel	-1	4,6	<b>5,6</b>
<b>AIF</b>	8,6	14,1	<b>6</b>
<b>s-AIF</b>	6,8	11,7	<b>5,1</b>
<b>d-AIF</b>	6,2	11,2	<b>5,3</b>
<b>APEC</b>	0,5	4,4	<b>3,9</b>
<b>AQMS</b>	2,9	5,4	<b>2,6</b>
<b>AQP</b>	-0,3	3,3	<b>3,7</b>
<b>RDOQ</b>	3,2	6,8	<b>3,7</b>
<b>IBDI</b>	9,3	14	<b>4,9</b>
<b>MVComp</b>	<b>2,6</b>		

TABLE 4.15 – Pourcentage de réduction de débit de chacune des méthodes intégrées dans le JM KTA, et de leur combinaison avec MVComp, pour l'ensemble des séquences 720p et les quatre QP VCEG dans la configuration High IPP.

mouvement sont plus précis. MVComp est donc naturellement plus efficace lorsque cet outil est activé. Pour les autres outils, l'augmentation d'efficacité est expliquée par le même phénomène. En effet, tous les outils présents dans le JM KTA (excepté MVComp) sont des outils d'amélioration du codage de la texture. Ils réduisent donc la proportion de l'information de texture dans le débit total. Par conséquent, la proportion de l'information de mouvement est plus élevée que pour la norme H.264/AVC. La méthode MVComp est donc forcément plus efficace lorsque ces outils sont activés.

## 4.4 Conclusion du chapitre

Dans ce chapitre, nous avons proposé deux méthodes de codage par compétition, l'une pour la prédiction des vecteurs mouvement et l'autre pour le mode Skip. La prédiction de chaque vecteur mouvement est sélectionnée à l'aide du critère débit-distorsion, qui prend en compte à la fois le débit de la prédiction et le débit de l'indice du prédicteur. La sélection du meilleur prédicteur pour le mode Skip est aussi basée sur le critère RD et considère le débit de l'indice du prédicteur et la distorsion engendrée par ce prédicteur. Ces deux compétitions permettent d'exploiter les redondances spatiales et temporelles des champs de vecteurs mouvement. De plus, des prédicteurs adaptés aux options des images de référence multiple et des images B ont été proposés. Ces adaptations sont des mises à l'échelle en fonction des distances temporelles parcourues par le vecteur courant et le prédicteur.

A ce jour, cette méthode a donné lieu à un stage de fin d'études dédié à l'implémentation des prédicteurs des images B hiérarchiques ainsi qu'à l'intégration de la méthode pour H.264/MVC

[Eno07]. De plus, deux stages de master recherche ont été consacrés au développement des deux compétitions dans un logiciel propriétaire [EB08].

La faible complexité et l'efficacité de codage engendrées par la méthode ont permis d'atteindre notre objectif de participer à l'élaboration de l'éventuelle future norme de l'ITU. En effet, le groupe VCEG a jugé les gains suffisamment importants pour que la méthode MVComp soit intégrée dans le JM KTA. Par ailleurs, les performances de la méthode sont améliorées lorsque la méthode est comparée aux outils déjà intégrés dans le JM KTA. Ce surcroît de performances augmente les possibilités que la méthode MVComp appartienne à une future norme de codage vidéo.

Dans le chapitre suivant nous allons proposer une adaptation de l'ensemble des prédicteurs du mode Skip défini dans ce chapitre, basé sur un critère a priori pour augmenter les performances de cette méthode.

## Chapitre 5

# Compétition a priori pour le mode Skip

*Dans le chapitre précédent, nous avons proposé une méthode basée sur une compétition exhaustive de prédicteurs pour le codage de l'information de mouvement. Les résultats expérimentaux ont montré que la compétition de deux prédicteurs spatiaux, pour le mode Skip, était plus efficace en moyenne sur les séquences utilisées. Cependant, pour certaines séquences, un ensemble de prédicteurs spatio-temporels s'est révélé plus performant. Le mode Skip est un mode particulier du codage Inter. Pour ce mode, aucune information n'est transmise au décodeur, excepté l'indice de mode. Une augmentation du nombre de modes Skip devrait augmenter l'efficacité de codage.*

*Dans ce but, nous proposons une adaptation d'ensembles de prédicteurs pour le mode Skip orientée contenu, afin, d'une part, de réduire l'information liée à la compétition exhaustive des prédicteurs et, d'autre part, d'offrir des prédicteurs adaptés aux contenus des séquences.*

### 5.1 Description de la méthode

Les résultats expérimentaux de la compétition de prédicteurs pour le mode Skip montrent que l'efficacité des prédicteurs mis en compétition varie en fonction du contenu de la séquence (cf. tableau 4.6 et tableau 4.7 du chapitre 4). Une sélection du meilleur ensemble de prédicteurs pour le mode Skip devrait améliorer les performances de la méthode MVComp. La solution optimale pour sélectionner le meilleur ensemble de prédicteurs est le test de chacun des ensembles, image par image. Cependant, cette solution présente une complexité de calcul très élevée au niveau de l'encodeur, puisqu'elle nécessite le ré-encodage de l'image courante pour chacun des ensembles de prédicteurs testés. Dans ce chapitre, nous proposons une alternative moins complexe, basée sur nos observations faites durant l'étude du meilleur ensemble de prédicteurs. Dans une première section, nous décrivons une adaptation de l'ensemble de prédicteurs pour le mode Skip. Nous noterons cette adaptation ADAPT 1 SET.

---

1. Publications de l'auteur associées : [Conf4], [ConfFR2] et [Expo1].

La sélection du mode Skip signifie qu'il est plus intéressant, au sens du critère débit-distorsion, de ne transmettre aucune information plutôt qu'un résiduel de bloc et un résiduel de vecteur mouvement. Ce mode a été créé pour les macroblocs ayant une forte probabilité d'obtenir un résiduel de vecteur et un résiduel de texture (transformé, quantifié) égaux à zéro. Les macroblocs concernés par ce mode ont un mouvement proche de leurs voisins précédemment encodés, et ils appartiennent à des objets qui ne subissent aucune déformation d'une image à l'autre. Dans le chapitre 3, nous avons remarqué que ce mode est souvent sélectionné pour les zones avec peu ou pas de mouvement. La sélection de ce mode dépend naturellement du paramètre de Lagrange utilisé dans le critère RD qui dépend, rappelons le, du pas de quantification ou du débit cible. Dans ce chapitre, nous proposons de segmenter grossièrement l'image en deux zones : une zone statique et une zone en mouvement. Chacune de ces zones a son propre ensemble de prédicteurs pour le mode Skip qui évolue indépendamment. Nous noterons cette méthode ADAPT 2 SETS.

### 5.1.1 Adaptation d'un ensemble de prédicteurs

Dans la méthode de compétition des vecteurs mouvement pour le mode Skip, pour chaque prédicteur  $i$  de l'ensemble  $\mathcal{P}_s$  composé de  $N$  prédicteurs, le coût débit-distorsion  $J_{SKIP}^i$  est calculé et comparé aux autres coûts débit-distorsion de toutes les possibilités de codage (Inter et Intra). Les prédicteurs utilisés dans ce chapitre appartiennent à l'ensemble défini dans le chapitre 4. Cependant, nous avons ajouté les vecteurs  $mv_{Sa}, mv_{Sb}, mv_{Sc}$  provenant de la mise à l'échelle des vecteurs  $mv_a, mv_b, mv_c$ . Ces vecteurs sont égaux aux vecteurs  $mv_a, mv_b, mv_c$ , si ces derniers pointent sur la première image de référence. Dans la méthode d'adaptation d'un ensemble de prédicteurs, onze prédicteurs ont été utilisés, ce qui peut paraître élevé. Cependant, ces vecteurs sont souvent égaux entre eux, ce qui limite l'impact sur le coût de l'indice des prédicteurs pour le mode Skip.

L'évolution d'un ensemble de prédicteurs pour le mode Skip se fait slice par slice à l'encodeur. Ainsi, l'ensemble des prédicteurs utilisés est transmis au début de chaque slice. L'impact sur le débit de cette adaptation d'ensemble est faible, car un seul bit pour chaque prédicteur de l'ensemble global est transmis au décodeur, soit 11 bits par slice pour notre étude (un seul slice par image dans les recommandations VCEG). L'ensemble courant dépend des statistiques de l'image précédente. Pour chaque macrobloc, tous les  $J_{SKIP}^j$  des  $M$  prédicteurs de l'ensemble global ( $M = 11$ ), appartenant ou non à l'ensemble utilisé, sont calculés. Notons  $RdCount^j$  le nombre de fois où le prédicteur  $p_j$  remplit la condition :

$$J_{SKIP}^j \leq J_{min} \quad (5.1)$$

où  $J_{min}$  est le coût minimum du critère débit-distorsion obtenu pour tous les modes de codages Inter et Intra. Ce critère est efficace, car il représente le nombre de fois où le prédicteur  $p_j$  est meilleur ou équivalent au codage sélectionné pour le macrobloc. Afin de choisir les prédicteurs qui vont être utilisés dans l'ensemble  $\mathcal{P}_s$  du slice suivant, définissons :

$$MaxRdCount = \max\{RdCount^j\}_{\forall j < M} \quad (5.2)$$

La sélection de l'ensemble  $\mathcal{P}_s$  pour le slice suivant se fait à la fin du slice courant.  $MaxRdCount$

est donc calculé à la fin du codage du slice courant. Au début du processus de mise à jour, aucun prédicteur n'appartient à l'ensemble courant. Un prédicteur  $p_j$  est ajouté dans l'ensemble des prédicteurs si :

$$MaxRdCount \leq \delta \times RdCount^j \quad (5.3)$$

Ceci signifie que le prédicteur  $p_j$  est ajouté si le nombre de fois où il minimise le critère RD du meilleur des codages est supérieur, en fonction de la constante  $\delta$ , au maximum de fois où un prédicteur a minimisé le critère RD de l'ensemble des codages. Notons que  $\delta \in [1; +\infty[$ . Si  $\delta = 1$ , seul le prédicteur qui maximise le nombre de minimisations du critère RD sera utilisé pour le slice suivant. Si  $\delta$  tend vers  $+\infty$ , l'ensemble  $\mathcal{P}_s$  du slice suivant sera composé des  $M$  prédicteurs.

L'adaptation décrite ci-dessus est utilisée dans la méthode ADAPT 1 SET pour chaque ensemble  $\mathcal{P}_s$  de chaque slice. Dans la méthode ADAPT 2 SETS, cette adaptation est aussi utilisée indépendamment pour chacun des deux ensembles  $\mathcal{P}_s$  de chaque slice. La segmentation des deux ensembles est décrite dans la section suivante.

### 5.1.2 Classification des macroblocs

L'analyse de la sélection du mode Skip du chapitre 3 a montré que le mode Skip était largement sélectionné pour les zones comportant peu de mouvement. Afin d'améliorer la méthode de compétition de prédicteurs pour le mode Skip, nous proposons de segmenter l'image en deux types de zones :

- les zones où le mode Skip a une forte probabilité de sélection, pour restreindre le coût de l'indice du prédicteur, en utilisant moins de prédicteurs ou des prédicteurs souvent égaux
- les zones où le mode Skip a une faible probabilité de sélection, pour augmenter la sélection de ce mode très efficace en utilisant plus de prédicteurs ou des prédicteurs qui exploitent des redondances différentes

Chaque macrobloc est donc classé soit dans la classe “Statique” représentant une zone statique avec peu ou pas de mouvement, soit dans une classe “Mouvement” représentant une zone avec du mouvement. La classification des macroblocs doit être calculable au décodeur afin de ne pas introduire une nouvelle information à transmettre. Nous avons donc choisi d'utiliser les deux images de référence précédentes qui sont disponibles au décodeur pour calculer ce critère. Ce critère est le SAD du macrobloc collocated de la première et de la seconde image de référence. Si ce SAD est inférieur à un seuil, le macrobloc courant appartient à la classe “Statique”, sinon à la classe “Mouvement”. Nous avons choisi volontairement cette classification grossière du mouvement au niveau macrobloc. Bien entendu, des méthodes plus fines d'analyse du mouvement existent. Cependant la classification proposée est causale, permettant ainsi de ne pas transmettre d'information supplémentaire.

Chaque classe a donc son propre ensemble de prédicteurs pour le mode Skip :  $\mathcal{P}_s^M$  pour la classe “Mouvement” et  $\mathcal{P}_s^S$  pour la classe “Statique”. Ces deux ensembles  $\mathcal{P}_s^M$  et  $\mathcal{P}_s^S$ , sont transmis slice par slice, soit 22 bits par image avec les recommandations VCEG . Ces deux ensembles évoluent avec le critère défini dans la section précédente. Afin que l'évolution des ensembles  $\mathcal{P}_s^M$  et  $\mathcal{P}_s^S$  soit indépendante, chaque classe a son propre  $MaxRdCount$ , son propre  $RdCount^j$  et son propre  $\delta$ . En

effet, pour que la classe “Statique” contienne peu de prédicteurs, le  $\delta_S$  est proche de 1 et le  $\delta_M$  de la classe “Mouvement” est tel que  $\delta_S < \delta_M$  pour que l’ensemble  $\mathcal{P}_s^M$  contienne un plus grand nombre de prédicteurs que l’ensemble  $\mathcal{P}_s^S$ .

## 5.2 Résultats expérimentaux

Les expériences ont été effectuées avec le profil Baseline IPP dans le logiciel JM KTA 1.1. Nous avons utilisé les conditions de tests recommandées par le groupe VCEG pour ce profil, excepté les pas de quantification que nous avons fixés à 27 32 37 42. En effet, nous ciblons les bas débits, où le nombre de macroblocs codés avec le mode Skip est plus élevé. Les gains en réduction de débit sont calculés avec la métrique de Bjøntegaard [Bjø01a]. Pour les trois configurations, MVComp, ADAPT 1 SET, ADAPT 2 SETS, deux prédicteurs sont utilisés pour la compétition de prédiction des vecteurs mouvement ( $mv_{H.264}$  et  $mv_{Scol_{R_0}}$ ). Pour la méthode MVComp, deux prédicteurs sont mis en compétition pour le mode Skip ( $mv_{spaEXT}$ ,  $mv_a$ ). Cet ensemble donne les meilleurs résultats en moyenne (cf. chapitre 4).

### 5.2.1 Augmentation de l’occurrence du mode Skip

La figure 5.1 représente le pourcentage d’augmentation du nombre de macroblocs codés avec le mode Skip, pour les trois méthodes (MVComp, ADAPT 1 SET, ADAPT 2 SETS) par rapport à la référence H.264/AVC sans compétition. Ce pourcentage d’augmentation du nombre de modes Skip est corrélé avec les types de séquences. Il est plus faible pour les séquences ayant un fond fixe comme *BBC news*, *Modo*, *Ice* et *Silent*, car la proportion de mode Skip est déjà très élevée (près de 80% en moyenne pour les 4 QP pour ces séquences). Pour chaque séquence, les deux méthodes d’adaptation de prédicteurs obtiennent un pourcentage d’augmentation d’occurrences de mode Skip deux fois supérieur à celui de la méthode MVComp. Le pourcentage d’augmentation du nombre de mode Skip, séquence par séquence, de chacune des méthodes ADAPT 1 SET, ADAPT 2 SETS est proche. Toutefois la méthode ADAPT 1 SET a un pourcentage moyen légèrement supérieur. On notera l’augmentation de 45% d’occurrences de macroblocs skippés pour la séquence *Mobile*, avec les deux méthodes d’adaptation, pour laquelle l’augmentation est seulement de 3,5% avec la méthode MVComp.

### 5.2.2 Analyse des ensembles $\mathcal{P}_s^M$ et $\mathcal{P}_s^S$

La figure 5.2 représente le nombre moyen de prédicteurs pour chacun des ensembles  $\mathcal{P}_s^M$  et  $\mathcal{P}_s^S$  de la méthode ADAPT 2 SETS, pour chaque séquence, pour l’ensemble des QP. Le nombre moyen de prédicteurs est plus élevé pour la classe des zones en mouvement que pour la classe des zones statiques. Ce nombre moyen de prédicteurs est lié aux  $\delta_M$  et  $\delta_S$  qui ont été fixés respectivement à 1,6 et 1,05. En moyenne, pour l’ensemble des séquences testées, l’ensemble  $\mathcal{P}_s^S$  a 3,2 prédicteurs et l’ensemble  $\mathcal{P}_s^M$  a 9,1 prédicteurs.

Pour les séquences *Modo*, *Silent* et *BBC news*, le nombre de prédicteurs pour les zones statiques est élevé. Cependant, ces séquences sont composées d’une large majorité de zones statiques. Les

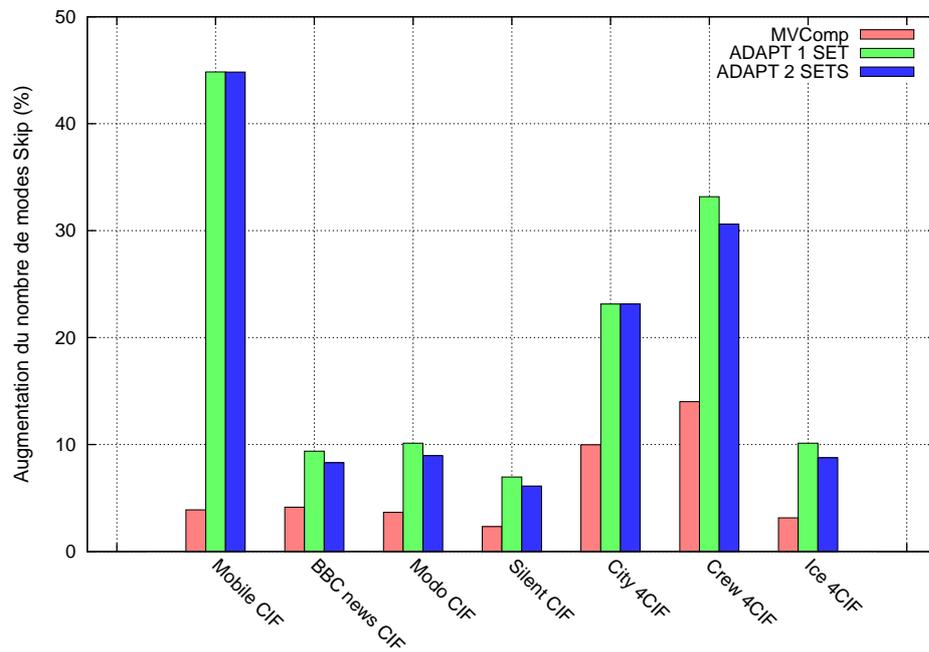


FIGURE 5.1 – Pourcentage d’augmentation moyen du nombre de macroblocs codés avec le mode Skip par séquence, pour MVComp, ADAPT 1 SET et ADAPT 2 SETS par rapport à la référence H.264/AVC en moyenne sur les 4 QP.

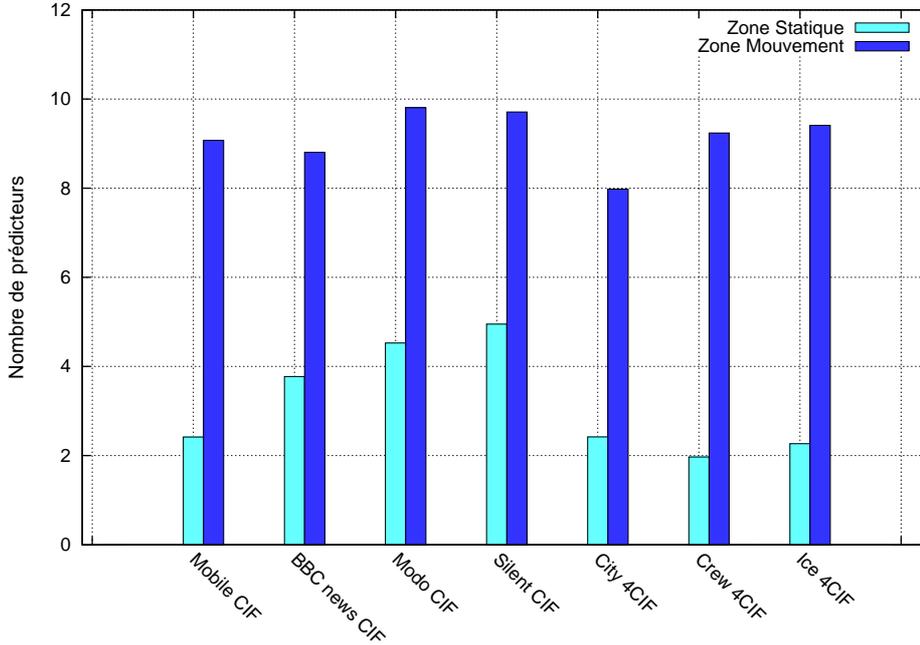


FIGURE 5.2 – Nombre moyen de prédicteurs utilisés pour chacun des ensembles de prédicteurs  $\mathcal{P}_s^M$  et  $\mathcal{P}_s^S$  de la méthode ADAPT 2 SETS, pour chaque séquence, en moyenne sur les 4 QP.

prédicteurs utilisés ont régulièrement la même valeur, généralement égale à zéro. Ceci est vérifié par les résultats du pourcentage de sélection des prédicteurs pour chacun des ensembles  $\mathcal{P}_s^M$  et  $\mathcal{P}_s^S$  donnés respectivement dans le tableau 5.2 et le tableau 5.1. Pour les zones statiques des séquences *Modo*, *Silent* et *BBC news*, les prédicteurs  $mv_{H.264}$ ,  $mv_{spaEXT}$ ,  $mv_{Scol_{R_0}}$ ,  $mv_{trans}$  et  $mv_0$  sont les plus sélectionnés. Pour les autres séquences, le mouvement est plus important. Par conséquent, les prédicteurs ont des valeurs différentes. Le nombre moyen de prédicteurs pour la classe “Statique” pour ces autres séquences est égal à 2,26 contre 4,41 pour les séquences *Modo*, *Silent* et *BBC news*. Pour les séquences avec un fond en mouvement (*Mobile*, *City* et *Crew*) aucun prédicteur n’a dominé la sélection des ensembles  $\mathcal{P}_s^S$ . La sélection des prédicteurs de l’ensemble  $\mathcal{P}_s^S$  de la séquence *Ice* est un peu différente de celle des autres séquences à fond fixe puisque, pour cette séquence, le prédicteur *collocated* est largement sélectionné. Il apparaît dans 94% des ensembles  $\mathcal{P}_s^S$ , dominant ainsi largement les autres prédicteurs (le deuxième prédicteur a 31% de sélection). Ceci est expliqué par la nature de cette séquence composée de nombreux objets en mouvement sur un fond fixe. Par conséquent, le vecteur *collocated* est le plus efficace pour les zones qui n’ont pas eu de mouvement dans les deux précédentes images (qui n’ont pas été traversées par un objet). Ce prédicteur est aussi le plus sélectionné pour l’ensemble  $\mathcal{P}_s^M$  de cette séquence des zones en mouvement.

Pour l’ensemble  $\mathcal{P}_s^M$  des séquences à fond fixe, le nombre de prédicteurs étant élevé, la sélection de chaque prédicteur est proche de l’équiprobabilité. Pour les séquences avec un fond en mouvement, la sélection du vecteur  $mv_0$  est faible, ce qui était prévisible. On remarquera que la sélection des

	Résolution	$mv_{H.264}$	$mv_{spaEXT}$	$mv_a$	$mv_b$	$mv_c$	$mv_0$	$mv_{Scol_{R_0}}$	$mv_{trans}$	$mv_{S_a}$	$mv_{S_b}$	$mv_{S_c}$
Mobile	CIF	33	20	21	21	13	25	22	24	22	22	18
BBC news	CIF	<b>70</b>	<b>57</b>	12	12	6	<b>58</b>	<b>71</b>	<b>63</b>	13	13	5
Modo	CIF	<b>82</b>	<b>78</b>	22	18	13	<b>65</b>	<b>63</b>	<b>63</b>	22	21	7
Silent	CIF	<b>93</b>	<b>91</b>	10	13	3	<b>87</b>	<b>87</b>	<b>86</b>	10	12	4
City	4CIF	34	36	36	21	15	14	13	9	27	23	16
Crew	4CIF	40	49	9	22	6	13	16	9	11	16	7
Ice	4CIF	31	24	9	6	1	24	<b>94</b>	29	7	2	1

TABLE 5.1 – Pourcentage d’appartenance à l’ensemble  $\mathcal{P}_s^S$  de chaque prédicteur pour la méthode ADAPT 2 SETS en fonction de chaque séquence en moyenne sur les 4 QP.

	Résolution	$mv_{H.264}$	$mv_{spaEXT}$	$mv_a$	$mv_b$	$mv_c$	$mv_0$	$mv_{Scol_{R_0}}$	$mv_{trans}$	$mv_{S_a}$	$mv_{S_b}$	$mv_{S_c}$
Mobile	CIF	<b>95</b>	<b>90</b>	<b>91</b>	<b>82</b>	<b>69</b>	13	<b>90</b>	<b>93</b>	<b>97</b>	<b>95</b>	<b>93</b>
BBC news	CIF	<b>83</b>	<b>86</b>	<b>84</b>	<b>84</b>	<b>76</b>	<b>72</b>	<b>79</b>	<b>81</b>	<b>84</b>	<b>81</b>	<b>72</b>
Modo	CIF	<b>96</b>	<b>94</b>	<b>93</b>	<b>91</b>	<b>83</b>	<b>91</b>	<b>81</b>	<b>85</b>	<b>92</b>	<b>92</b>	<b>83</b>
Silent	CIF	<b>94</b>	<b>94</b>	<b>93</b>	<b>90</b>	<b>81</b>	<b>88</b>	<b>86</b>	<b>82</b>	<b>92</b>	<b>91</b>	<b>82</b>
City	4CIF	<b>99</b>	<b>99</b>	<b>99</b>	<b>99</b>	<b>98</b>	1	5	2	<b>99</b>	<b>99</b>	<b>99</b>
Crew	4CIF	<b>96</b>	<b>97</b>	<b>96</b>	<b>97</b>	<b>94</b>	57	<b>64</b>	44	<b>95</b>	<b>97</b>	<b>89</b>
Ice	4CIF	<b>91</b>	<b>90</b>	<b>91</b>	<b>85</b>	<b>77</b>	<b>80</b>	<b>95</b>	<b>77</b>	<b>89</b>	<b>87</b>	<b>79</b>

TABLE 5.2 – Pourcentage d’appartenance à l’ensemble  $\mathcal{P}_s^M$  de chaque prédicteur pour la méthode ADAPT 2 SETS en fonction de chaque séquence en moyenne sur les 4 QP.

prédicteurs temporels pour la séquence City est minime par rapport à celles des prédicteurs spatiaux.

La figure 5.3 donne le nombre de bits moyen pour les ensembles  $\mathcal{P}_s^M$  et  $\mathcal{P}_s^S$  de la méthode proposée et celui de la méthode MVComp, séquence par séquence pour l’ensemble des QP étudiés. Dans le cadre de ces travaux, nous avons modifié le codage de l’indice du meilleur prédicteur sélectionné par rapport au codage utilisé et intégré dans le JM KTA. Pour les méthodes ADAPT 1 SET et ADAPT 2 SETS le nombre de bits transmis pour le meilleur prédicteur dépend du nombre de valeurs prises par l’ensemble des prédicteurs en compétition et non du nombre  $n$  de prédicteurs de l’ensemble. Dans ce chapitre, la configuration MVComp de référence utilise deux prédicteurs. Par conséquent, le processus d’élimination d’indices de prédicteurs par égalité dans le chapitre précédent est équivalent à celui développé dans ce chapitre. Le nombre de bits nécessaire au codage de l’indice du prédicteur pour les zones statiques est proche de celui de la méthode MVComp. En moyenne, 0,3 bits ont été utilisés pour le codage de l’indice du prédicteur pour la méthode MVComp contre

0,2 bits pour les zones statiques de la méthode proposée. L'objectif de réduction du coût des indices de prédicteur pour l'ensemble  $\mathcal{P}_s^S$  est donc atteint. Pour les zones en mouvement, le codage de l'indice de prédicteurs pour le mode Skip est beaucoup plus coûteux, atteignant presque 2 bits pour la séquence *Ice* et 1,4 bits en moyenne pour l'ensemble des tests. Ce chiffre est bien entendu lié au nombre de prédicteurs utilisés pour l'ensemble  $\mathcal{P}_s^M$ .

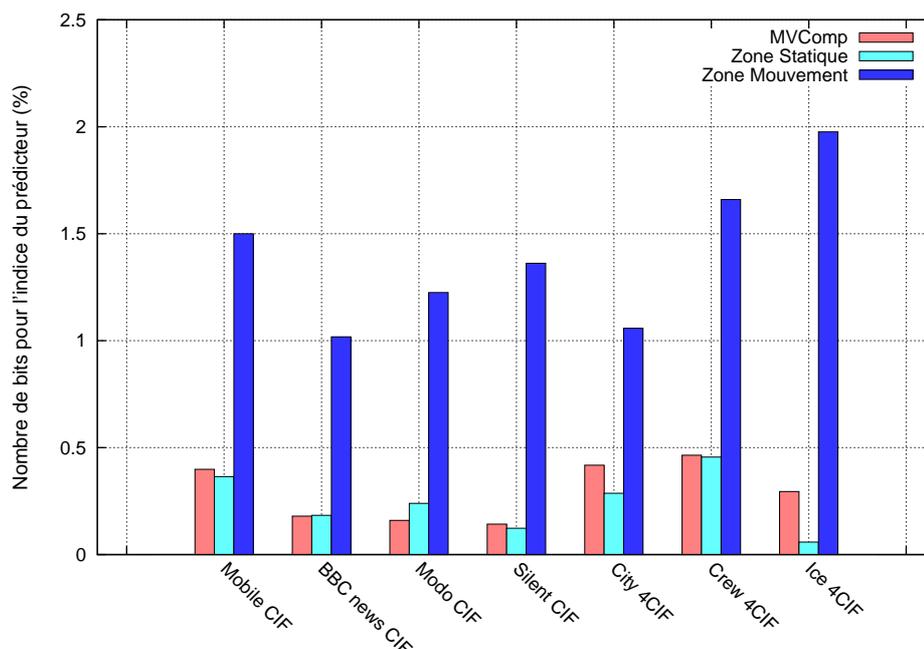


FIGURE 5.3 – Nombre de bits moyen par macrobloc nécessaire au codage de l'indice du prédicteur du mode Skip, pour chaque séquence, pour la méthode MVComp et pour chacun des ensembles  $\mathcal{P}_s^M$  et  $\mathcal{P}_s^S$  de la méthode d'adaptation d'ensemble de prédicteurs, en moyenne sur les 4 QP.

### 5.2.3 Analyse de la classification des macroblocs

La méthode proposée a pour but d'adapter l'ensemble des prédicteurs pour le mode Skip en fonction du contenu de la séquence. Il est par conséquent important que la segmentation corresponde aux résultats escomptés. La figure 5.4 est une représentation visuelle de la segmentation de l'image 3 de la séquence *Ice* 4CIF. La segmentation est correcte car les macroblocs composant la glace sont classifiés dans la zone statique et les macroblocs composant les personnes en mouvement sont classifiés dans la zone en mouvement.

La figure 5.5 représente la proportion des macroblocs sautés pour les zones statiques et les zones en mouvement par rapport au nombre total de macroblocs codés en Skip pour la méthode proposée. Pour les séquences comportant un fond en mouvement (*Mobile*, *City*, *Crew*), le mode Skip

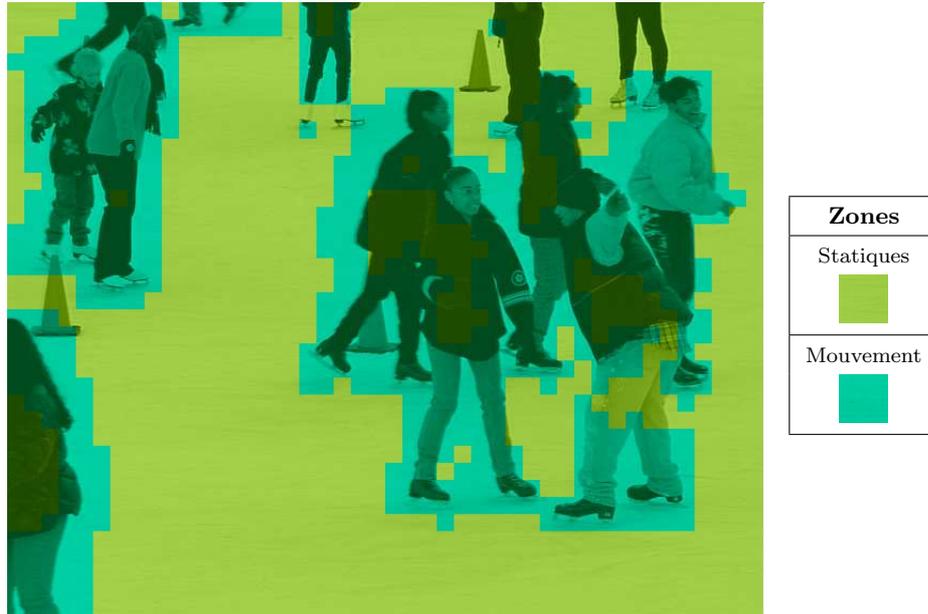


FIGURE 5.4 – Représentation spatiale de la classification des macroblocs en zones statiques ou en zones de mouvement, de l'image 3 de la séquence *Ice* 4CIF.

a été sélectionné pour la classe des zones en mouvement, ce qui est cohérent car la quasi totalité des macroblocs est classée dans ces zones. Pour les séquences à fond fixe, une majorité des macroblocs sautés se trouve dans les zones statiques.

#### 5.2.4 Gains sur le débit total

La figure 5.6 donne la réduction de débit en pourcentage par rapport à la configuration Baseline IPP de la référence H.264/AVC, de chaque séquence, pour les méthodes MVComp, ADAPT 1 SET, ADAPT 2 SETS. La constante  $\delta$  a été fixée empiriquement à 1,2 pour ADAPT 1 SET. Cette méthode d'adaptation est décodable par un décodeur incluant la méthode MVComp implémentée de la même manière que dans le JM KTA. En effet, ADAPT 1 SET n'est qu'un choix de codage, un choix de l'ensemble des prédicteurs du mode Skip slice par slice. La méthode ADAPT 2 SETS requiert, elle, le calcul de la classification des macroblocs sautés au décodeur et la transmission de deux ensembles slice par slice au lieu d'un seul dans les deux autres méthodes.

La réduction de débit est en moyenne de 5,9 % pour la méthode MVComp, de 6,4% pour l'adaptation d'un ensemble, et de 8% pour l'adaptation de deux ensembles de prédicteurs. La méthode ADAPT 1 SET obtient de meilleures performances que la méthode MVComp, excepté pour les séquences *BBC news* et *Ice*. La méthode ADAPT 2 SETS est, elle, systématiquement meilleure que la méthode MVComp et obtient un gain de 2,2% par rapport à cette dernière. La réduction de débit, par rapport à MVComp, est plus élevée pour les séquences à fond fixe.

Nous avons proposé dans [JLPP07] une adaptation d'un ensemble de prédicteurs composé de

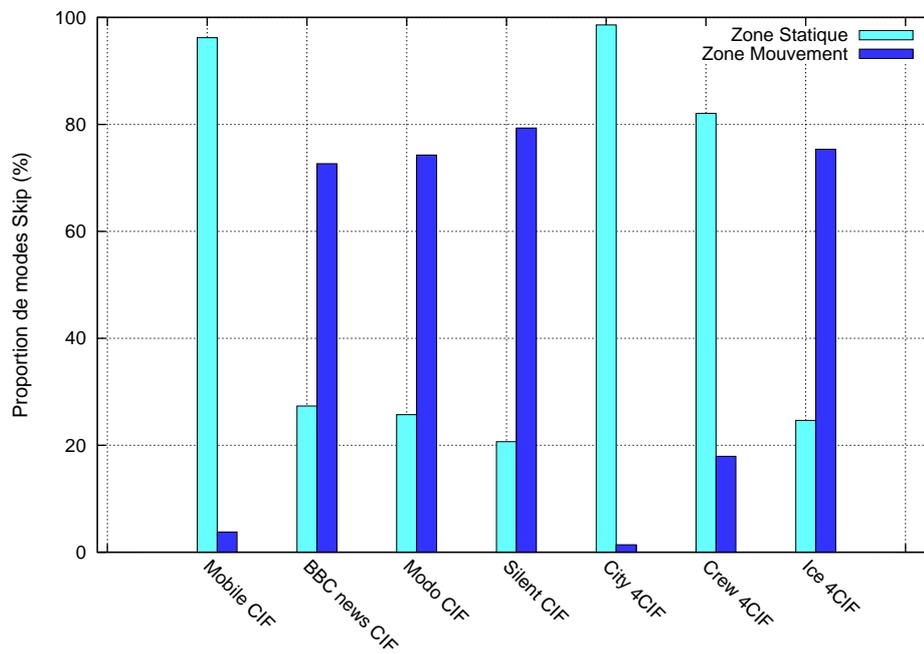


FIGURE 5.5 – Proportion de macroblocs skippés sélectionnés dans l'ensemble  $\mathcal{P}_s^M$  ou l'ensemble  $\mathcal{P}_s^S$  par rapport au nombre total de macroblocs codés en mode Skip, pour chaque séquence en moyenne sur les 4 QP.

trois prédicteurs avec un critère équivalent à celui présenté dans la section 5.1.1. Cette adaptation d'un seul ensemble de trois prédicteurs donne des gains systématiques par rapport à la méthode MVComp. Les réductions de débit sont en moyenne équivalentes à celles proposées pour ADAPT 1 SET dans ce chapitre.

L'adaptation des ensembles  $\mathcal{P}_s^M$  et  $\mathcal{P}_s^S$  est plus efficace que l'adaptation d'un seul ensemble. Cependant, pour les séquences *Mobile* et *City*, les deux méthodes sont équivalentes. Pour ces deux séquences, les macroblocs sautés se trouvent, pour la quasi-totalité, dans la même classe pour la méthode d'adaptation orientée contenu (cf. figure 5.5).

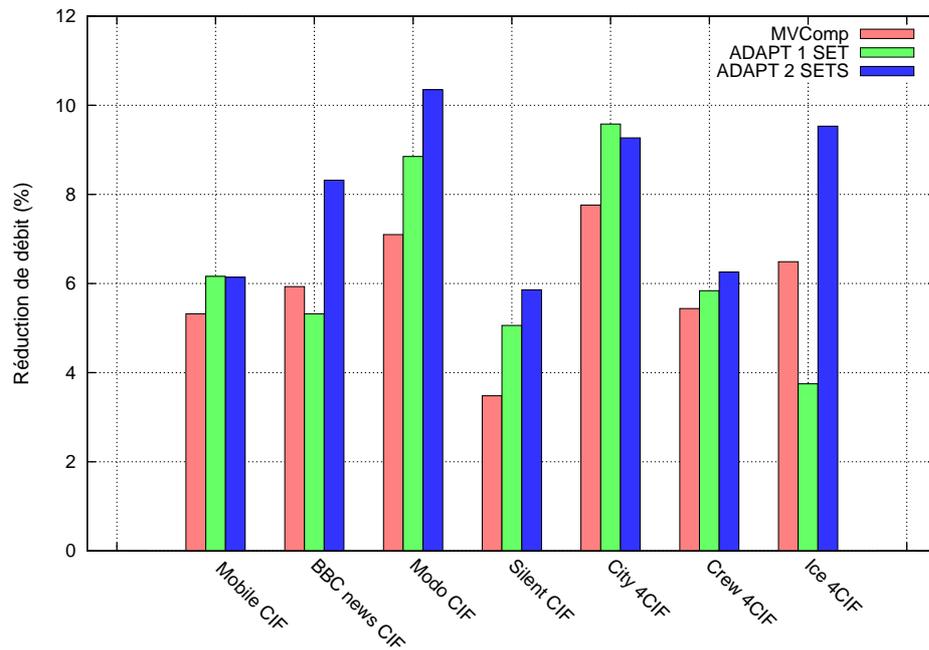


FIGURE 5.6 – Réduction du débit pour la méthode MVComp, l'adaptation d'un ensemble de prédicteurs et l'adaptation de deux ensembles de prédicteurs pour le mode Skip, pour chaque séquence en moyenne sur les 4 QP.

### 5.3 Conclusion

Dans ce chapitre, nous avons proposé une méthode d'adaptation d'ensembles de prédicteurs pour le mode Skip induite par les résultats de la méthode de compétition exhaustive des prédicteurs de vecteurs mouvement pour ce mode. En effet, les prédicteurs utilisés dans la meilleure configuration de la méthode du chapitre précédent ne sont pas les meilleurs pour l'ensemble des séquences. Nous proposons donc une adaptation d'un ensemble de prédicteurs pour le mode Skip basée sur le nombre de minimisations du critère débit-distorsion pour tous les prédicteurs définis. L'adaptation

de l'ensemble est faite slice par slice avec ce critère a priori.

En remarquant que le mode Skip est largement sélectionné pour les zones comportant peu ou pas de mouvement, nous avons proposé une adaptation orientée contenu des séquences en utilisant deux ensembles de prédicteurs. Le but de cette méthode est de diminuer le coût de l'indice du prédicteur où le mode Skip est déjà largement sélectionné et d'augmenter l'occurrence de ce mode très efficace en terme de compromis débit-distorsion pour les zones où le Skip est moins sélectionné en utilisant un ensemble plus large de prédicteurs. Les résultats sont au rendez-vous avec une réduction de 2,2% de débit par rapport à la meilleure configuration de la méthode de compétition des prédicteurs de vecteurs mouvement.

# Chapitre 6

## Intra 1D

*Le coût d'un slice Intra dans le standard H.264/AVC est élevé par rapport au coût d'un slice Inter (P ou B) comme on l'a vu dans le chapitre 1. Certes les slices Intra exploitent uniquement les redondances spatiales alors que les slices Inter exploitent à la fois les redondances spatiales et temporelles. Cependant le nombre de possibilités de codage des slices Intra est faible par rapport à celui des slices P et des slices B. D'ailleurs l'augmentation de ce nombre de possibilités de codage est une voie de recherche largement étudiée, comme on a pu le voir dans l'état de l'art (cf. chapitre 1). Dans ce chapitre nous proposons un nouveau mode de codage prédictif Intra permettant de réduire la distance spatiale entre le signal de référence utilisé et les pixels à prédire. Cette méthode utilise un partitionnement 1D du macrobloc obtenu par un algorithme de morphologie mathématique. De plus, pour accroître le nombre de compétitions disponibles pour le codage Intra nous utilisons trois ordres de parcours de ces partitionnements 1D, au lieu de l'unique parcours lexicographique de la référence.*

### 6.1 Description de la méthode

Le standard H.264/AVC utilise un partitionnement carré du macrobloc. Ce partitionnement est adapté à la transformée carrée (DCT  $n \times n$ ) qui est utilisée dans la quasi-totalité des standards vidéo et d'image. Pour un partitionnement carré, avec l'ordre de parcours "lexicographique", l'erreur de prédiction en bas à droite d'un bloc est généralement plus grande que l'erreur en haut à gauche, car la distance spatiale entre les pixels à prédire et les pixels de référence (ligne du haut et colonne de gauche) est plus élevée pour les pixels en bas à droite du bloc courant. Nous proposons de résoudre ce problème en utilisant un découpage du macrobloc en partitions 1D pour lequel le pixel à prédire a au moins un pixel voisin appartenant au signal de référence lors d'un parcours lexicographique. Définissons tout d'abord la mesure  $ANRS_E$  (Average Nearest Reference Signal) donnant la moyenne de la distance spatiale minimum entre tous les pixels à prédire d'une partition Intra et le signal de référence utilisé :

---

1. Publications de l'auteur associées : [Conf2], [ConfFR1], [Brev5] et [Brev6].

$$ANRS_E = \frac{\sum_{i,j} Near(p_{i,j}, RS)}{card(E)} \quad (6.1)$$

où  $p_{i,j}$  est le pixel à prédire dans la partition  $E$  à la position  $(i, j)$ ,  $RS$  est le signal de référence utilisé pour le calcul des prédicteurs de la partition courante. La fonction  $Near(p_{i,j}, RS)$  donne la distance euclidienne (en nombre de pixels parcourus) entre le pixel  $p_{i,j}$  et le pixel le plus proche du signal de référence  $RS$ . Cette mesure est indépendante de la construction des prédicteurs. Pour l'Intra  $4 \times 4$  de la norme H.264/AVC la mesure  $ANRS_{4 \times 4}$  est égale à 1,9 et la mesure  $ANRS_{16 \times 16}$  de l'Intra  $16 \times 16$  est égale à 5,8. Avec un parcours lexicographique des partitions 1D, la mesure  $ANRS_{1D}$  d'une partition 1D est systématiquement égale à 1 ce qui signifie que le signal à prédire est plus proche du signal de référence que pour les partitions bloc de l'Intra d'H.264/AVC. Pour augmenter l'efficacité globale de la méthode, nous proposons deux autres ordres de parcours en plus du parcours lexicographique : un parcours "bidirectionnel" et un parcours "hiérarchique". Des prédicteurs adaptés à ce partitionnement 1D et à ces trois ordres de parcours ont aussi été définis. Enfin, les résiduels sont transformés à l'aide d'une DCT 1D pour que le codage soit adapté aux partitionnements 1D proposés.

### 6.1.1 Partitionnement 1D

Dans cette section nous proposons un algorithme basé sur la morphologie mathématique permettant d'obtenir un partitionnement 1D du macrobloc. Tout d'abord, définissons  $Par_i, \forall i = 1..P$  la partition numéro  $i$  d'un macrobloc découpé en  $P$  partitions. L'ensemble des partitions remplit le macrobloc et ses partitions ne se chevauchent pas ( $\forall i = 1..P, \forall j = 1..P, i \neq j, Par_i \cap Par_j = \emptyset$ ). De plus les partitions n'ont pas forcément le même nombre de pixels. Enfin notons  $Par_0$  le signal de référence en dehors du macrobloc courant. Dans notre algorithme la partition initiale  $Par_1$  est une forme linéaire comme une colonne, une ligne, un 'L', un pixel, un contour (extrait de l'image originale ou des images précédemment encodées), etc. Le partitionnement 1D est obtenu par propagation de cette partition initiale  $Par_1$ . La propagation du rang  $i$  qui donne la partition  $Par_i$  est obtenue par un opérateur de dilatation morphologique à partir de la partition  $Par_{i-1}$  donnée par la formule suivante :

$$Par_i = Par_{i-1} \oplus B, \forall i = 2..P \quad (6.2)$$

où  $\oplus$  est l'opérateur de dilatation morphologique. L'élément structurant  $B$  qui permet d'obtenir un partitionnement 1D peut être par exemple une croix de taille  $3 \times 3$  tel que :

0	1	0
1	1	1
0	1	0

Cet élément structurant permet une dilatation 4-connexe. La figure 6.1 donne des exemples de dilatations obtenues avec cet opérateur en fonction de différentes partitions initiales. Notons que pour la dilatation d'un pixel (cf. figure 6.1(d)) l'élément structurant est un carré plein de taille  $3 \times 3$ .

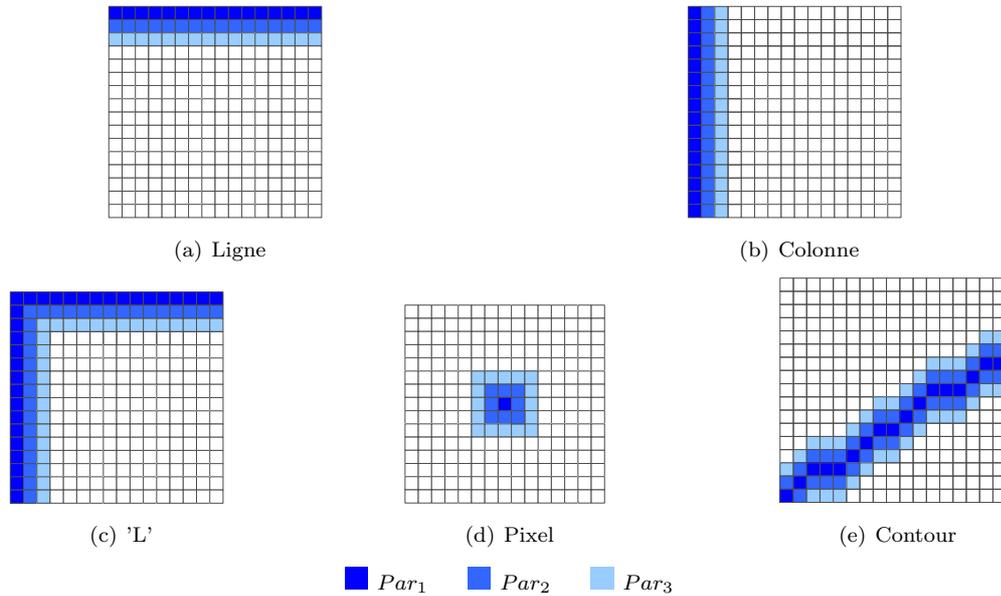


FIGURE 6.1 – Exemple de partition initiale  $Par_1$  et les partitions  $Par_2$  et  $Par_3$  engendrées par la dilatation au rang 2 et 3.

Dans cette figure 6.1, les partitions initiales Ligne et Colonne peuvent être adaptées au codage de macrobloc composé de formes ligne ou colonne. Le partitionnement 'L', lorsque l'ordre de parcours lexicographique est utilisé, permet d'avoir l'ensemble du signal de référence le plus proche dans la partition de rang inférieur, servant de base au calcul des prédicteurs. Le partitionnement Pixel permet de concentrer l'énergie d'un contour ou d'un objet sur un seul pixel et de coder les autres partitionnements à l'aide d'un signal de référence provenant de deux objets. Enfin, lorsque l'on est capable d'estimer la présence et la forme d'un contour dans un macrobloc, le partitionnement initial Contour permet de diffuser ce contour efficacement dans les autres partitions.

Cette propagation est un algorithme déterministe qui produit un seul partitionnement possible du macrobloc. Par conséquent, seule la partition  $Par_1$  doit être transmise au décodeur, ce qui correspond à la transmission d'un index si un ensemble de partitions initiales est connu du codeur et du décodeur.

### 6.1.2 Ordre de parcours des partitions 1D

Nous proposons trois ordres de parcours pour nos partitionnements 1D : un parcours lexicographique, un parcours bidirectionnel et un parcours hiérarchique. L'ordre lexicographique est un parcours séquentiel des partitions obtenues avec l'algorithme de morphologie mathématique, soit un parcours des partitions  $Par_i$  pour  $i$  de 1 à  $P$ . Pour l'ordre de parcours bidirectionnel, une partition sur deux est codée avec un retard d'une partition, dans le but d'utiliser deux partitions directement voisines pour le calcul des prédicteurs des partitions retardées. Cet ordre est représenté dans la figure 6.2(a). La possible efficacité de cet ordre de parcours réside sur le fait qu'il est parfois

plus efficace, au sens du critère débit-distorsion, de coder certaines partitions à l'aide d'un signal de référence plus éloigné puis de traiter les partitions restantes avec un signal de référence provenant de plusieurs directions engendrant des corrélations spatiales plus élevées. Ceci a été exploité dans la méthode BIP de [STC07] mais aussi pour les images B qui ont comme images de référence des images dans le passé et des images dans le futur (une image B est entourée par deux images P). Les images B hiérarchiques étendent cette idée. Pour la hiérarchie de ces images, on utilise en général [SMW05] l'ordre pyramidal. Nous proposons d'utiliser cet ordre hiérarchique pyramidal pour le parcours de nos partitions 1D, comme représenté dans la figure 6.2(b).

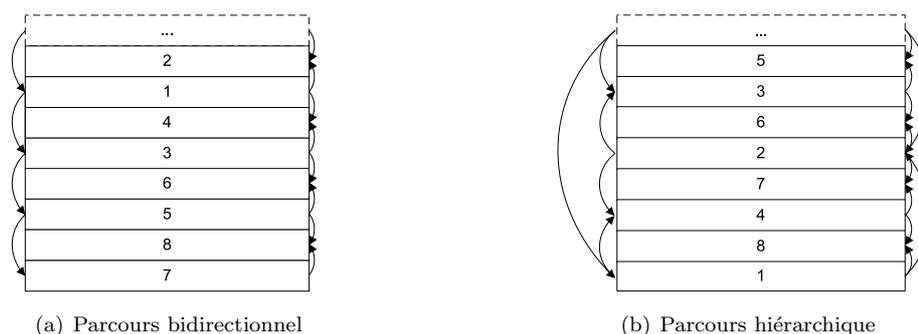


FIGURE 6.2 – Ordres de parcours bidirectionnel et hiérarchique pour un partitionnement Ligne.

### 6.1.3 Prédicteurs

Les partitions 1D proposées et les trois ordres de parcours impliquent la définition de prédicteurs spécifiques. Dans cette section, nous définissons, pour chacun des ordres de parcours, des prédicteurs correspondant à un partitionnement en lignes du macrobloc. Les définitions de ces prédicteurs peuvent se généraliser aux autres partitionnements.

#### 6.1.3.1 Ordre de parcours lexicographique

Dans l'ordre de parcours lexicographique le signal de référence utilisé pour une partition  $Par_i$  est la partition  $Par_{i-1}$  ou tout autre pixel déjà décodé. Le signal de référence utilisé pour une partition ligne est représenté dans la figure 6.3. Si ces pixels n'existent pas (pixels en dehors de l'image), ils sont remplacés par la valeur 128 dans notre implémentation. Les prédicteurs définis pour l'Intra Ligne pour le parcours lexicographique sont les suivants :

- *Prev\_line* : la ligne juste au dessus de la ligne courante, soit la partition  $Par_{i-1}$ . Ce prédicteur contient les pixels de  $A$  à  $P$ .
- *Prev\_line<sup>2d</sup>* : la deuxième ligne précédente, soit la partition  $Par_{i-2}$ . Ce prédicteur contient les pixels de  $A'$  à  $P'$ .
- *Left\_pix* : le pixel à gauche de la ligne courante ( $S$ ) ; tous les pixels de ce prédicteur sont égaux à  $S$ .

- $Left\_pix + 1$  : le pixel en bas du pixel de gauche ( $T$ ) ; tous les pixels de ce prédicteur sont égaux à  $T$ .
- $Left\_pix - 1$  : le pixel en haut du pixel de gauche ( $R$ ) ; tous les pixels de ce prédicteur sont égaux à  $R$ .
- $Prev\_Left\_S$  : la ligne précédente décalée vers la gauche. Ce prédicteur contient les pixels de  $R$  à  $O$ . Si le pixel  $O$  n'existe pas il est remplacé par le pixel  $A$ .
- $Prev\_Right\_S$  : la ligne précédente décalée vers la droite. Ce prédicteur contient les pixels de  $B$  à  $Q$ . Si le pixel  $Q$  n'existe pas ou s'il n'a pas encore été décodé il est remplacé par le pixel  $P$ .
- $DC_L$  : la moyenne de la ligne précédente.
- $Balanced_L$  : la moyenne pondérée en fonction de la distance spatiale entre le pixel de gauche  $S$  et les pixels de la ligne précédente (pixels de  $A$  à  $P$ ).
- $Left\_2Prev$  : la moyenne pondérée entre le pixel de gauche  $S$  et les pixels de la ligne précédente. Chaque pixel de ce prédicteur est égal à la moyenne entre  $S$  et deux fois le pixel juste au dessous.
- $3\_Prev\_Line$  : la moyenne pour chaque pixel, des pixels provenant des prédicteurs  $Prev\_line$ ,  $Prev\_Left\_S$ ,  $Prev\_Right\_S$ .
- $A\_DC\_Prev\_line$  : moyenne pour chaque pixel entre les prédicteurs  $Prev\_line$  et  $DC$ .
- $Sub\_Prev\_line$  : le demi pixel calculé avec les pixels de  $R$  à  $Q$ .
- Nous avons aussi utilisé les prédicteurs de l'Intra  $16 \times 16$  pour définir les prédicteurs  $VERT\_16 \times 16_L$ ,  $DC\_16 \times 16_L$ ,  $PLANE\_16 \times 16_L$ . Chacun de ces prédicteurs est une copie de la ligne du macrobloc prédicteur Intra  $16 \times 16$  ayant la même position que la ligne courante. Remarquons que nous n'avons pas défini le prédicteur  $HOR\_16 \times 16_L$  car il est identique au prédicteur  $Left\_pix$ .

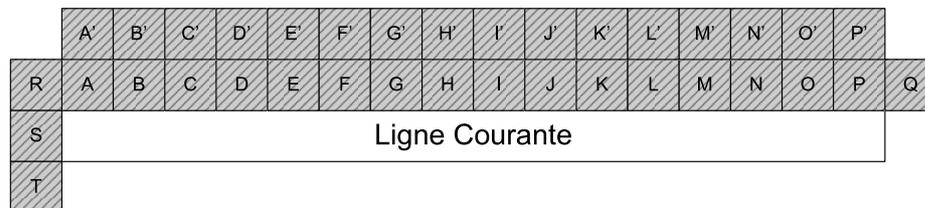


FIGURE 6.3 – Pixels du signal de référence utilisés pour le calcul des prédicteurs du partitionnement Ligne pour le parcours lexicographique.

### 6.1.3.2 Ordre de parcours bidirectionnel

Pour le parcours bidirectionnel, nous pouvons considérer deux types de prédiction :

- la prédiction pour laquelle le signal de référence pour la partition courante peut être la partition  $Par_{i-2}$  ou tout autre pixel déjà décodé. Pour faciliter la compréhension, nous modifierons notre notation des partitions  $Par_i$  par  $Par_i^P$  pour ce type de prédiction.
- la prédiction pour laquelle le signal de référence pour la partition courante  $Par_i$  peut être

$Par_{i-1}$  et  $Par_{i+1}$  ou tout autre pixel déjà décodé. Nous noterons les partitions de ce type de prédiction  $Par_i^B$ .

Les prédicteurs définis pour les partitions  $Par_i^P$  du partitionnement Ligne sont les mêmes que ceux de l'ordre de parcours lexicographique. Cependant, ces prédicteurs utilisent, comme signal de référence, la deuxième ligne au-dessus et non la première, car cette dernière n'a pas encore été décodée. Pour les partitions  $Par_i^B$ , nous avons défini neuf prédicteurs utilisant les pixels des partitions précédentes  $Par_{i-1}^P$  et  $Par_{i+1}^P$  représentés dans la figure 6.4 pour un partitionnement ligne. Ces prédicteurs sont les suivants :

- $B\_Prev\_line$  : la ligne juste au dessus de la ligne courante soit la partition  $Par_{i-1}^P$ . Ce prédicteur contient les pixels de  $A$  à  $P$ .
- $B\_Follo\_line$  : la ligne juste en dessous de la ligne courante soit la partition  $Par_{i+1}^P$ . Ce prédicteur contient les pixels de  $A''$  à  $P''$ .
- $B\_Left\_pix$  : le pixel à gauche de la ligne courante ( $S$ ).
- $B\_Left\_pix + 1$  : le pixel en bas du pixel de gauche ( $T$ ).
- $B\_Left\_pix - 1$  : le pixel en haut du pixel de gauche ( $R$ ).
- $B\_DCL$  : la moyenne des prédicteurs  $B\_Prev\_line$  et  $B\_Follo\_line$ .
- $B\_Prev\_Follo\_lines$  : la moyenne pixel à pixel entre les prédicteurs  $B\_Prev\_line$  et  $B\_Follo\_line$ .
- $B\_Prev\_Follo\_Left\_S$  : la moyenne pixel à pixel entre les pixels de la ligne au-dessus décalée vers la gauche ( $R$  à  $O$ ) et la ligne au-dessous décalée vers la droite soient les pixels  $B''$  à  $P''$ . en ajoutant un pixel virtuel pour le dernier pixel de cette ligne égale à la valeur de  $P''$ .
- $B\_Prev\_Follo\_Right\_S$  : la moyenne pixel à pixel entre les pixels de la ligne au-dessus décalée vers la droite ( $B$  à  $Q$ ) et la ligne au-dessous décalée vers la gauche soient les pixels  $T$  à  $O''$ . Si le pixel  $Q$  n'existe pas ou s'il n'a pas encore été décodé, il est remplacé par le pixel  $P$ .

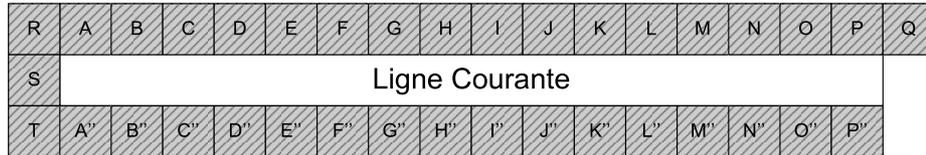


FIGURE 6.4 – Pixels du signal de référence utilisés pour le calcul des prédicteurs du partitionnement Ligne pour le parcours bidirectionnel dans le cas des partitions utilisant les partitions  $Par_i^B$ .

### 6.1.3.3 Ordre de parcours hiérarchique

La partition courante  $Par_i$  pour l'ordre de parcours hiérarchique a pour signal de référence les partitions de niveau supérieur dans la hiérarchie ou tout autre pixel déjà décodé. La partition courante  $Par_i$  est entourée par deux partitions ; ce qui correspond aux partitions  $Par_i^B$  de l'ordre de parcours bidirectionnel. On notera ces partitions  $Par_i^{BH}$ . Par conséquent, les prédicteurs de l'ordre de parcours hiérarchique ont la même définition que les prédicteurs du parcours bidirectionnel pour les partitions  $Par_i^B$ . On notera donc ces prédicteurs comme ceux du parcours bidirectionnel en

remplaçant la lettre  $B$  par  $BH$  devant chacun de ces prédicteurs. Les pixels de référence utilisés pour le calcul de ces prédicteurs sont représentés dans la figure 6.5. Il existe cependant une exception : la première partition de la hiérarchie  $Par_N^P$ . Son signal de référence est  $Par_0$ . Dans ce cas, les prédicteurs des partitions  $Par_i^P$  du parcours bidirectionnel se généralisent à ce cas particulier de l'ordre de parcours hiérarchique. Le signal de référence pour ces prédicteurs est alors la partition  $Par_0$  (la ligne voisine du macrobloc courant) et non la seconde ligne précédente (parcours bidirectionnel).

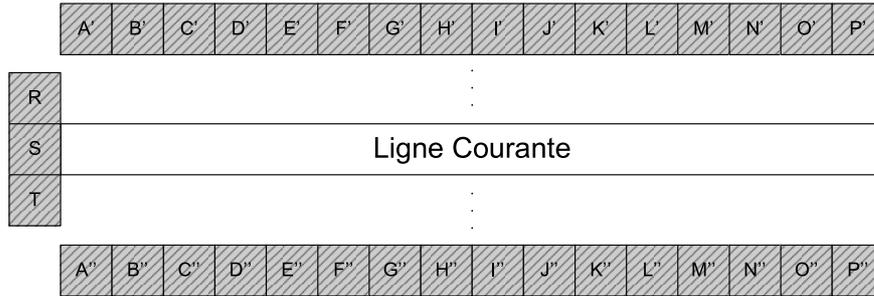


FIGURE 6.5 – Pixels du signal de référence utilisés pour le calcul des prédicteurs du partitionnement Ligne pour le parcours hiérarchique dans le cas des partitions  $Par_i^{BH}$ .

#### 6.1.4 Transformée DCT 1D

Le standard H.264/AVC utilise une transformée DCT carrée qui peut être de taille  $4 \times 4$  ou  $8 \times 8$  pour coder les résiduels de bloc des prédictions Intra et Inter. Afin de coder au mieux les résiduels obtenus lors de l'utilisation de nos partitionnements linéaires, nous utilisons une transformée DCT 1D donnée par la formule suivante :

$$X_k = \alpha(k) \sum_{n=0}^{N-1} x_n \cos \left[ \frac{\pi}{N} \left( n + \frac{1}{2} \right) k \right], \quad 0 \leq k \leq N-1 \quad (6.3)$$

où

$$\alpha(0) = \frac{1}{\sqrt{2}}, \quad \alpha(k) = 1, \quad 1 \leq k \leq N-1$$

Pour quantifier nos résiduels transformés, nous utilisons la quantification de la DCT  $4 \times 4$  pour le codage Intra, mais nous n'utilisons pas l'outil non normatif "adaptive quantization offset" inclus dans le logiciel de référence [JM]. L'offset de la quantification est donc égal à zéro.

## 6.2 Résultats expérimentaux

### 6.2.1 Conditions de tests

#### 6.2.1.1 Partitionnements 1D actifs

Nous avons sélectionné deux partitions initiales  $Par_1$  : une ligne et une colonne. Ces deux partitions initiales nous donnent respectivement un partitionnement Ligne et un partitionnement Colonne du macrobloc. Ces deux nouveaux partitionnements Intra sont en compétition avec les partitionnements Intra de la norme (Intra  $16 \times 16$ ,  $4 \times 4$  et  $8 \times 8$  pour le profil High). En considérant que les modes de codage Intra 1D comportent autant de partitions que l'Intra  $4 \times 4$ , nous avons ajouté l'indice de la compétition après le codage du mode Intra  $4 \times 4$ . Cet indice indique donc que le macrobloc est codé en Intra  $4 \times 4$  ou en Intra 1D. Le codage des modes Intra 1D nécessite donc plus de bits que les autres modes Intra de la norme. Il s'agit d'un choix volontaire qui permet de montrer que la sélection d'un mode Intra 1D par le critère RD réside dans une réelle amélioration du codage par rapport au mode Intra de la norme.

Dans la section 6.1.1, nous avons cité, en plus des partitionnements Ligne et Colonne, les partitionnements Contour, Pixel et 'L' (qui n'est qu'un cas particulier du partitionnement Pixel). Nous n'avons pas implémenté ces partitionnements car leur développement est très long. En effet, ils engendrent des partitions qui n'ont pas le même nombre de pixels. Par conséquent, l'implémentation dans le JM KTA devient très compliquée puisqu'il est nécessaire de modifier à la fois la quantification, le codage des coefficients et tous les autres outils définis pour un codage de 16 ou 64 coefficients. Il faut alors déterminer, ou re-déterminer, un ensemble de paramètres optimaux beaucoup plus nombreux que ceux que nous avons estimés pour les partitionnements Ligne et Colonne. L'implémentation des autres partitionnements nécessite donc une multitude de tests supplémentaires par rapport à ceux que nous allons décrire dans les sections suivantes. Cependant, notre hypothèse d'utilisation de partitionnements linéaires pour l'amélioration du codage Intra est validée par les résultats que nous décrivons ci-dessous. L'implémentation des autres partitionnements devrait améliorer ces résultats. Cette perspective sera probablement envisagée par notre équipe.

#### 6.2.1.2 Ordre de parcours actifs

Pour chacun des modes Intra 1D, nous avons utilisé les trois ordres de parcours définis dans la section 6.1.2. Cependant, des tests complémentaires sur l'efficacité des modes Intra 1D ont révélé qu'il était plus efficace d'utiliser deux ensembles de prédicteurs pour le parcours lexicographique et un seul pour les ordres de parcours bidirectionnel et hiérarchique. Ces tests complémentaires ont été effectués avec le profil Baseline et sont reportés dans l'annexe B. L'indice permettant de distinguer les ordres de parcours de chacun des deux modes Ligne et Colonne et les deux ensembles de prédicteurs pour l'ordre de parcours lexicographique est écrit après celui indiquant l'utilisation de l'Intra 1D. Cet indice est codé systématiquement sur 4 bits pour le CAVLC et binéarisé avec 4 bits pour le CABAC.

### 6.2.1.3 Prédicteurs actifs

Les prédicteurs que nous avons définis dans la section 6.1.3 pour l'Intra Ligne, sont facilement généralisables au cas de l'Intra Colonne. Pour l'ordre de parcours lexicographique du mode Colonne, ces prédicteurs sont respectivement :  $Prev\_Col$ ,  $Prev\_Col^{2d}$ ,  $Up\_pix$ ,  $Up\_pix + 1$ ,  $Up\_pix - 1$ ,  $Prev\_Up\_S$ ,  $Prev\_Down\_S$ ,  $DC_C$ ,  $Balanced_C$ ,  $Up\_2Prev$ ,  $3\_Prev\_Col$ ,  $A\_DC\_Prev\_Col$ ,  $Sub\_Prev\_Col$ ,  $HOR.16 \times 16_C$ ,  $DC.16 \times 16_C$ ,  $PLANE.16 \times 16_C$ . Pour l'ordre de parcours bidirectionnel, les prédicteurs pour les partitions  $Par_i^B$  sont respectivement :  $B\_Prev\_Col$ ,  $B\_Follo\_Col$ ,  $B\_Up\_pix$ ,  $B\_Up\_pix + 1$ ,  $B\_Up\_pix - 1$ ,  $B\_DC_C$ ,  $B\_Prev\_Follo\_Cols$ ,  $B\_Prev\_Follo\_Up\_S$ ,  $B\_Prev\_Follo\_Down\_S$ . De même, pour le parcours hiérarchique les prédicteurs définis seront notés de la même manière que ceux des partitions  $Par_i^B$  du parcours bidirectionnel en remplaçant la lettre  $B$  par  $BH$ .

Les tests complémentaires sur les modes Intra 1D, proposés dans l'annexe B, nous ont permis de définir le nombre et les prédicteurs à utiliser pour chacun des modes Intra 1D et ordres de parcours. La meilleure configuration du nombre de prédicteurs et de leur ordonnancement est décrite dans le tableau 6.1. Nous avons donc sélectionné deux ensembles de prédicteurs pour l'ordre de parcours lexicographique du mode Ligne avec 2 et 9 prédicteurs, deux ensembles de prédicteurs pour l'ordre de parcours lexicographique du mode Colonne avec 1 et 9 prédicteurs. Le mode bidirectionnel du mode Ligne a 2 prédicteurs pour les partitions  $Par_i^P$  et 1 prédicteur pour les partitions  $Par_i^B$ . Le mode Colonne pour l'ordre de parcours bidirectionnel a, quant à lui, 9 prédicteurs pour les partitions  $Par_i^P$  et 2 prédicteurs pour les partitions  $Par_i^B$ . Enfin, l'ordre de parcours hiérarchique a 5 prédicteurs pour le mode Ligne et 2 prédicteurs pour le mode Colonne. Notons que pour l'ensemble des ordres de parcours, le premier prédicteur est toujours la ligne (dans le cas du mode Ligne) ou la colonne (dans le cas du mode Colonne) la plus proche pour l'ordre de parcours lexicographique et une moyenne des lignes les plus proches ou des colonnes les plus proches pour les deux autres ordres de parcours. Le deuxième prédicteur est toujours le pixel à gauche pour le mode Ligne et le pixel au dessus pour le mode Colonne. Il est à noter que ces prédicteurs ont été ordonnancés en fonction de leur nombre de sélections et qu'on les retrouve systématiquement dans les deux premières positions, quelque soit le nombre de prédicteurs utilisés et l'ordre de parcours.

Les prédicteurs des modes Intra 1D, pour l'ensemble des ordres de parcours, sont codés de manière équivalente au codage des prédicteurs de l'Intra  $4 \times 4$  ou de l'Intra  $8 \times 8$  que nous avons décrit dans la section 1.2 du chapitre 1. Cependant, nous n'avons pas forcément 9 prédicteurs comme dans la référence pour chacun des partitionnements et ordres de parcours. Lorsque que 17, 9, 5 et 3 prédicteurs sont utilisés pour ces nouveaux modes, le même codage prédictif des indices de la référence est appliqué. Avec 2 prédicteurs, l'utilisation d'une prédiction d'indice de prédicteur n'est pas utile. L'indice est alors codé sur 1 bit et lorsque l'on utilise un seul prédicteur il n'est pas nécessaire d'utiliser d'indice. Le prédicteur le plus probable pour l'ordre de parcours bidirectionnel est le prédicteur utilisé pour le codage de la partition  $Par_{i-2}^B$  pour les partitions  $Par_i^B$ , et le prédicteur de la partition  $Par_{i-2}^P$  pour les partitions  $Par_i^P$ . Pour les autres cas, le prédicteur de la partition  $Par_{i-1}$  est utilisé comme prédicteur le plus probable pour les partitions  $Par_i$ . Si  $i = 1$ , le prédicteur le plus probable est égal à celui de la partition  $Par_0$  si celle-ci est du même type et du même ordre de parcours, sinon l'indice de ce prédicteur est égal à 0.

Indice	Lexicographique		Bidirectionnel		Hiérarchique
	Ensemble 1	Ensemble 2	$Par^P$	$Par^B$	
<b>Ligne</b>					
1	<i>Prev_line</i>	<i>Prev_line</i>	<i>Prev_line</i>	<i>B_Prev_Follo_lines</i>	<i>BH_Prev_Follo_lines</i>
2	<i>Left_pix</i>	<i>Left_pix</i>	<i>Left_pix</i>		<i>BH_Left_pix</i>
3	<i>Balanced<sub>L</sub></i>				<i>BH_Prev_Follo_Left_S</i>
4	<i>Left_2Prev</i>				<i>BH_Prev_line</i>
5	<i>Prev_Right_S</i>				<i>BH_Follo_line</i>
6	<i>Left_pix + 1</i>				
7	<i>A_DC_Prev_line</i>				
8	<i>Sub_Prev_line</i>				
9	<i>Prev_Left_St</i>				
<b>Colonne</b>					
1	<i>Prev_Col</i>	<i>Prev_Col</i>	<i>Prev_Col</i>	<i>B_Prev_Follo_Cols</i>	<i>BH_Prev_Follo_Cols</i>
2	<i>Up_pix</i>		<i>Up_pix</i>	<i>B_Prev_Col</i>	<i>BH_Up_pix</i>
3	<i>Prev_Up_St</i>		<i>Up_2Prev</i>		
4	<i>Balanced<sub>C</sub></i>		<i>Balanced<sub>C</sub></i>		
5	<i>Prev_Down_St</i>		<i>Up_pix + 1</i>		
6	<i>3_Prev_Col</i>		<i>Sub_Prev_Col</i>		
7	<i>Up_2Prev</i>		<i>3_Prev_Col</i>		
8	<i>A_DC_Prev_Col</i>		<i>A_DC_Prev_Col</i>		
9	<i>DC<sub>C</sub></i>		<i>DC<sub>C</sub></i>		

TABLE 6.1 – Prédicteurs utilisés pour chacun des ordres de parcours de chacun des modes Intra 1D et chacun des ensembles de prédicteurs.

## 6.2.2 Analyses de l'efficacité des ordres de parcours

Avant de présenter les résultats expérimentaux de l'Intra 1D dans la configuration présentée ci-dessus, il est intéressant de proposer un test d'efficacité des différents ordres de parcours. En effet, les ordres de parcours n'ont pas la même mesure  $ANRS$ . Pour l'ordre de parcours lexicographique la mesure  $ANRS_L$  est systématiquement égale à 1, que ce soit pour le partitionnement Ligne ou Colonne (cf. section 6.1.1). Pour l'ordre bidirectionnel,  $ANRS_B$  est égale à 1,5. Cette moyenne pour l'ensemble du macrobloc reste inférieure à  $ANRS_{4 \times 4}$  qui est égale à 1,9. Pour l'ordre de parcours hiérarchique l' $ANRS_H$  est égale à 3. Ce chiffre est supérieur à l' $ANRS_{4 \times 4}$  mais reste inférieur à l' $ANRS_{16 \times 16}$ . Notons que l'ordre de parcours hiérarchique offre la possibilité pour  $N - 1$  partitions du macrobloc d'avoir comme signal de référence des lignes ou des colonnes qui entourent la partition, ce qui permet d'augmenter les redondances spatiales et par conséquent de réduire l'entropie. Même remarque pour les partitions  $Par_i^B$  de l'ordre hiérarchique.

Pour nous convaincre de l'efficacité d'utiliser plusieurs ordres de parcours nous avons mesuré l'entropie de chacun des modes Intra ayant 16 partitions (Intra  $4 \times 4$ , Intra Ligne, Intra Colonne) pour chaque ordre de parcours et chacune des partitions. Ces entropies sont reportées dans le tableau 6.2 pour la première image de la séquence *Tempete* encodée en profil Baseline avec un QP égale à 27. Pour chaque partition nous utilisons neuf prédicteurs. Chaque entropie est calculée pour l'ensemble des macroblocs. Pour les trois parcours lexicographique l'entropie de chaque partition est équivalente. De plus les entropies moyennes sont proches pour chacun des trois modes Intra.

Pour le parcours bidirectionnel du mode Ligne et du mode Colonne, les entropies des partitions  $Par_i^B$  sont inférieures aux entropies des  $Par_i^P$ . Les entropies de toutes les partitions  $Par_i^B$  sont proches comme les entropies des partitions  $Par_i^P$ . Notons que les partitions  $Par_i^B$  sont inférieures aux entropies des partitions des parcours lexicographiques alors que l'entropie des partitions  $Par_i^P$  leur est supérieure. Dans le cas des partitions  $Par_i^B$  le signal de référence entoure le signal à prédire. Les corrélations spatiales sont donc plus grandes, par conséquent l'entropie est plus faible que dans le cas du parcours lexicographique pour lequel le signal de référence sont les pixels du haut ou les pixels de gauche. Pour le cas des partitions  $Par_i^P$ , la distance moyenne entre un pixel à prédire et son pixel de référence étant systématiquement plus élevée que dans le cas du parcours lexicographique, l'entropie est par conséquent plus élevée.

L'ordre de parcours hiérarchique obtient une entropie moyenne supérieure aux entropies moyennes des deux ordres de parcours. Ceci était prévisible, car l' $ANRS_H$  de cet ordre est plus élevée que celle des deux autres ordres. Cependant la différence est faible car pour l'ordre hiérarchique, toutes les partitions, exceptée la partition  $Par_N$ , ont un signal de référence provenant de trois directions. Dans le tableau 6.2 nous pouvons voir, pour cet ordre de parcours, que plus une partition est élevée dans la hiérarchie plus son entropie est grande.

Enfin les entropies moyennes de chaque mode et de chaque ordre de parcours sont quasi équivalentes même si ces entropies moyennes sont ordonnées en fonction de leurs mesures  $ANRS$ . Ce n'est pas l'efficacité globale de chacun des modes ou des ordres de parcours qui engendre l'efficacité de la méthode que nous allons décrire dans les sections suivantes, mais l'efficacité conjointe liée à la compétition de ces possibilités de codage.

<b>Intra</b>	<b>4×4</b>	<b>Ligne</b>			<b>Colonne</b>		
<b>Partition</b>	“Lexico”	“Lexico”	Bidirectionnel	Hiérarchique	“Lexico”	Bidirectionnel	Hiérarchique
<i>Par</i> <sub>1</sub>	5.89	5.85	6.17	5.31	5.81	6.16	5.36
<i>Par</i> <sub>2</sub>	5.79	5.75	5.24	5.99	5.70	5.36	5.89
<i>Par</i> <sub>3</sub>	5.92	5.71	6.15	5.24	5.75	6.09	5.38
<i>Par</i> <sub>4</sub>	5.81	5.73	5.23	6.26	5.74	5.34	6.15
<i>Par</i> <sub>5</sub>	5.79	5.72	6.16	5.18	5.75	6.16	5.37
<i>Par</i> <sub>6</sub>	5.79	5.75	5.19	5.95	5.76	5.33	5.91
<i>Par</i> <sub>7</sub>	5.68	5.73	6.17	5.23	5.78	6.10	5.34
<i>Par</i> <sub>8</sub>	5.75	5.73	5.21	6.44	5.71	5.34	6.43
<i>Par</i> <sub>9</sub>	5.88	5.72	6.14	5.21	5.76	6.09	5.35
<i>Par</i> <sub>10</sub>	5.72	5.70	5.22	5.95	5.70	5.34	5.86
<i>Par</i> <sub>11</sub>	5.80	5.70	6.13	5.25	5.76	6.16	5.41
<i>Par</i> <sub>12</sub>	5.80	5.74	5.22	6.21	5.74	5.41	6.20
<i>Par</i> <sub>13</sub>	5.78	5.78	6.20	5.25	5.78	6.08	5.40
<i>Par</i> <sub>14</sub>	5.77	5.74	5.23	5.96	5.75	5.38	5.84
<i>Par</i> <sub>15</sub>	5.79	5.75	6.20	5.26	5.72	6.09	5.36
<i>Par</i> <sub>16</sub>	5.77	5.85	5.24	6.75	5.67	5.32	6.83
Moyenne	5.83	5.77	5.77	5.80	5.77	5.79	5.83

TABLE 6.2 – Entropie de chaque  $Par_i$  pour le mode Intra 4×4, le mode Ligne et ses trois ordres de parcours, le mode Colonne et ses trois ordres de parcours pour l’ensemble de la première image de la séquence *Tempete* CIF codée avec le profil Baseline pour le QP 27. Chaque mode et chacun de ces ordres de parcours a neuf prédicteurs.

## 6.2.3 Configuration Intra

Dans cette section nous donnons les résultats d'une configuration ne comportant que des slices Intra que nous noterons III. Cette configuration est utilisée pour le codage de séquences vidéo pour le cinéma numérique. Les résultats reportés concernent des séquences composées uniquement de la composante de luminance.

### 6.2.3.1 Gain sur le débit total

Le tableau 6.3 donne le pourcentage de gain sur le débit, calculé avec la métrique de Bjøntegaard [Bjø01a]. Nous avons testé trois configurations de la norme H.264/AVC : une configuration correspondant au profil Baseline III (Intra  $16 \times 16$ ,  $4 \times 4$ ), une correspondant au profil High III (Intra  $16 \times 16$ ,  $8 \times 8$  et  $4 \times 4$ ) et une configuration High III- $\{8 \times 8\}$  qui est la même que la précédente excepté que l'Intra  $8 \times 8$  est désactivé. Remarquons que ces trois configurations sont décodables par un décodeur standard. Les paramètres utilisés pour ces configurations sont décrits dans l'annexe A.

Le gain le plus faible est obtenu pour le profil High III avec 0,1% de pertes pour la séquence 720p *Cyclists* à bas débit et la séquence *Station2* à moyen débit. Le meilleur résultat est obtenu pour la séquence Schumacher avec 18,2% de gain à bas débit. Les gains sont en moyenne plus élevés pour le profil Baseline III que pour le profil High III qui obtient tout de même 3% de gain moyen à bas débit. Ceci est lié d'une part à l'efficacité de codage déjà engendrée par le profil High (le meilleur profil d'H.264/AVC) mais aussi par l'efficacité de l'Intra  $8 \times 8$ . En effet les résultats obtenus pour le profil High III- $\{8 \times 8\}$  (sans l'Intra  $8 \times 8$ ) sont plus élevés avec 5% de gain à moyen débit et 5,4% à bas débit. De plus nous avons paramétré le nombre et l'ordre des prédicteurs (cf. annexe B), pour chacun des partitionnements et des ordres de parcours, pour le profil Baseline. Le compromis est donc optimal pour ce dernier. Notre méthode obtient, pour ce profil, des gains élevés avec respectivement 5%, 6,8% et 8,2% pour les hauts, moyens et bas débits.

### 6.2.3.2 Sélection des Modes Intra

Le tableau 6.4 donne le pourcentage de sélection des modes Intra  $16 \times 16$ ,  $4 \times 4$  et 1D pour la méthode proposée et la référence pour le profil Baseline en fonction des QP. L'Intra 1D est plus souvent sélectionné que les modes de la référence H.264/AVC pour les quatre QP représentant les moyens débits. Le maximum de sélection de l'Intra 1D est atteint pour le QP 22 avec 60,6% de sélection. On note que par rapport à la référence, l'Intra 1D remplace davantage l'Intra  $4 \times 4$  que l'Intra  $16 \times 16$  car la baisse de sélection moyenne de l'Intra  $16 \times 16$ , pour la méthode proposée par rapport à la référence, est de 22% contre 59.1% pour l'Intra  $4 \times 4$ .

Pour le QP 17, l'Intra  $4 \times 4$  reste le plus sélectionné, comme dans la référence. Ceci est expliqué par plusieurs phénomènes. Premièrement, l'efficacité de codage de la transformée DCT  $4 \times 4$  à haut débit dans le logiciel de référence de la norme [JM] avec l'outil "Adaptive Rounding Offset" obtient son maximum de performance à haut débit [SLS05]. Nous n'avons pas implémenté cet outil pour la DCT 1D. Deuxièmement, nous utilisons la matrice de quantification définie pour la DCT  $4 \times 4$  pour la quantification des résiduels Intra 1D. Ce qui pénalise les modes Intra 1D notamment à haut débit où la précision de la quantification est plus importante. De meilleurs résultats devraient être

QP	Baseline III			High III			High- $\{8 \times 8\}$ III		
	17-32	22-37	27-42	17-32	22-37	27-42	17-32	22-37	27-42
	<b>QCIF</b>								
Container	4,2	5,9	7,8	3,2	4,4	5,1	3,2	4,5	5,6
Foreman	3,2	4,2	5,9	2,0	2,2	2,2	2,6	3,2	4,1
Silent	4,0	5,2	7,0	3,4	3,9	4,5	3,8	4,5	5,3
	<b>CIF</b>								
Foreman	3,2	4,1	5,5	2,1	2,3	2,3	2,5	2,9	3,4
Mobile	2,6	3,2	3,3	2,5	2,8	2,3	3,2	3,7	3,3
Paris	3,9	5,0	6,3	3,9	4,4	4,4	4,0	4,6	4,9
Tempete	3,1	3,9	4,8	2,8	3,0	2,4	3,3	4,2	4,4
Carphone	5,4	7,3	9,4	4,0	5,0	5,5	4,5	6,0	7,3
Crew	3,2	5,1	7,5	1,6	1,9	1,9	2,8	3,9	4,9
Ice	4,3	4,6	5,4	2,7	3,0	3,3	3,0	3,5	4,1
Schumacher	12,9	15,7	<b>18,2</b>	13,4	15,6	<b>17,5</b>	13,9	16,5	<b>18,7</b>
	<b>4CIF</b>								
City	8,0	9,4	11,2	3,5	3,4	3,0	8,0	8,3	8,1
Ice	6,5	6,7	6,7	3,2	3,8	3,6	3,6	4,5	4,7
Soccer	8,9	9,7	10,7	2,2	1,7	1,3	7,9	7,3	6,5
	<b>720p</b>								
Bigships	5,4	7,5	10,3	2,1	2,1	2,2	3,7	4,5	5,2
Shuttlestart	4,5	5,4	5,1	0,5	0,3	0,3	1,9	1,8	1,2
City	5,4	7,1	9,1	0,8	0,8	0,7	5,1	5,8	5,7
Crew	4,5	7,2	8,7	0,3	0,6	1,2	2,3	3,7	4,8
Night	4,9	6,4	7,5	2,6	2,7	2,4	4,8	5,4	5,2
Raven	3,0	3,0	2,1	0,1	0,0	0,0	1,3	0,8	0,1
Stockholm	4,7	6,9	9,1	3,0	3,6	3,6	3,9	5,2	6,0
Cyclists	6,2	9,0	9,9	1,0	0,6	-0,1	3,3	4,0	3,6
Jets	9,1	11,4	11,5	1,4	1,2	0,9	5,1	6,2	6,2
Panslow	4,6	8,8	12,1	4,0	7,4	8,4	4,6	7,9	10,9
Sailormen	7,1	9,1	11,1	3,9	3,9	3,1	5,9	7,1	7,4
Spincalendar	4,6	7,3	9,8	3,3	4,5	4,9	3,7	5,3	6,5
	<b>1080p</b>								
Blue sky	6,6	7,0	6,5	5,3	3,8	1,6	7,2	6,5	4,7
Intotree	1,7	3,1	6,5	0,2	0,4	1,3	0,4	0,9	2,1
Station2	5,2	7,2	7,2	0,3	-0,1	0,0	3,0	3,7	-3,2
Toys and calendar	4,5	8,1	9,7	1,6	1,9	1,0	3,2	4,8	4,6
<b>Moyenne</b>	<b>5,2</b>	<b>6,8</b>	<b>8,2</b>	<b>2,7</b>	<b>3,0</b>	<b>3,0</b>	<b>4,2</b>	<b>5,0</b>	<b>5,4</b>

TABLE 6.3 – Pourcentage de réduction de débit pour l'ensemble des séquences proposées incluant l'ensemble des séquences VCEG pour le profil Baseline III, High III et High III- $\{8 \times 8\}$  (sans Intra  $8 \times 8$ ) pour la configuration Intra.

QP	Référence H.264		Méthode Proposée		
	Intra16×16	Intra4×4	Intra16×16	Intra4×4	Intra1D
17	7,5	<b>92,5</b>	4,7	<b>57,9</b>	37,4
22	15,6	<b>84,4</b>	8,0	31,4	<b>60,6</b>
27	26,7	<b>73,3</b>	18,1	24,0	<b>57,9</b>
32	38,3	<b>61,7</b>	29,5	21,0	<b>49,5</b>
37	<b>50,9</b>	49,1	41,9	16,2	<b>41,9</b>
42	<b>67,8</b>	32,2	<b>59,5</b>	10,5	30,0
<b>Moyenne</b>	34,5	<b>65,5</b>	27,0	26,8	<b>46,2</b>

TABLE 6.4 – Pourcentage de sélection, en fonction des QP, des modes Intra 16×16 et 4×4 de la référence H.264/AVC et des modes Intra 16×16, 4×4 et 1D, pour la méthode proposée, codée avec le profil Baseline III, calculé pour l’ensemble des séquences.

obtenus en utilisant un vecteur de quantification adapté à cette transformée. Enfin, à haut débit, les résiduels de texture représentent la quasi totalité du débit (cf. chapitre 3). Le débit des indices des prédicteurs de l’Intra 4×4 a un faible impact sur le débit total. De plus, ce nombre élevé de prédicteurs engendre une meilleure prédiction du résiduel. Dans la référence, l’Intra 4×4 est toujours plus sélectionné pour les quatre premiers QP. Cette tendance s’inverse à partir du QP 37 car le coût des indices devient élevé en terme de débit-distorsion. Dans notre méthode seul l’ordre de parcours lexicographique pour le premier ensemble de prédicteurs a autant de prédicteurs que l’Intra 4×4, par conséquent les autres ordres de parcours ou les autres ensembles de prédicteurs sont défavorisés et sont donc moins sélectionnés. Des tests supplémentaires utilisant plus de prédicteurs pour le mode Intra 1D donnent des gains plus élevés à haut débit. Cependant, à moyen et bas débits, le nombre de prédicteurs utilisé dans ce chapitre reste plus efficace.

La sélection plus faible de l’Intra 1D pour le QP 42 s’explique par le coût plus élevé des indices relatifs à la compétition. En considérant le compromis débit-distorsion, les modes de codage à faible débit d’indice comme l’Intra 16×16 sont plus sélectionnés, même si d’autres modes obtiennent une meilleure entropie de résiduels de texture. Le débit lié à la signalisation des partitionnements et des ordres de parcours de l’Intra 1D étant élevé, le mode Intra 16×16 reste le plus sélectionné.

En moyenne pour l’ensemble des QP, l’Intra 1D est plus sélectionné que l’Intra 16×16 et 4×4 avec 46,2% de sélection. Ce pourcentage élevé de sélection provient du nombre important de nouvelles possibilités de codage engendrées par la méthode avec les deux partitionnements et les trois ordres de parcours proposés. Notons que cette sélection provient d’un choix optimisé en débit-distorsion. De plus nous avons délibérément choisi un codage ne défavorisant pas les modes Intra de la norme puisqu’un seul bit est ajouté à chaque macrobloc codé en Intra 4×4 en profil Baseline. Ce pourcentage élevé prouve donc l’efficacité de la méthode. Une réorganisation du codage des indices des modes Intra devrait encore augmenter l’efficacité de codage.

Le tableau 6.5 donne le pourcentage de sélection de chacun des modes Intra pour la référence et la méthode proposée pour le profil High III. La sélection de l’Intra 1D est en moyenne de 25,9% et atteint un maximum de sélection pour le QP 27 avec 36,4% de sélection. Pour ce QP, l’Intra 1D

QP	Référence H.264			Méthode Proposée			
	Intra16×16	Intra8×8	Intra4×4	Intra16×16	Intra8×8	Intra4×4	Intra1D
17	4,5	<b>51,8</b>	43,7	3,9	<b>47,0</b>	28,0	21,1
22	5,9	<b>58,1</b>	36,0	4,2	<b>43,7</b>	15,8	36,3
27	16,7	<b>49,6</b>	33,7	14,0	35,0	14,6	<b>36,4</b>
32	22,5	<b>51,5</b>	26,0	21,3	<b>37,1</b>	13,2	28,4
37	28,9	<b>52,3</b>	18,8	30,1	<b>38,2</b>	10,6	21,1
42	34,8	<b>55,8</b>	9,4	39,1	<b>42,7</b>	6,3	11,9
<b>Moyenne</b>	18,9	<b>53,2</b>	27,9	18,8	<b>40,6</b>	14,7	25,9

TABLE 6.5 – Pourcentage de sélection, en fonction des QP, des modes Intra 16×16, 8×8 et 4×4 de la référence H.264/AVC et des modes Intra 16×16, 8×8, 4×4 et 1D pour la méthode proposée, codée avec le profil High III, calculé pour l’ensemble des séquences.

est plus sélectionné que les autres modes Intra de la norme. Pour cette configuration, l’Intra 1D a remplacé plus fortement l’Intra 4×4 que l’Intra 8×8 car la baisse de sélection, entre la méthode et la référence, est de 47% pour l’Intra 4×4 contre 23,7% pour l’Intra 8×8. Ceci était prévisible car le nombre de pixels codés pour une partition Intra 1D est égal à celui de l’Intra 4×4. Il serait intéressant d’utiliser un partitionnement 1D Ligne sur des blocs de taille 64×16, comme proposé dans [IY09], ou un partitionnement 1D colonne sur des blocs de taille 16×64, afin que les partitions 1D aient autant de pixels que les partitions 8×8. Les partitionnements Ligne et Colonne, pour un macrobloc, sont plus efficaces que le partitionnement 4×4, a priori la méthode devrait se révéler plus efficace que le partitionnement 8×8 s’il le même nombre de pixels (64) est utilisé. De même, des partitionnements en L ou provenant d’une dilatation d’un pixel devraient concurrencer l’Intra 8×8, car ils engendrent des partitions comportant un nombre équivalent voir plus élevé de pixels.

### 6.2.3.3 Sélection des partitions 1D

Le tableau 6.6 donne le pourcentage de sélection moyen de chaque partitionnement et ordre de parcours en fonction des QP. Le mode Ligne est plus souvent sélectionné que le mode Colonne. Les ordres de parcours pour chacun des ces partitionnements ayant un nombre de prédicteurs équivalent, cette différence de sélection provient des redondances spatiales contenues dans les images des séquences utilisées. Lorsque l’on analyse la sélection des partitionnements, séquence par séquence, il apparaît que le nombre de sélections du partitionnement Colonne dépasse celui du partitionnement Ligne pour certaines d’entre elles.

Les sélections des ordres de parcours sont liées à leurs *ANRS*, car elles sont inversement ordonnées en fonction de ces mesures ( $ANRS_H > ANRS_B > ANRS_L$ ). Cependant l’ordre de parcours lexicographique regroupe deux ensembles de prédicteurs par partitionnement. Il est par conséquent avantage par rapport au deux autres ordres. Pour vérifier l’ordonnancement de la sélection des ordres de parcours, nous avons testé une configuration avec un seul ensemble de prédicteurs pour le parcours lexicographique. Les sélections des parcours restent inversement ordonnées en fonction de leurs *ANRS*.

QP	Ligne			Colonne			Moyenne
	L×2	B	H	L×2	B	H	
17	15,9	3,2	8,4	4,2	4,1	1,6	37,4
22	18,7	9,4	10,3	9,1	8,9	4,2	60,6
27	14,3	12,0	7,5	9,1	9,8	5,2	57,9
32	9,8	12,2	5,1	8,7	8,7	5,0	49,5
37	6,6	11,8	3,3	8,3	7,5	4,4	41,9
42	3,7	10,1	1,9	6,3	4,9	3,1	30,0
<b>Moyenne</b>	11,5	9,8	6,1	7,6	7,3	3,9	46,2

TABLE 6.6 – Pourcentage de sélection des ordres de parcours des partitionnements Ligne et Colonne en fonction des QP pour le profil Baseline III, calculé pour l’ensemble des séquences. Le parcours lexicographique est noté L. le symbole ×2 signifie que deux ensembles de prédicteurs sont utilisés pour ce parcours. Le parcours bidirectionnel est noté B et le parcours hiérarchique est noté H.

Prédicteurs	Référence H.264	Méthode Proposée
Vertical	25,6	<b>29,0</b>
Horizontal	<b>29,3</b>	26,4
DC	<b>25,7</b>	24,0
Plane	19,4	<b>20,6</b>

TABLE 6.7 – Pourcentage de sélection de chaque prédicteur Intra 16×16 pour la référence et pour la méthode proposée en profil Baseline III, calculé pour l’ensemble des séquences.

#### 6.2.3.4 Sélection des prédicteurs Intra

La nature des partitionnements Intra Ligne et Colonne devrait théoriquement engendrer une baisse de la sélection des prédicteurs “horizontal” et “vertical” des modes Intra 16×16 et 4×4. Le tableau 6.7 et le tableau 6.8 nous montrent respectivement le pourcentage de sélection de chaque prédicteur des modes Intra 16×16 et 4×4, pour la référence et pour la méthode proposée en profil Baseline. On peut remarquer que le prédicteur “horizontal” et le prédicteur “DC” ont une proportion de sélection moins élevée lorsque notre méthode est activée. On peut faire la remarque inverse, à savoir que le prédicteur “vertical” est proportionnellement plus sélectionné avec notre méthode. Ce résultat mitigé par rapport à la théorie peut être hypothétiquement lié à la nature des résultats présentés. En effet, ces résultats ne sont que des proportions par rapport à l’ensemble des modes. Lorsque l’Intra 1D est activé, le nombre de sélections des modes Intra 16×16 et 4×4 est inférieur à la référence.

Le tableau 6.9 donne le pourcentage de sélection de chaque prédicteur de l’ordre de parcours lexicographique pour les partitionnements Ligne et Colonne. Le prédicteur *Prev\_line* pour le partitionnement Ligne et le prédicteur *Prev\_Col* pour le partitionnement Colonne ont un pourcentage de sélection plus élevé que les autres prédicteurs, d’une part parce qu’ils appartiennent aux deux en-

Prédicteurs	Référence H.264	Méthode Proposée
Vertical	23,0	<b>30,1</b>
Horizontal	<b>24,6</b>	15,3
DC	<b>18,4</b>	14,0
Diagonaldown-left	5,1	<b>6,2</b>
Diagonaldown-right	6,5	<b>8,5</b>
Vertical-right	5,1	<b>6,4</b>
Horizontal-down	7,0	<b>8,2</b>
Vertical-left	4,5	<b>5,5</b>
Horizontal-up	<b>5,8</b>	<b>5,8</b>

TABLE 6.8 – Pourcentage de sélection de chaque prédicteur Intra  $4 \times 4$  pour la référence et pour la méthode proposée en profil Baseline III, calculé pour l’ensemble des séquences.

sembles de prédicteurs (deux ensembles de prédicteurs pour le parcours lexicographique), et d’autre part parce que la distance spatiale avec le signal à prédire est égale à 1. Il est à noter que lorsque l’on utilise un seul ensemble de prédicteurs pour l’ordre de parcours lexicographique, ces deux prédicteurs sont aussi les plus sélectionnés. Pour le mode Ligne, on notera que les redondances spatiales avec les pixels à gauche de la ligne courante sont importantes puisque l’addition de la sélection des prédicteurs *Left\_pix* et *Left\_pix + 1* atteint 16,4%. De plus, les prédicteurs provenant de moyennes entre la ligne précédente et des pixels à gauche (*Balanced<sub>L</sub>*, *Left\_2Prev*) obtiennent 19,6% de sélection ; soit 36% des prédicteurs sélectionnés exploitent totalement ou en partie les redondances spatiales provenant des pixels de gauche de la ligne courante pour le mode Ligne. Une remarque similaire peut être faite pour le partitionnement Colonne puisque le prédicteur *Up\_pix* représente 11,7% de sélection, les prédicteurs utilisant à la fois la colonne précédente et des pixels au dessus *Balanced<sub>C</sub>*, *Up\_2Prev* représentent 20,2% des sélections ; soit 31,9% des prédicteurs sélectionnés exploitent totalement ou en partie les redondances spatiales provenant des pixels en haut de la colonne courante pour le mode Colonne.

### 6.2.3.5 Sélection spatiale des modes Intra 1D

Les figures 6.6, 6.7, 6.8 représentent la sélection spatiale des modes Intra, notamment ceux des partitionnements et ordres de parcours de l’Intra 1D pour trois séquences. La séquence *Panslow 720p* est composée de plusieurs objets texturés sur un fond uni. Pour l’image 51 codée avec le profil Baseline QP 37, l’Intra 1D est sélectionné dans les parties texturées qui sont les plus coûteuses en terme de débit-distorsion pour la référence (cf. chapitre 3). Notre objectif de proposer un nouveau codage efficace pour les macroblocs ayant un coût élevé en débit-distorsion est donc atteint. En revanche, dans cette séquence, il est difficile de caractériser l’efficacité de l’un des deux partitionnements testés en fonction du type de texture. Dans la figure 6.7, représentant la sélection des modes Intra pour l’image 22 de la séquence *Stockholm 720p* codée avec profil Baseline pour le QP 22, on remarque que le partitionnement Ligne est davantage sélectionné que le partitionnement Colonne. Ceci s’explique par le contenu de la séquence qui est composée de nombreux motifs en ligne, comme la route en

Prédicteurs	Ligne “Lexico”	Prédicteurs	Colonne “Lexico”
<i>Prev_line</i>	23,5	<i>Prev_Col</i>	22,5
<i>Left_pix</i>	10,0	<i>Up_pix</i>	11,7
<i>Balanced<sub>L</sub></i>	11,3	<i>Prev_Up_St</i>	12,9
<i>Left_2Prev</i>	8,3	<i>Balanced<sub>C</sub></i>	11,6
<i>Prev_Right_S</i>	11,1	<i>Prev_Down_St</i>	11,1
<i>Left_pix + 1</i>	6,4	<i>3_Prev_Col</i>	8,6
<i>A_DC_Prev_line</i>	10,4	<i>Up_2Prev</i>	7,6
<i>Sub_Prev_line</i>	9,6	<i>A_DC_Prev_Col</i>	7,3
<i>Prev_Left_St</i>	9,4	<i>DC<sub>C</sub></i>	6,7

TABLE 6.9 – Pourcentage de sélections de chaque prédicteur Intra Ligne et Intra Colonne de l’ordre de parcours lexicographique pour le profil Baseline III, calculé pour l’ensemble des séquences.

premier plan ou les façades des immeubles, les vagues de la rivière et les toits des immeubles au loin. On remarque que le mode Colonne a été choisi pour une partie des macroblochs à cheval sur le bord de l’image et la colonne noire à gauche. La figure 6.8 représente la sélection spatiale de l’image 91 de la séquence Harbour 4CIF codée avec le profil Baseline pour le QP 32. Dans cette figure, le mode Colonne est largement sélectionné sur les mâts des bateaux qui sont des motifs colonnes. On notera que le partitionnement Ligne est sélectionné en majorité sur les macroblochs contenant uniquement les vagues de la mer. Après analyse des ces trois images, on peut conclure que la sélection des partitionnements est liée aux motifs contenus dans les séquences : le partitionnement Ligne est sélectionné pour les macroblochs contenant des textures composées de lignes et le partitionnement Colonne pour les macroblochs contenant des textures composées de colonnes. Enfin, aucune corrélation entre les motifs contenus dans les séquences et les ordres de parcours n’apparaît visuellement.

### 6.2.3.6 Coût du codage de la compétition Intra 1D

L’ajout de nouveaux modes de codage par compétition engendre la transmission de cette information de compétition du codeur au décodeur. Le tableau 6.10 donne la proportion de débit nécessaire au codage des modes Intra et des prédicteurs Intra 4×4 pour la référence et lorsque la méthode Intra 1D est activée en fonction des QP pour le profil Baseline. Comme on a pu le voir dans le chapitre 3, la proportion dans le débit total des prédicteurs Intra augmente avec le QP. En moyenne la proportion du coût de la compétition de tous les modes Intra et de leurs prédicteurs est équivalente à celle de la référence. C’est-à-dire que la méthode Intra 1D n’augmente pas la proportion de débit liée au codage des modes Intra.

Le tableau 6.11 donne les mêmes informations que le tableau 6.10 mais pour la configuration High III. On peut faire la même remarque sur l’évolution de la proportion en fonction des QP pour les trois modes, excepté pour le QP 42 pour les modes Intra 1D et Intra 4×4. Ceci s’explique par le faible nombre de sélections de ces modes pour ce QP (cf. tableau 6.5). En revanche la proportion de l’information de la compétition des modes et de leurs prédicteurs est différente entre notre méthode et la référence, passant de 15,1% à 17,5%. Nous avons ajouté dans ce tableau le nombre moyen

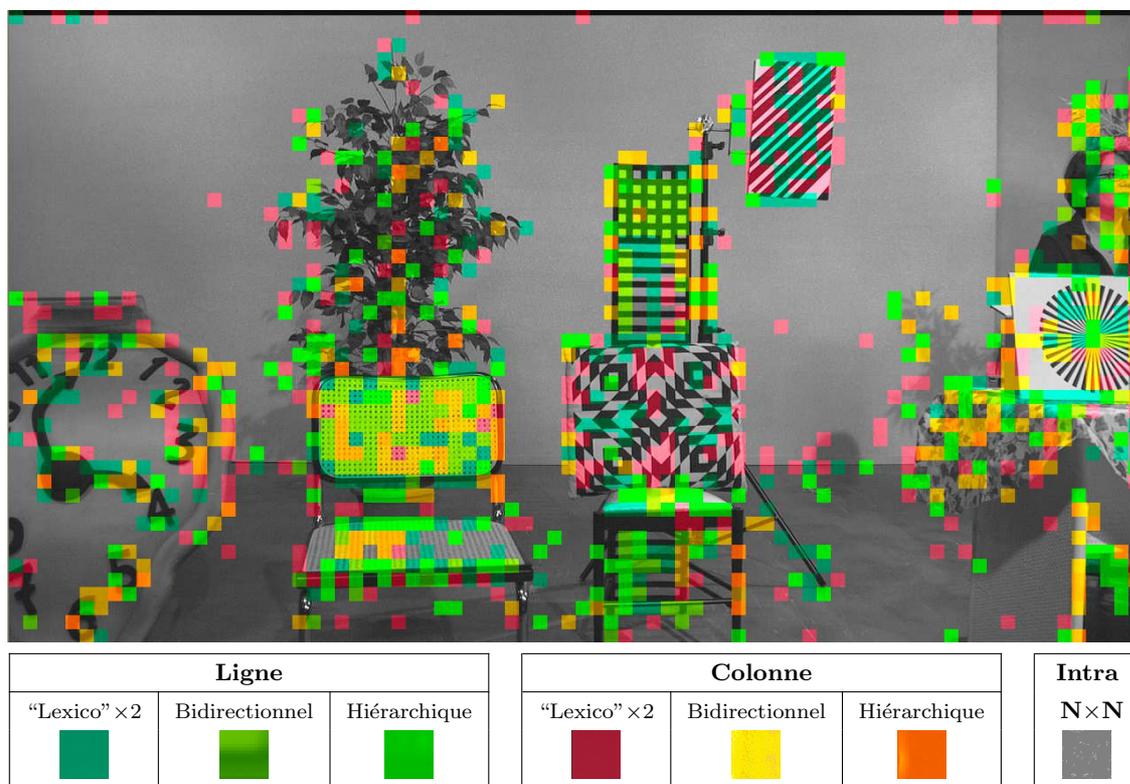


FIGURE 6.6 – Représentation spatiale de la sélection des modes Intra pour l'image 51 de la séquence *Panslow* 720p codée avec le profil Baseline III à QP 37 pour la méthode proposée.

QP	Référence H.264		Méthode Proposée			
	Mode	Prédicteurs	Mode	Mode 1D	Prédicteurs	
		Intra 4×4			Intra 4×4	Intra 1D
17	0,4	9,4	0,4	0,5	7,7	3,0
22	0,9	12,6	0,6	1,2	6,3	6,1
27	1,7	16,6	1,4	1,9	5,2	8,2
32	3,4	21,2	3,0	2,7	6,2	10,3
37	6,3	25,0	6,0	3,8	8,5	12,4
42	13,0	25,3	13,0	4,5	9,9	12,9
<b>Moyenne</b>	4,3	18,3	4,1	2,5	9,9	8,8
$\Sigma$	22,6		23,0			

TABLE 6.10 – Proportion de débit des indices de modes et des indices de prédicteurs dans le débit total pour la référence et la méthode proposée pour le profil Baseline III.

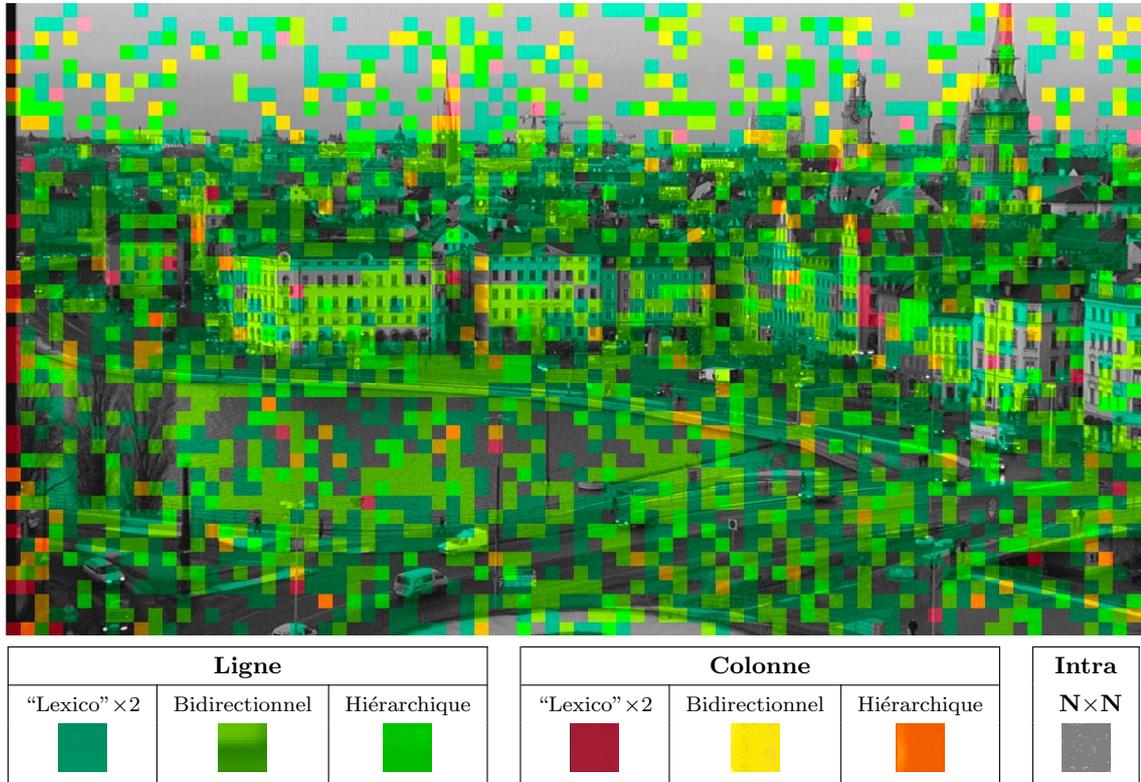


FIGURE 6.7 – Représentation spatiale de la sélection des modes Intra pour l’image 2 de la séquence *Stockholm* 720p codée avec le profil Baseline III à QP 22 pour la méthode proposée.



FIGURE 6.8 – Représentation spatiale de la sélection des modes Intra pour l’image 91 de la séquence *Harbour* 4CIF codée avec le profil Baseline III à QP 32 pour la méthode proposée.

QP	Référence H.264			Méthode Proposée				
	Mode	Prédicteurs		Mode	Mode 1D	Prédicteurs		
		Intra 8×8	Intra 4×4			Intra 8×8	Intra 4×4	Intra 1D
		<b>8,4bpm</b>	<b>30,6bpm</b>			<b>8,8 bpm</b>	<b>31,1 bpm</b>	<b>20,1 bpm</b>
17	0,2	1,7	4,1	0,2	0,3	1,6	2,8	1,8
22	0,4	2,7	4,9	0,5	0,7	2,3	2,5	3,7
27	1,2	3,8	6,7	1,2	1,1	3,1	3,4	4,9
32	2,4	6,1	7,8	2,6	1,6	5,0	4,5	5,2
37	4,5	9,0	7,8	5,0	2,2	7,4	4,8	5,3
42	9,0	13,1	5,5	10,3	2,3	11,1	3,8	4,2
<b>Moyenne</b>	3,0	6,1	6,1	3,3	1,4	5,1	3,6	4,2
$\Sigma$		15,1			17,5			

TABLE 6.11 – Proportion de débit des indices de modes et des indices de prédicteurs dans le débit total pour la référence et la méthode proposée pour le profil High III. Le nombre de bits par macrobloc (*bpm*), relatif au codage des indices de prédicteurs, par mode Intra est indiqué en gras.

de bits par macrobloc nécessaire au codage des prédicteurs engendrés par chacun des modes Intra. Concernant la référence, ce nombre de bits moyen pour l’Intra 4×4 est plus de trois fois supérieur à celui des prédicteurs de l’Intra 8×8. On remarque que le nombre de bits moyen par macrobloc, pour les modes Intra 1D, se situe entre ceux des deux modes de la norme. Ceci explique en partie le gain plus faible obtenu avec la méthode pour le profil High III. De plus la sélection du nombre de prédicteurs par mode Intra 1D a été effectuée avec le profil Baseline sur un ensemble de séquences CIF. Il serait donc intéressant de refaire une étude équivalente pour le profil High afin de déterminer le nombre optimal de prédicteurs ou encore mieux d’utiliser une méthode de sélection automatique du nombre de prédicteurs en fonction des caractéristiques de ces prédicteurs. Nous proposerons dans le chapitre 8 une telle méthode.

### 6.2.3.7 Intra 1D par rapport à JPEG 2000

Dans la partie I on a vu que le profil High III de la norme H.264/AVC est plus efficace que JPEG 2000 pour le codage d’images fixes. Le profil Baseline III de la norme H.264/AVC obtient une meilleure efficacité de codage que JPEG 2000 en moyenne sur l’ensemble des séquences VCEG, mais JPEG 2000 est plus performant pour les séquences 720p de l’ensemble testé (cf. tableau 3.5 du chapitre 3).

Dans le tableau 6.12 nous reportons les gains en pourcentage de réduction de débit du profil Baseline III de la norme H.264/AVC et ceux obtenus avec notre méthode (Baseline III Intra 1D) par rapport à la norme JPEG 2000. Nous donnons ces résultats en moyenne pour les séquences QCIF et CIF et en détail pour les séquences 720p. Rappelons qu’à basse résolution, H.264/AVC Baseline III est plus efficace que JPEG 2000. Notre méthode étant plus efficace que ce profil, nous n’avons donc pas reporté ces résultats. De même, notre méthode est plus efficace que le profil High III de la

	Baseline III			Baseline III Intra1D		
	17-32	22-37	27-42	17-32	22-37	27-42
<b>QCIF</b>	13,0	17,9	23,9	16,6	22,4	29,5
<b>CIF</b>	6,6	8,2	9,7	9,7	11,9	14,3
Bigships	-3,0	-4,0	-7,0	2,5	3,7	3,6
Shuttlestart	-11,2	-12,3	-13,5	-6,4	-6,2	-7,4
City_corr	-3,3	-4,8	-9,0	2,2	2,5	0,4
Crew	-4,4	-6,8	-10,2	0,1	0,7	-0,7
Night	0,1	0,3	-0,6	5,0	6,6	6,8
<b>Moyenne 720p</b>	-4,4	-5,5	-8,1	0,7	1,4	0,5
<b>Moyenne</b>	<b>3,6</b>	<b>4,9</b>	<b>5,9</b>	<b>7,6</b>	<b>10,2</b>	<b>12,3</b>

TABLE 6.12 – Pourcentage de réduction de débit de la configuration Baseline III et Baseline III Intra 1D par rapport à JPEG 2000 pour l'ensemble des séquences VCEG.

norme et ce dernier est systématiquement plus efficace que JPEG 2000. Ces résultats ne seront donc pas reportés. Dans ce tableau 6.12, le pourcentage de gain moyen sur les séquences VCEG 720p, obtenu avec notre méthode, est positif (contrairement à la référence Baseline III). JPEG 2000 est cependant plus efficace que notre méthode pour une séquence (*Shuttlestart part2 720p*).

La figure 6.9 représente les courbes débit-distorsion de la séquence Schumacher CIF 30Hz pour les cinq codages étudiés. On voit sur ces courbes que le codage High III Intra 1D de cette séquence est le plus efficace, mais surtout que le profil Baseline III Intra 1D est plus efficace que les profils Baseline III et High III de la norme H.264/AVC.

#### 6.2.4 Configuration Broadcast TV

Notre méthode offre des gains significatifs sur les slices Intra (cf. tableau 6.3). Cependant, un flux vidéo, dans la plupart des applications est composé de slices Inter. Nous proposons donc, dans cette section, de calculer les gains de l'Intra 1D dans une configuration GOP 12 (cf. annexe A) correspondant à une application de transmission de flux TV. Dans cette configuration une image Intra est codée pour 11 images Inter. Pour le Baseline GOP 12 une image Intra est codée pour 11 images P (11 P pour le Baseline IPP, 3 P et 8 B pour le High IBBP). Le tableau 6.13 donne le pourcentage de réduction de débit de la méthode proposée pour les configurations Baseline GOP 12 IPP et High GOP 12 IBBP pour l'ensemble des séquences utilisées incluant l'ensemble des séquences VCEG. Les gains sont naturellement moins élevés que pour la configuration cinéma, cependant ils restent positifs puisque nous obtenons 4% de gain à bas débit pour le profil Baseline et 2,4% pour le profil High. Dans le chapitre 3 nous avons vu que dans une configuration High GOP 12 IBBP, le débit des slices Intra représente entre 30% et 50% du débit total en fonction du QP utilisé. Les gains avec notre méthode sont de 3,0% à bas débit pour les slices I et le gain pour la configuration GOP 12 est de 2,4%. On note alors que l'amélioration apportée par l'Intra 1D se répercute sur les autres images P et B. En effet, comme nous l'avons expliqué dans le chapitre 3, pour les recommandations de test,

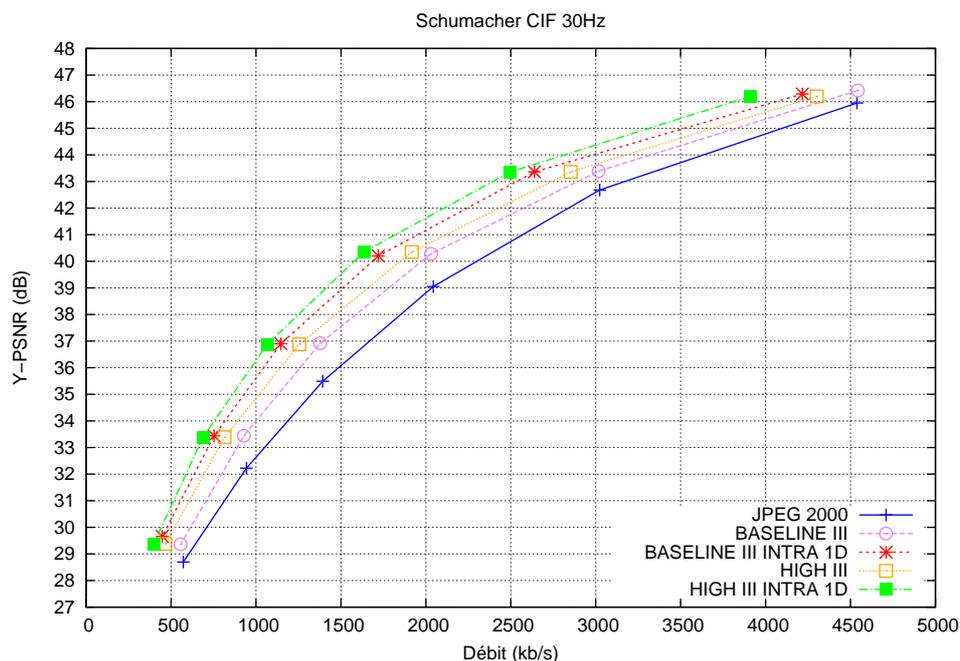


FIGURE 6.9 – Courbes débit-distorsion pour la référence JPEG 2000, H.264/AVC Baseline III et High III et la méthode proposée en Baseline III et High III pour la séquence Schumacher CIF 30Hz.

la distorsion des images I (lorsqu'il n'y a pas de changement de scène) se propage sur l'ensemble de la séquence.

### 6.3 Conclusion et perspectives

Dans ce chapitre nous avons présenté une méthode de codage Intra. Cette méthode est basée sur un partitionnement 1D du macrobloc qui offre une distance spatiale plus faible entre les pixels à prédire et le signal de référence que le partitionnement carré des modes Intra de la norme H.264/AVC. Le partitionnement 1D est obtenu par dilatation morphologique de la partition initiale. Afin d'augmenter l'efficacité de ce partitionnement trois ordres de parcours du partitionnement sont utilisés : hiérarchique, bidirectionnel et lexicographique. Les résiduels de texture de ces nouveaux partitionnements sont ensuite codés avec une transformée DCT 1D.

Les modes Intra 1D testés (Ligne et Colonne) sont plus sélectionnés que les modes Intra  $4 \times 4$  et  $16 \times 16$  de la norme pour le profil Baseline. Le gain obtenu est de 3% pour le profil High et 8,2% à bas débit pour le profil Baseline et atteint un maximum de 18,2% pour une séquence. Ces gains élevés pour les images Intra se reportent dans une configuration de la norme exploitant aussi les redondances temporelles puisque pour la configuration GOP 12 notre méthode atteint 4,6% de gains moyens.

L'efficacité de la méthode peut être cependant améliorée. En effet, le nombre de prédicteurs

QP	Baseline GOP 12 IPP			High GOP 12 IBBP		
	17-32	22-37	27-42	17-32	22-37	27-42
	<b>QCIF</b>					
Container	3,2	4,8	6,2	2,8	4,3	5,2
Foreman	1,2	1,7	2,2	1,4	1,5	1,3
Silent	3,2	4,3	5,2	2,6	3,0	3,4
	<b>CIF</b>					
Foreman	1,4	2,1	3,4	1,4	1,8	2,2
Mobile	0,8	1,3	1,7	1,1	1,4	1,4
Paris	2,5	3,3	4,7	2,9	3,8	4,5
Tempete	0,9	1,3	2,1	1,1	1,4	1,4
Carphone	2,3	3,3	4,6	2,3	3,2	4,1
Crew	1,7	2,4	2,9	0,7	1,1	1,6
Ice	1,3	1,6	1,9	1,1	1,4	1,5
Schumacher	5,9	7,6	9,3	5,3	7,1	9,1
	<b>4CIF</b>					
City	2,6	3,8	5,6	1,9	2,2	2,3
Ice	2,5	2,4	2,6	1,8	2,0	2,1
Soccer	3,2	4,0	4,8	1,4	1,0	0,9
	<b>720p</b>					
Bigships	2,7	4,4	6,9	1,2	1,5	1,8
Shuttlestart	2,3	2,4	2,2	1,0	0,6	0,3
City_corr	2,3	3,9	5,6	1,1	1,5	1,6
Crew	3,1	3,9	4,1	0,7	1,0	1,1
Night	1,9	2,8	3,6	1,5	2,0	2,0
Raven	1,3	1,0	0,8	0,2	0,1	0,0
Stockholm	1,9	4,7	6,9	2,1	3,4	3,5
Cyclists	2,3	3,9	4,8	0,8	0,4	0,1
Jets	7,3	8,4	8,1	2,2	1,7	1,3
Panslow	2,2	6,9	9,2	1,6	6,6	8,2
Sailormen	2,8	4,3	6,0	1,9	2,1	2,1
Spincalendar	2,1	4,8	7,2	2,1	4,1	4,8
	<b>1080p</b>					
Blue_sky	2,7	2,6	2,4	2,8	1,9	0,9
Intotree	0,7	2,0	4,2	0,4	0,4	1,2
Station2	3,3	4,5	4,4	0,4	0,3	0,3
Toys and calendar	2,5	4,6	5,1	1,2	1,5	0,7
<b>Moyenne</b>	<b>2,5</b>	<b>3,6</b>	<b>4,6</b>	<b>1,6</b>	<b>2,1</b>	<b>2,4</b>

TABLE 6.13 – Pourcentage de réduction de débit pour l'ensemble des séquences proposées incluant l'ensemble des séquences VCEG, pour les profils Baseline GOP 12 IPP et High GOP 12 IBBP.

pour chaque mode Intra 1D a été fixé dans une configuration incluant l’Intra  $4 \times 4$  et l’Intra  $16 \times 16$  ; il serait intéressant de fixer un nombre de prédicteurs optimal pour la compétition avec ces deux modes et l’Intra  $8 \times 8$  pour améliorer l’efficacité de la méthode pour le profil High. La quantification de la transformée 1D est déterminée à partir de la matrice de quantification de la DCT  $4 \times 4$ . Une recherche d’une quantification adaptée à nos partitionnements devrait être bénéfique. On peut aussi envisager de remplacer cette étape par une “transformée sans transformée” comme dans l’outil APEC [NM06] intégré dans le KTA. Une autre voie de recherche intéressante est de fixer des QP différents pour chaque niveau de hiérarchie pour l’ordre de parcours hiérarchique, comme pour les images B hiérarchiques. De même des QP différents pour les partitions  $Par_i^P$  et  $Par_i^B$  devraient aussi améliorer l’efficacité de codage de la méthode. Enfin, une dernière perspective qui devrait engendrer des gains est la mise en compétition de partitionnements 1D autres que les deux actuellement testés et de mettre tous ces partitionnements en compétition.

Les gains obtenus par l’Intra 1D offrent une perspective réaliste d’être intégré dans une éventuelle prochaine norme. A ce jour, ces travaux ont donné lieu à un stage de master recherche afin de tester plusieurs des perspectives décrites précédemment [Thi08].

## Troisième partie

# Que doit-on vraiment transmettre ?



## Chapitre 7

# Suppression d'indices de prédictions pour la méthode MVComp

*Dans le chapitre 4, nous avons proposé une méthode de codage des vecteurs mouvement basée sur une compétition exhaustive des prédictions de ces vecteurs. Cette méthode introduit une nouvelle information de compétition : l'indice du prédicteur sélectionné. Le coût de cette information atteint 3,6% du débit total à bas débit et 15% du nouveau débit de l'information de mouvement. Nous éliminons une partie de cette information en considérant l'égalité des prédicteurs mis en compétition. En effet, si tous les prédicteurs sont égaux, l'indice de prédiction n'est pas transmis. Cependant, il est possible d'éliminer encore d'autres indices, comme nous allons le voir.*

*Dans ce chapitre, nous proposons une méthode de suppression d'indices de prédicteurs de vecteurs mouvement basée sur le critère de choix de la méthode de compétition. Pour chaque compétition de prédicteurs de vecteur mouvement, l'encodeur détermine les indices qui peuvent ne pas être transmis. Le décodeur reproduit ensuite la même opération afin de déterminer les prédicteurs "implicites".*

### 7.1 Description de la méthode

La méthode présentée ci-dessous permet de deviner certains prédicteurs de la compétition de vecteurs mouvement. Pour cela, il faut que le décodeur soit capable de reproduire le processus de sélection du prédicteur du vecteur mouvement réalisé par l'encodeur. L'encodeur et le décodeur utilisent alors le même critère pour discerner les prédicteurs à éliminer.

Nous expliquerons tout d'abord le procédé d'élimination de l'indice lorsque deux prédicteurs sont en compétition, puis le problème sera étendu au cas d'ensembles composés de  $N$  prédicteurs.

---

1. Publications de l'auteur associées : [RevueFR1], [Brev9] et [Brev8].

### 7.1.1 Deux prédicteurs

Dans le chapitre 4, nous avons noté  $mv$  le vecteur mouvement courant et  $\varepsilon_{mv_i}$  le résiduel du vecteur  $mv$  associé au prédicteur  $p_i$ . Le critère de sélection d'un prédicteur d'un ensemble de  $N$  prédicteurs est donné dans l'équation (4.13) du chapitre 4, défini par  $R_{mv/mm} = \min\{\zeta(\varepsilon_{mv_i}) + \zeta(i)\}_{i \in \{1, \dots, N\}}$ . Dans cette équation,  $R_{mv/mm}$  représente le débit des vecteurs mouvement et de l'indice du prédicteur utilisé et  $\zeta(x)$  est le coût en débit réel ou estimé de l'information  $x$ .

Dans la méthode de codage des vecteurs mouvement par compétition de prédictions, pour chaque vecteur, le décodeur calcule les prédicteurs mis en compétition au codeur. Le décodeur teste l'égalité des prédicteurs. Si les prédicteurs ne sont pas égaux, l'indice du prédicteur utilisé est lu dans le flux et le vecteur  $mv$  est obtenu par l'addition du résiduel  $\varepsilon_{mv}$  et du prédicteur correspondant à l'indice décodé. Considérons maintenant que le décodeur connaît le critère de choix utilisé pour le codage.

Le codeur a déterminé le meilleur prédicteur de la manière suivante :

Si

$$\min\{\zeta(\varepsilon_{mv_1}) + \zeta(1), \zeta(\varepsilon_{mv_2}) + \zeta(2)\} = \zeta(\varepsilon_{mv_1}) + \zeta(1) \quad (7.1)$$

alors le prédicteur  $p_1$  est sélectionné, sinon

$$\min\{\zeta(\varepsilon_{mv_1}) + \zeta(1), \zeta(\varepsilon_{mv_2}) + \zeta(2)\} = \zeta(\varepsilon_{mv_2}) + \zeta(2) \quad (7.2)$$

et par conséquent,  $p_2$  est le prédicteur sélectionné.

Notons  $mv_1 = \varepsilon_{mv} + p_1$  le vecteur mouvement décodé avec le prédicteur  $p_1$ , et  $mv_2 = \varepsilon_{mv} + p_2$  le vecteur mouvement décodé avec le prédicteur  $p_2$ . Le vecteur original est soit  $mv_1$  soit  $mv_2$ . Si  $p_1$  a été choisi alors :

$$\min\{\zeta(mv_1 - p_1) + \zeta(1), \zeta(mv_1 - p_2) + \zeta(2)\} = \zeta(mv_1 - p_1) + \zeta(1) \quad (7.3)$$

Si l'on ajoute à cette condition que :

$$\min\{\zeta(mv_2 - p_1) + \zeta(1), \zeta(mv_2 - p_2) + \zeta(2)\} = \zeta(mv_2 - p_1) + \zeta(1) \quad (7.4)$$

alors il est certain que le prédicteur  $p_1$  est le prédicteur qui a été choisi par le codeur, car il est sélectionné pour les deux vecteurs possibles,  $mv_1$  et  $mv_2$ . Le décodeur peut effectuer le même processus. Il n'est donc pas nécessaire de transmettre l'indice du prédicteur si l'équation (7.3) et l'équation (7.4) sont vérifiées.

### 7.1.2 $N$ prédicteurs

Généralisons notre critère au cas où l'ensemble des prédicteurs de vecteurs mouvement contient  $N$  prédicteurs.  $mv_i = \varepsilon_{mv} + p_i$  est le vecteur mouvement décodé avec le prédicteur  $p_i$ . Si l'indice du prédicteur n'est pas transmis, il existe alors  $N$  vecteurs possibles. Le prédicteur  $p_i$  est le prédicteur utilisé et il n'est pas nécessaire de transmettre son indice  $i$ , si et seulement si :



courant dans le flux, on teste l'équation (7.5). Si l'on obtient un prédicteur implicite, on n'extrait pas du flux l'indice du prédicteur, mais on calcule donc directement la prédiction inverse avec ce prédicteur implicite.

Au niveau de l'encodeur, deux implémentations sont possibles. Dans la première, on peut appliquer l'algorithme uniquement au moment de l'écriture de l'indice du prédicteur dans le flux. Cette implémentation fournira exactement la même sélection des prédicteurs de vecteurs mouvement et les mêmes modes de codage que ceux obtenus avec la méthode MVComp pour une configuration identique. La seule différence est le nombre d'indices transmis au décodeur qui est au maximum égal à celui de la méthode MVComp. Par conséquent, les distorsions engendrées par les deux méthodes sont égales, seul le débit est différent. Dans cette implémentation, aucune perte, en termes de débit-distorsion, ne peut être observée par rapport à MVComp. La deuxième implémentation consiste à appliquer la méthode à chaque calcul d'un résiduel de vecteur, notamment pendant la phase de sélection du meilleur vecteur mouvement pour un bloc. Dans ce cas, le critère de suppression influe donc sur les choix de codage. Théoriquement, cette deuxième implémentation est plus efficace car pour chaque bloc le meilleur choix de codage est sélectionné. Ceci se vérifie en pratique. Nous présentons donc les résultats avec cette seconde implémentation.

## 7.2.2 Réduction de débit des indices de prédicteurs

La figure 7.1 donne la proportion des indices de prédicteurs de vecteurs mouvement qui sont transmis et ceux qui sont éliminés par la méthode proposée et par le test d'égalité des prédicteurs. On remarque que le pourcentage d'indice non transmis représente plus d'un quart des indices atteignant près de 30% pour le QP 17. Nous noterons que l'égalité des prédicteurs n'est qu'un cas particulier du critère défini dans ce chapitre. La réduction de débit des indices de prédicteurs dépasse 10% pour les QP recommandés par le groupe VCEG dans la configuration Baseline IPP et 8% pour la configuration High IPP. Cette réduction est plus importante à haut débit avec 13,5% de réduction pour le QP 17 contre 8,7% pour le QP 42. Cependant, le débit des indices de prédicteurs représente 15% du débit de l'information de mouvement et seulement 2,6% du débit total, comme on a pu le voir dans le tableau 4.10 du chapitre 4. Par conséquent, l'amélioration d'efficacité de codage est peu élevée avec respectivement 0,2%, 0,3%, 0,3% de réduction de débit par rapport à MVComp pour les ensembles de QP représentant les hauts, moyens et bas débits du profil Baseline IPP.

## 7.2.3 Analyse

Le tableau 7.1 donne la proportion de sélection du prédicteur médian et du prédicteur collocated, ainsi que leur égalité en fonction des QP lorsque la méthode de suppression des prédicteurs décrite dans ce chapitre est activée. Ces résultats sont à comparer avec ceux de la figure 4.8 du chapitre 4. On remarque que la proportion de sélection diffère par rapport à la méthode MVComp. En effet, pour MVComp le collocated était le prédicteur le plus sélectionné pour les trois derniers QP. Avec l'ajout de cette méthode, le prédicteur temporel est maintenant le plus sélectionné pour les quatre derniers QP. Cette proportion de sélection du prédicteur collocated augmente pour tous les QP. En moyenne, cette sélection est de 45,4% contre 43,7% dans la référence. L'impact de la méthode

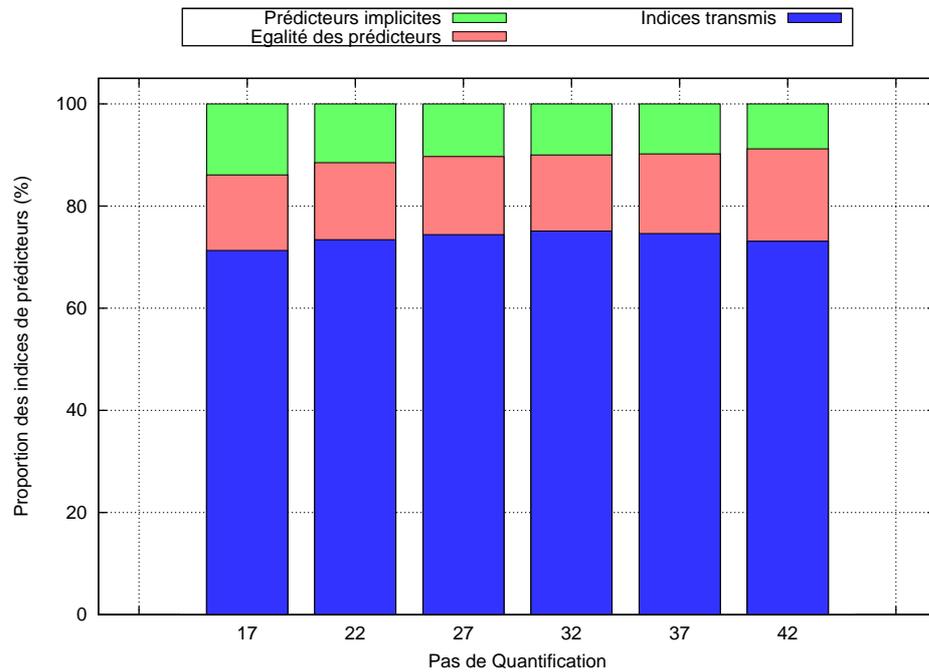


FIGURE 7.1 – Proportion des indices de prédicteurs des vecteurs mouvement transmis et non transmis grâce au test d'égalité et ceux éliminés par le critère des prédicteurs implicites, pour la configuration Baseline IPP, en fonction des six QP utilisés.

QP	Prédicteur spatial <i>mv<sub>H.264</sub></i>	Prédicteur temporel <i>mv<sub>Scol<sub>R0</sub></sub></i>	Egalité des prédicteurs (même valeur)
<b>17</b>	44,7	40,2	15,1
<b>22</b>	44,1	40,8	15,1
<b>27</b>	42,2	42,5	15,3
<b>32</b>	39,1	46,0	14,9
<b>37</b>	34,4	50,0	15,6
<b>42</b>	28,6	53,3	18,1
<b>Moyenne</b>	<b>38,9</b>	<b>45,4</b>	<b>15,7</b>

TABLE 7.1 – Proportion de sélection des prédicteurs  $mv_{Scol_{R_0}}$  et  $mv_{H.264}$ , et de leur égalité de l'ensemble  $\mathcal{P}$  lorsque la méthode proposée est activée, de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences et des six QP utilisés.

est relativement faible sur la sélection quantitative et spatiale des modes de codage. De même, la proportion de la sélection des blocs prédicteurs dans les images de référence n'a pas évolué par rapport à la méthode MVComp. On peut considérer que l'augmentation de la proportion de prédicteurs collocated est probablement liée à l'évolution de la sélection du meilleur vecteur mouvement. En effet, cette sélection peut être modifiée par l'introduction de la méthode car le critère de suppression est calculé pour chaque estimation du meilleur prédicteur, et donc de chaque estimation du débit de l'information de mouvement. On notera que la proportion moyenne d'égalité des prédicteurs a, elle aussi, augmenté, passant de 14,4% (cf. tableau 4.12 du chapitre 4) à 15,7%.

Il est aussi intéressant de constater que l'efficacité de l'algorithme est corrélée avec les images de référence. Le tableau 7.2 donne le pourcentage d'élimination des indices supprimés par le test d'égalité et ceux supprimés par la méthode proposée, en fonction des images de référence. On remarque que le critère d'élimination des prédicteurs implicites évince une plus grande proportion de prédicteurs lorsque le prédicteur de bloc est sélectionné sur la première image de référence. La proportion de suppression ne cesse de diminuer lorsque la distance temporelle entre le bloc courant et le bloc prédicteur augmente. Pour la première image de référence, 30,2% des indices ont été éliminés contre 17,1% pour l'image de référence *Ref<sub>3</sub>*.

#### 7.2.4 Elimination et fonctions de choix

Les pourcentages d'élimination d'indices de prédicteurs obtenus avec la méthode sont naturellement liés au critère de choix du meilleur prédicteur défini dans l'équation (4.13) du chapitre 4. En effet, si on remplace dans cette équation la fonction d'estimation de coût  $\zeta(x)$  par une fonction avec des caractéristiques différentes, ces pourcentages d'élimination évoluent. Pour nous convaincre

Images de référence	Egalité des prédicteurs (même valeur)	Prédicteurs implicites	Total indices non transmis
$Ref_0$	17,8	12,4	30,2
$Ref_1$	11,8	7,9	19,7
$Ref_2$	11,2	7,3	18,5
$Ref_3$	10,8	6,3	17,1

TABLE 7.2 – Proportion de suppression des indices des prédicteurs avec l'égalité des prédicteurs et la méthode proposée et le pourcentage total d'indices non transmis en fonction des images de référence de la configuration Baseline IPP, en moyenne pour l'ensemble des séquences et des six QP utilisés.

de cette évolution, nous avons substitué la fonction d'estimation de coût  $\zeta(x)$  correspondant à une table de Golomb dans le logiciel de référence, par d'autres fonctions. Nous noterons  $Golomb(x)$  cette fonction. Les pourcentages d'élimination engendrés par chacune de ces fonctions, pour la séquence *Foreman* QCIF 15Hz codée avec le profil Baseline à QP 32, sont reportés dans le tableau 7.3. L'ensemble des fonctions testées ont un pourcentage de suppression d'indices de prédicteurs supérieur à celui obtenu avec la fonction  $Golomb(x)$  de référence. Pour les fonctions  $|x|$ ,  $x^2$  et  $e^x$ , la raison de cette supériorité est liée à leurs croissances. La figure 7.2 est la représentation graphique de l'ensemble des fonctions que nous avons testées. Si l'on exclut le cas de la fonction  $\text{floor}(\log(|x|))$ , les pourcentages de suppression sont ordonnés selon la rapidité de croissance de ces fonctions. Cependant, la fonction qui obtient la plus forte proportion de suppression de prédicteurs est la fonction  $\text{floor}(\log(|x|))$  qui a justement la croissance la plus faible. Avec cette fonction, les coûts des deux résiduels  $\varepsilon_{mv_1}$  et  $\varepsilon_{mv_2}$  sont très souvent égaux, donc le prédicteur numéro 1 est souvent sélectionné. De même, les coûts de  $mv_1 - p_1$  et  $mv_1 - p_2$  ou  $mv_2 - p_1$  et  $mv_2 - p_2$  ont la même valeur dans de nombreux cas. Par conséquent, les équations (7.1) et (7.2) sont plus souvent vérifiées que si l'on utilise la fonction  $Golomb(x)$ .

Bien que toutes les fonctions testées engendrent une proportion plus élevée d'élimination d'indices de prédicteurs, elles ne fournissent pas un choix optimal, au sens du critère débit-distorsion, du meilleur prédicteur du vecteur mouvement. Dès lors, la réduction de débit obtenue par la baisse du débit des indices de prédicteurs ne comble pas, en moyenne, le choix sous optimal du prédicteur du vecteur mouvement de chacune de ces fonctions.

### 7.3 Conclusion

L'information liée à la compétition des prédictions de vecteurs de la méthode MVComp proposée dans le chapitre 4 représente en moyenne 15% du débit de l'information de mouvement. Dans ce chapitre, nous avons proposé de réduire cette information de compétition en utilisant une méthode de suppression d'indice de prédicteurs basée sur le critère de sélection du meilleur prédicteur. La

Fonction de choix	Prédicteurs implicites
$Golomb(x)$	12,7
$ x $	14,7
$x^2$	16,9
$e^x$	30,8
$\text{floor}(\log( x ))$	40,1

TABLE 7.3 – Proportion de suppression des indices des prédicteurs obtenue avec la méthode proposée, en fonction de plusieurs fonctions de sélection du meilleur prédicteur, pour la séquence *Foreman* QCIF 15Hz codée avec le profil Baseline à QP 32.

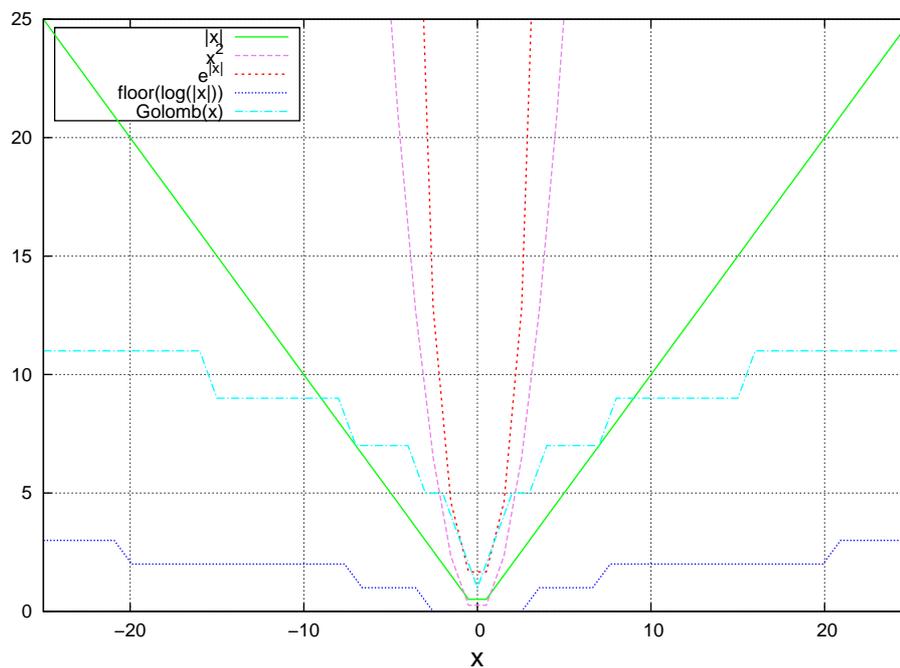


FIGURE 7.2 – Représentation graphique des fonctions  $Golomb(x)$ ,  $|x|$ ,  $x^2$ ,  $e^x$ ,  $\text{floor}(\log(|x|))$ .

méthode permet de supprimer plus de 10% des indices de prédicteurs. Si l'on ajoute à ce pourcentage les indices supprimés par l'égalité des prédicteurs, plus d'un quart de l'information de compétition des prédicteurs n'est pas transmise au décodeur. Cette information a cependant une faible proportion dans le débit total. Par conséquent, l'impact de cette modification est faible sur l'efficacité globale de codage.

La réduction de l'information de compétition est une voie de recherche essentielle pour l'amélioration incrémentale de la norme de référence, car cette information représente la majeure partie du débit total pour les moyens et bas débits. Le transfert d'une partie de l'intelligence du système (concentrée en majorité au codeur) vers le décodeur répond à ce besoin de réduction, comme nous venons de le vérifier dans ce chapitre. Dans le chapitre suivant, nous proposons une méthode de suppression d'indices de prédicteurs de blocs, basée sur ce transfert d'intelligence.



## Chapitre 8

# Suppression de prédicteurs pour le codage Intra

*Dans une méthode de codage par compétition, la pertinence des modes de prédiction mis en compétition ne dépend pas uniquement de leur efficacité intrinsèque, mais aussi de leur complémentarité. La méthode proposée dans ce chapitre repose sur cette remarque, en considérant que certains prédicteurs mis en compétition sont trop similaires pour combler le coût de l'information de compétition engendrée par ces multiples prédicteurs. L'information liée à la compétition des prédicteurs Intra atteint 25% du débit total d'un slice Intra à bas débit pour l'ensemble des séquences VCEG (cf. chapitre 3). Le coût de la compétition des prédicteurs Intra n'est donc pas négligeable. Pour tenter de réduire ce débit, nous proposons une méthode permettant de réduire cette information de compétition, en éliminant des prédicteurs de l'ensemble des prédicteurs. La méthode utilise la proximité des prédicteurs dans l'espace transformé afin d'obtenir un ensemble comprenant des prédicteurs suffisamment différents pour ne pas être redondants en terme de codage. La notion de proximité est déterminée à l'aide du pas de quantification utilisé pour le codage, permettant à la méthode d'être adaptée aux différents débits. Le critère d'élimination retenu étant reproductible au décodeur, la méthode est décodable.*

### 8.1 Description de la méthode

L'objectif de notre méthode est de supprimer de l'ensemble des prédicteurs, les prédicteurs superflus pour le codage. Les prédicteurs superflus sont les prédicteurs obtenant un codage équivalent à un autre prédicteur. Nous allons, dans cette section, définir plusieurs critères d'équivalence, avant de déterminer le critère reproductible au décodeur qui sera utilisé dans la méthode.

---

1. Publications de l'auteur associées : [Revue1] et [Brev4].

### 8.1.1 Critère idéal

Notons  $X \in E$  le groupe de pixels original à prédire, où  $E$  représente l'ensemble des partitions possibles de l'image. C'est-à-dire que  $X$  peut être un bloc  $4 \times 4$ ,  $8 \times 8$  ou  $16 \times 16$  dans le cas du codage Intra de la norme H.264/AVC, ou une ligne ou une colonne dans le cas de notre implémentation de l'Intra 1D. Considérons que pour chacune des partitions,  $N$  prédicteurs  $P_i$  sont mis en compétition pour la prédiction de ce signal. Nous noterons  $\varepsilon_{P_i}$  le résiduel associé au prédicteur  $P_i$  tel que  $\varepsilon_{P_i} = X - P_i$ . Les différentes transformées utilisées dans la norme (DCT  $4 \times 4$  et DCT  $8 \times 8$ ) et la transformée DCT 1D dans le cas de l'Intra 1D seront notées  $DCT(\varepsilon_{P_i})$ . Les opérations  $Q$  et  $Q^{-1}$  de l'équation (1.6) du chapitre 1 représentent respectivement l'opération de quantification et de déquantification. On notera  $\tilde{\varepsilon}_{P_i} = DCT^{-1}(Q^{-1}Q(DCT(\varepsilon_{P_i})))$  le résiduel transformé, quantifié, déquantifié et détransformé, c'est à dire le résiduel décodé. Enfin  $\tilde{X}_{P_i} = \tilde{\varepsilon}_{P_i} + P_i$  est le signal reconstruit associé au prédicteur  $P_i$ .

Idéalement, pour un signal  $X$  donné, on souhaiterait supprimer un prédicteur  $P_j$  si un autre prédicteur  $P_i$  de l'ensemble des prédicteurs, avec  $i \neq j$ , conduit exactement au même bloc décodé. Dans la suite du chapitre, on notera ce critère  $C_{ideal}$  qui est défini par l'équation suivante :

$$\tilde{X}_{P_i} = \tilde{X}_{P_j} \quad (8.1)$$

Ce critère est, comme son nom l'indique, idéal mais pas réaliste puisqu'il utilise le signal de référence  $X$ . Par conséquent, il n'est pas reproductible au décodeur.

### 8.1.2 Critère d'égalité des résiduels reconstruits

Considérons que l'on veut supprimer un prédicteur  $P_j$  s'il existe un prédicteur  $P_i$  dans l'ensemble des prédicteurs, avec  $i \neq j$ , si les résiduels reconstruits sont égaux. Dans la suite du chapitre, on notera ce critère  $C_{err}$ , qui est défini par l'équation suivante :

$$\tilde{\varepsilon}_{P_i} = \tilde{\varepsilon}_{P_j} \quad (8.2)$$

Avec ce critère, on considère uniquement la réduction d'entropie apportée par la prédiction, contrairement au critère  $C_{ideal}$  qui considère la réduction d'entropie et la qualité de la reconstruction. Par conséquent, le critère  $C_{err}$  devrait éliminer plus de prédicteurs.

Dans l'équation (8.2), le signal original  $X$  n'étant pas disponible au décodeur, nous allons transformer notre critère afin d'éliminer cette valeur de nos équations. L'équation (8.2) peut être développée comme suit :

$$DCT^{-1}(Q^{-1}Q(DCT(X - P_i))) = DCT^{-1}(Q^{-1}Q(DCT(X - P_j))) \quad (8.3)$$

Par la propriété  $DCT^{-1}(Z) = DCT^{-1}(Y) \Leftrightarrow Z = Y$ , l'équation (8.3) devient :

$$Q^{-1}Q(DCT(X - P_i)) = Q^{-1}Q(DCT(X - P_j)) \quad (8.4)$$

Dans cette équation ajoutons  $DCT(X - P_i) - DCT(X - P_i)$  et  $DCT(X - P_j) - DCT(X - P_j)$  respectivement dans le terme à gauche et le terme à droite de l'équation (8.4) :

$$\begin{aligned} & Q^{-1}Q(DCT(X - P_i)) + DCT(X - P_i) - DCT(X - P_i) \\ & = Q^{-1}Q(DCT(X - P_j)) + DCT(X - P_j) - DCT(X - P_j) \end{aligned} \quad (8.5)$$

En développant cette équation on obtient :

$$\begin{aligned} & Q^{-1}Q(DCT(X - P_i)) - DCT(X - P_i) - Q^{-1}Q(DCT(X - P_j)) + DCT(X - P_j) \\ & = DCT(X - P_j) - DCT(X - P_i) \end{aligned} \quad (8.6)$$

La transformée DCT étant linéaire, le terme  $DCT(X - P_j) - DCT(X - P_i)$  est égal à  $DCT(X) - DCT(P_j) - DCT(X) + DCT(P_i) = DCT(P_i - P_j)$ . L'équation (8.2) est donc équivalente à :

$$\begin{aligned} & Q^{-1}Q(DCT(X - P_i)) - DCT(X - P_i) - Q^{-1}Q(DCT(X - P_j)) + DCT(X - P_j) \\ & = DCT(P_i - P_j) \end{aligned} \quad (8.7)$$

Les propriétés des opérations conjointes de quantification/déquantification engendrent la relation suivante :

$$|Q^{-1}Q(DCT(Z)) - DCT(Z)| \leq q/2 \quad (8.8)$$

où  $q$  représente soit les coefficients de la matrice de quantification dans le cas de la transformée DCT carrée, soit les éléments du vecteur de quantification dans le cas de la DCT 1D. Grâce à cette propriété, nous pouvons donc écrire :

$$\begin{aligned} & |Q^{-1}Q(DCT(X - P_i)) + DCT(X - P_i) - Q^{-1}Q(DCT(X - P_j)) + DCT(X - P_j)| \\ & \leq |Q^{-1}Q(DCT(X - P_i)) - DCT(X - P_i)| + |-(Q^{-1}Q(DCT(X - P_j)) + DCT(X - P_j))| \\ & \leq 2 \times q/2 \end{aligned} \quad (8.9)$$

donc

$$\Leftrightarrow |DCT(P_i - P_j)| \leq 2 \times q/2 \quad (8.10)$$

Or, par propriété de l'opération conjointe de quantification et de déquantification  $|DCT(Z)| \leq q/2 \Leftrightarrow Q^{-1}Q(DCT(Z)) = 0$

L'équation (8.10) est donc équivalente à une quantification pour laquelle on utiliserait des coefficients de la matrice ou du vecteur de quantification deux fois plus grands que les coefficients  $q$  utilisés pour le codage. L'équation (8.10) peut donc s'écrire :

$$Q_{\times 2}^{-1}Q_{\times 2}(DCT(P_i - P_j)) = 0 \quad (8.11)$$

où  $Q_{\times 2}$  et  $Q_{\times 2}^{-1}$  représentent respectivement l'opération de quantification et de déquantification, avec des coefficients multipliés par deux par rapport aux coefficients de quantification utilisés pour

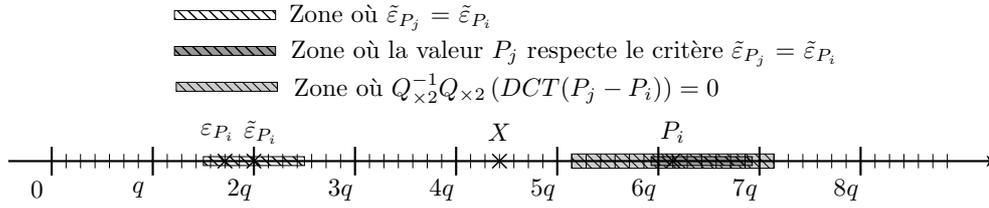


FIGURE 8.1 – Représentation graphique de  $X$ ,  $P_i$ ,  $\varepsilon_{P_i}$ ,  $\tilde{\varepsilon}_{P_i}$  et des zones où le critère de l'équation (8.11) est vérifié lorsque  $E$  correspond à un pixel.

le codage. On sait que le doublement des coefficients de quantification correspond à l'ajout de 6 au QP courant. Donc, les opérations  $Q_{\times 2}$  et  $Q_{\times 2}^{-1}$  correspondent aux opérations  $Q$  et  $Q^{-1}$  avec un pas de quantification (que l'on notera  $QP_{max}$ ) égal à  $QP+6$ , où  $QP$  est le pas de quantification utilisé pour le codage.

### 8.1.3 Approximation du critère optimal

Dans la section précédente, on a montré que si l'équation (8.11) ( $C_{err}$ ) est vérifiée, la relation de l'équation (8.2) peut être vraie, et que si l'on choisit des coefficients de quantification supérieurs à  $2q$ , le critère  $C_{err}$  n'est jamais vérifié. C'est à dire que l'équation (8.2) implique l'équation (8.11) mais l'implication de l'équation (8.11) vers l'équation (8.2) n'est pas vérifiée. Dans cette section, nous allons proposer de rendre cette implication vraie dans un certain nombre de cas, en considérant les propriétés de répartition des valeurs de la différence entre  $DCT(\varepsilon_{P_i})$  et  $DCT(\tilde{\varepsilon}_{P_i})$ . Le but est alors de trouver le QP optimal ( $QP_{opt}$ ) qui obtiendra la meilleure efficacité. Nous noterons ce critère  $C_{opt}$  défini par l'équation :

$$Q_{opt}^{-1}Q_{opt}(DCT(P_i - P_j)) = 0 \quad (8.12)$$

L'équation (8.11) nous donne la borne maximale de  $QP_{opt}$ . On cherche  $QP_{opt}$  tel que  $0 \leq QP_{opt} \leq QP_{max}$ . La figure 8.1 illustre ce problème de sélection du  $QP_{opt}$ , pour une partition correspondant à un pixel. Dans ce cas  $DCT(Z) = Z$ . La notation DCT ne sera donc pas reportée dans ce schéma. L'axe représente l'amplitude des pixels et de leurs différences. On voit clairement dans ce schéma que la zone où l'équation (8.11) est vérifiée, est deux fois plus grande que la zone où le critère  $C_{err}$  est vérifié. C'est-à-dire que l'implication de (8.12) vers (8.2) n'est valide que dans 50% des cas si l'on considère que l'on a une équiprobabilité de valeurs  $DCT(\tilde{\varepsilon}_{P_i}) - DCT(\tilde{\varepsilon}_{P_j})$  et  $P_i - P_j$ .

La figure 8.2 représente l'histogramme des valeurs de la différence  $DCT(\tilde{\varepsilon}_{P_i}) - DCT(\tilde{\varepsilon}_{P_j})$  du coefficient d'indice de fréquence numéro 8 avec l'ordre de parcours zigzag (cf. figure 1.2 du chapitre 1), pour tous les prédicteurs du mode Ligne de la séquence *Foreman* CIF encodée avec le QP 42. Nous avons choisi cette valeur de QP car nous savons que la méthode proposée sera plus efficace à bas débit. En effet, la proportion du débit des indices des prédicteurs dans le débit total est plus élevée à bas débit qu'à haut débit. Dans cette figure, on voit que l'ensemble des valeurs de

$DCT(\tilde{\varepsilon}_{P_i}) - DCT(\tilde{\varepsilon}_{P_j})$  comprises entre  $-q/4$  et  $q/4$  représente 99% des cas. Par conséquent, nous pouvons considérer pour notre test que les composantes  $|Q^{-1}Q(DCT(X - P_i)) - DCT(X - P_i)|$  et  $|Q^{-1}Q(DCT(X - P_j)) - DCT(X - P_j)|$  sont bornées par  $q/4$  dans 99% des cas ( $q/2$  dans 100% des cas). On peut alors, dans 99% de cas, considérer que  $q_{opt} = q$ . Par conséquent, la valeur de  $QP_{opt}$  sera comprise entre la valeur du QP utilisé pour le codage et 0 (et non entre  $QP_{max}$  et 0).

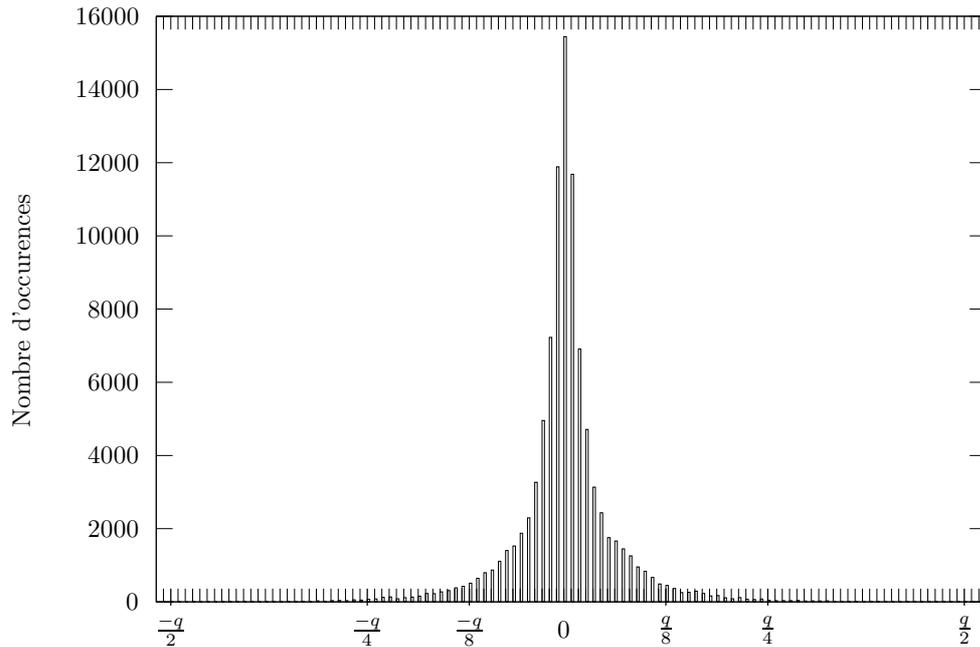


FIGURE 8.2 – Histogramme des valeurs de  $|Q^{-1}Q(DCT(X - P_i)) - DCT(X - P_i)|$  de l'ensemble des prédicteurs mis en compétition pour le mode Intra 1D Ligne de la première image de la séquence *Foreman* CIF codée avec le profil Baseline à QP 42.

Il faut noter que l'implication de (8.12) vers (8.2) n'est pas vérifiée (pour la totalité des cas) lorsque l'inégalité  $0 \leq QP_{opt} \leq QP$  est vraie. De plus, une partie des cas où (8.2)  $\Rightarrow$  (8.12) est éliminée avec cette inégalité. Il faut aussi considérer que le  $QP_{opt}$  dépend de la répartition de la différence de  $DCT(\tilde{\varepsilon}_{P_i}) - DCT(\tilde{\varepsilon}_{P_j})$  qui est variable en fonction des QP utilisés pour le codage, mais aussi du partitionnement, de l'ordre de parcours et de l'efficacité de chacun des prédicteurs, comme on a pu le voir dans le tableau 6.2 du chapitre 6. Enfin, le critère  $C_{err}$  de l'équation (8.2) considère uniquement la réduction de débit. En effet, avec ce critère on élimine un prédicteur s'il en existe un autre qui obtient le même résiduel reconstruit. On ne prend donc pas en compte la distorsion apportée par chacun des deux prédicteurs. L'efficacité de la compétition de prédictions ne dépend pas uniquement des résiduels engendrés, mais aussi de la proximité entre les prédicteurs et le signal original  $X$ . Pour toutes ces raisons, nous avons fixé de manière empirique un  $QP_{opt}$  pour

chacun des modes Intra que l'on va utiliser, en sachant que  $0 \leq QP_{opt} \leq QP_{max}$  avec une forte probabilité d'avoir une borne maximum égale à  $QP$ .

## 8.2 Résultats expérimentaux

Dans cette section, nous décrivons, dans un premier temps, l'algorithme de suppression des prédicteurs. Dans un second temps, nous reportons les résultats empiriques qui nous ont permis de fixer le  $QP_{opt}$  du critère  $C_{opt}$ . Puis, nous comparons l'efficacité de codage de chacun des critères pour le mode Intra  $4 \times 4$  avant de donner les résultats de la méthode avec le critère  $C_{opt}$  par rapport à la norme et à notre méthode de l'Intra 1D, pour l'ensemble des séquences utilisées dans les configurations Intra recommandées par le groupe VCEG.

### 8.2.1 Implémentation de l'algorithme

Pour les trois critères que nous avons définis, chaque prédicteur  $P_i$ ,  $0 < i \leq N - 1$  est comparé au prédicteur  $P_j$  avec  $i < j \leq N$ . Si le couple  $P_i, P_j$  respecte le critère, le prédicteur  $P_j$  est éliminé de l'ensemble des prédicteurs. A la fin de cette étape, dans le nouvel ensemble de prédicteurs, aucun couple de prédicteurs ne vérifie le critère.

Afin d'optimiser le codage des prédicteurs, nous avons ajouté comme contrainte à notre algorithme, l'obligation d'avoir 1, 2, 3, 5, ou 9 prédicteurs dans l'ensemble des prédicteurs. Ces nombres correspondent au nombre de bits optimum pour le codage des prédicteurs Intra, comme nous l'avons expliqué dans le chapitre 1. Par exemple, si 4 prédicteurs appartiennent au nouvel ensemble avec la méthode proposée, le codage s'effectue soit sur 1 bit (si le prédicteur sélectionné est égal au "Most Probable Mode") soit sur 3 bits. Nous devons donc coder 3 prédicteurs avec 2 bits alors que nous pouvions en coder 4. A la fin du calcul du nouvel ensemble, si le nombre de prédicteurs n'est pas égal à 1, 2, 3, 5 ou 9, nous ajoutons dans l'ensemble de prédicteurs les derniers prédicteurs éliminés pour atteindre un nombre optimal de prédicteurs par rapport aux nombres de bits alloués. Précisons que le codage des prédicteurs Intra avec le CABAC utilise la même binarisation que le CAVLC. Par conséquent, nous utilisons le procédé décrit précédemment pour le CABAC et le CAVLC.

### 8.2.2 Sélection de $QP_{opt}$

Afin de sélectionner le  $QP_{opt}$  du critère  $C_{opt}$ , nous avons testé, pour les modes Intra  $4 \times 4$ , Intra  $8 \times 8$  et Intra 1D, l'ensemble des valeurs tel que  $QP - 6 \leq QP_{opt} \leq QP + 6$ . Ces tests ont été effectués indépendamment pour chacun des modes. Les résultats, en pourcentage de réduction de débit, sont reportés dans la figure 8.3 pour l'Intra  $4 \times 4$ , la figure 8.4 pour l'Intra  $8 \times 8$  et la figure 8.5 pour l'Intra 1D. Seuls les résultats de l'ensemble des QP 27-42 représentant les bas débits sont reportés, car la proportion du débit des prédicteurs Intra est plus élevée pour ces débits. Dans les configurations testées, le mode Intra  $4 \times 4$  est en compétition avec l'Intra  $16 \times 16$  en profil Baseline. Le mode Intra  $8 \times 8$  est en compétition avec l'Intra  $4 \times 4$  et l'Intra  $16 \times 16$  en profil High. Et le mode Intra 1D est en compétition avec l'Intra  $4 \times 4$  et l'Intra  $16 \times 16$  en profil Baseline.

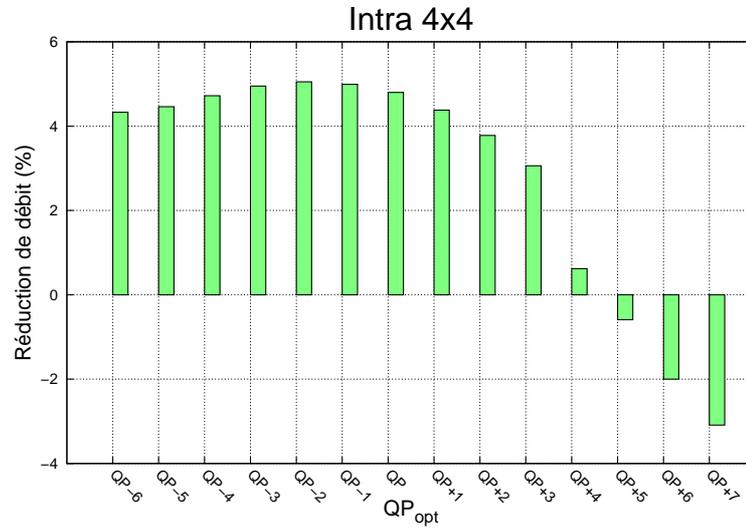


FIGURE 8.3 – Pourcentage de réduction de débit obtenu avec l’utilisation du critère  $C_{opt}$  sur les prédicteurs de l’Intra  $4 \times 4$ , pour plusieurs valeurs de  $QP_{opt}$ , en moyenne pour l’ensemble des séquences VCEG codées avec le profil Baseline pour l’ensemble de QP 27-42.

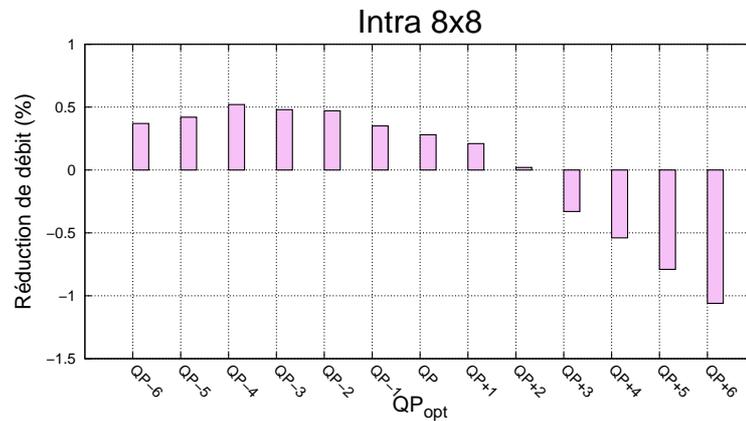


FIGURE 8.4 – Pourcentage de réduction de débit obtenu avec l’utilisation du critère  $C_{opt}$  sur les prédicteurs de l’Intra  $8 \times 8$ , pour plusieurs valeurs de  $QP_{opt}$ , en moyenne pour l’ensemble des séquences VCEG codées avec le profil High pour l’ensemble de QP 27-42.

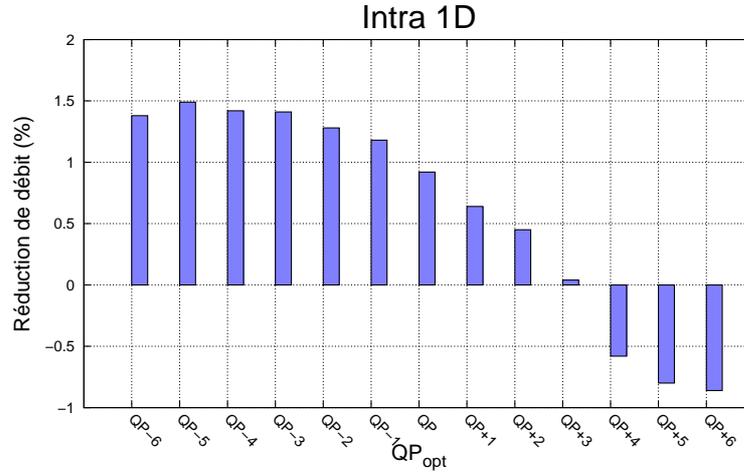


FIGURE 8.5 – Pourcentage de réduction de débit obtenu avec l’utilisation du critère  $C_{opt}$  sur les prédicteurs de l’Intra 1D, pour plusieurs valeurs de  $QP_{opt}$ , en moyenne pour l’ensemble des séquences VCEG codées avec le profil Baseline pour l’ensemble de QP 27-42.

On remarque que  $QP_{max}$ ,  $QP+6$ ,  $QP+4$  et  $QP+5$ , donnent systématiquement des pourcentages de gains négatifs. L’Intra  $4 \times 4$  atteint son maximum de gain pour  $QP_{opt} = QP - 2$ , l’Intra  $8 \times 8$  pour  $QP_{opt} = QP - 4$ , et l’Intra 1D pour  $QP_{opt} = QP - 5$ . On remarque, pour chaque mode Intra, que le pourcentage de gain croît entre  $QP_{max}$  et  $QP_{opt}$ , en atteignant son maximum pour  $QP_{opt}$ , et que, après ce maximum, les gains décroissent. De plus, ces valeurs maximales sont différentes, car elles dépendent de la nature (taille, partitionnement, nombre de prédicteurs par partition) des trois modes Intra. Nous utiliserons, dans la suite du chapitre, ces trois valeurs de  $QP_{opt}$ .

### 8.2.3 Comparaison des critères

Dans cette section, nous comparons les trois critères proposés ( $C_{ideal}$ ,  $C_{err}$  et  $C_{opt}$ ). Nous avons testé ces critères pour le profil Baseline, pour les prédicteurs de l’Intra  $4 \times 4$ , sur l’ensemble des séquences VCEG. Dans cette configuration, l’Intra  $4 \times 4$  est en compétition avec l’Intra  $16 \times 16$ . Les résultats sont reportés dans le tableau 8.1.

Le critère  $C_{opt}$  est le plus efficace pour l’ensemble des QP. Le critère  $C_{err}$  obtient la deuxième position. Le critère  $C_{ideal}$  est le critère le moins efficace, mais il obtient tout de même un gain pour l’ensemble des débits. Ce faible gain par rapport aux deux autres critères est expliqué par le nombre moins élevé de cas où l’équation (8.1) est validée. Cependant, ce critère élimine toujours les bons prédicteurs : ceux qui donnent la même reconstruction. Le critère  $C_{err}$ , quant à lui, supprime les prédicteurs offrant le même résiduel reconstruit qu’un autre prédicteur de l’ensemble. Il peut donc évincer des prédicteurs qui offraient une meilleure prédiction en terme de distorsion. Le critère  $C_{err}$  élimine donc plus de prédicteurs que le critère  $C_{ideal}$  mais il peut se “tromper”. Le critère  $C_{opt}$  obtient de meilleurs résultats que le critère  $C_{err}$ , car nous avons fixé un QP optimal. En effet, si on utilise le même QP que le critère  $C_{err}$ , le critère  $C_{opt}$  est naturellement moins bon (cf. figure 8.3)

		Baseline I		
QP		17-32	22-37	27-42
$C_{ideal}$		0,4	0,7	1,1
$C_{err}$		1,2	2,2	4,1
$C_{opt}$		1,5	2,8	5,1

TABLE 8.1 – Pourcentage de réduction de débit avec les critères  $C_{ideal}$ ,  $C_{err}$  et  $C_{opt}$  appliqués sur les prédicteurs de l’Intra 4×4, pour le profil Baseline sur l’ensemble des séquences VCEG.

puisqu’il supprime un plus grand nombre de prédicteurs sans tenir compte du signal d’entrée. Grâce au  $QP_{opt}$  inférieur au QP du codage, le nombre d’éliminations de prédicteurs efficaces est restreint.

## 8.2.4 Application à H.264/AVC

### 8.2.4.1 Gain sur le débit total

Le tableau 8.2 donne le pourcentage de réduction de débit obtenu avec la méthode, séquence par séquence, pour le profil Baseline III et le profil High III. Pour le profil Baseline III, le mode Intra 4×4 est en compétition avec l’Intra 16×16 et pour le profil High III, les modes Intra 4×4 et Intra 8×8 sont en compétition avec l’Intra 16×16. La méthode est activée uniquement pour l’Intra 4×4 et l’Intra 8×8. Le gain est supérieur pour le profil Baseline III car, dans ce profil, l’Intra 8×8 n’est pas activé. En effet, nous avons remarqué dans le chapitre 6 que le débit par macrobloc des prédicteurs Intra 4×4 était plus élevé que celui de l’Intra 8×8. En moyenne la méthode obtient respectivement 1,8%, 3,4% et 6% de gain sur le débit pour les hauts, moyens et bas débits. Ces résultats sont élevés par rapport à la proportion de débit des prédicteurs Intra dans le débit total (cf. tableau 6.10 du chapitre 6). Remarquons que la réduction maximale de débit est de 9,4% pour la séquence *Crew* 720p. Pour le profil High III, la méthode obtient respectivement 0,5%, 0,8% et 1,5% de gain moyen pour les hauts, moyens et bas débits. Pour les deux configurations, les gains sont supérieurs à bas débit, car la proportion de débit des indices des prédicteurs Intra y est plus élevée.

### 8.2.4.2 Analyses

Le tableau 8.3 donne le pourcentage de sélection des modes Intra pour la méthode proposée. Le mode Intra 4×4 maximise la sélection pour l’ensemble des QP du profil Baseline III, ce qui n’est pas le cas de la référence, comme on peut le voir dans le tableau 6.4 du chapitre 6. En effet, le mode Intra 16×16 est le plus sélectionné pour les QP 37 et 42 dans la référence. Pour l’ensemble des QP, on observe une augmentation de la proportion du mode Intra 4×4. Cependant, cet accroissement est plus important à bas débit, passant de 32,2% à 61,8% pour le QP 42. En moyenne, cette sélection passe de 65,5% à 78,7%.

Pour le profil High III, le mode Intra 8×8 reste le mode le plus sélectionné, avec une légère hausse par rapport à la référence (cf. tableau 6.5 du chapitre 6). Le pourcentage de sélection de l’Intra 4×4

QP	Baseline III			High III		
	17-32	22-37	27-42	17-32	22-37	27-42
	<b>QCIF</b>					
Container	0,9	1,4	2,2	0,3	0,3	0,2
Foreman	1,1	2,3	5,0	0,8	1,4	2,3
Silent	1,0	2,3	5,3	0,8	1,1	1,6
	<b>CIF</b>					
Foreman	1,1	2,4	5,1	0,5	1,0	1,9
Mobile	0,3	0,5	1,2	0,2	0,4	0,9
Paris	0,8	1,4	2,8	0,3	0,6	1,0
Tempete	0,4	0,7	2,0	0,2	0,3	0,5
Carphone	2,2	4,4	7,5	1,2	2,3	<b>3,8</b>
Crew	2,2	4,6	8,5	0,8	1,4	2,1
Ice	3,5	4,4	7,1	1,6	2,3	3,7
Schumacher	1,8	2,6	3,9	0,5	0,7	1,0
	<b>4CIF</b>					
City	0,7	2,0	5,0	0,2	0,4	0,9
Ice	4,4	5,9	9,0	1,5	2,2	3,3
Soccer	1,1	2,5	5,0	0,2	0,4	0,8
	<b>720p</b>					
Bigships	1,5	3,6	7,3	0,4	0,8	1,6
Shuttlestart	4,3	6,5	9,1	0,7	1,3	2,4
City	0,8	2,2	5,2	0,2	0,4	0,8
Crew	3,5	6,1	<b>9,4</b>	0,3	0,7	1,4
Night	1,3	2,9	5,7	0,3	0,6	1,3
Raven	3,0	4,6	6,3	0,3	0,5	0,9
Stockholm	0,6	1,9	5,3	0,2	0,3	1,1
Cyclists	3,7	6,3	8,9	0,3	0,5	0,9
Jets	3,2	5,1	7,8	0,0	0,1	0,7
Panslow	0,5	1,7	4,0	0,2	0,4	1,2
Sailormen	0,9	2,4	5,5	0,2	0,5	1,1
Spincalendar	0,6	1,8	3,8	0,2	0,7	1,6
	<b>1080p</b>					
Blue sky	2,5	3,6	5,9	1,0	1,3	1,3
Intotree	1,2	3,0	8,0	0,2	0,6	1,7
Station2	3,1	6,2	8,9	0,2	0,6	1,2
Toys and calendar	2,6	5,8	9,2	0,2	0,6	1,5
<b>Moyenne</b>	<b>1,8</b>	<b>3,4</b>	<b>6,0</b>	<b>0,5</b>	<b>0,8</b>	<b>1,5</b>

TABLE 8.2 – Pourcentage de réduction de débit, avec le critère  $C_{opt}$ , de l'ensemble des séquences, pour le profil Baseline III et le profil High III, pour trois ensembles de QP, lorsque la méthode est activée pour les modes Intra  $4 \times 4$  et  $8 \times 8$ .

	Méthode Proposée				
	Baseline III		High III		
QP	Intra 16×16	Intra 4×4	Intra 16×16	Intra 8×8	Intra 4×4
17	5,4	<b>94,6</b>	4,3	<b>51,4</b>	44,3
22	11,4	<b>88,6</b>	5,2	<b>57,5</b>	37,3
27	17,7	<b>82,3</b>	11,8	<b>50,1</b>	38,1
32	24,5	<b>75,5</b>	15,2	<b>51,9</b>	32,9
37	30,8	<b>69,2</b>	19,6	<b>53,6</b>	26,8
42	38,2	<b>61,8</b>	23,1	<b>60,5</b>	16,4
<b>Moyenne</b>	21,3	<b>78,7</b>	13,2	<b>54,2</b>	32,6

TABLE 8.3 – Pourcentage de sélection, en fonction des QP, des modes Intra 16×16, 4×4 et 8×8 pour la méthode proposée, codée avec le profil Baseline III et le profil High III, calculé pour l'ensemble des séquences.

subit une augmentation un peu plus élevée, passant de 27,9% à 32,6%.

Le tableau 8.4 donne le nombre de bits par macrobloc (*bpm*) moyen pour le codage des indices des prédicteurs de l'Intra 4×4 et de l'Intra 8×8, pour les deux profils testés de la référence et de la méthode proposée. Pour le profil Baseline III, le *bpm* moyen de l'Intra 4×4 a été réduit de presque 10 bits. Cette diminution est plus forte pour les bas débits, passant de 27,8 *bpm* pour la référence à 11,2 *bpm* pour la méthode proposée pour le QP 42, soit 60% de réduction. L'objectif de réduire l'information des prédicteurs à bas débit est donc atteint.

Pour l'Intra 8×8 du profil High, la réduction du nombre de *bpm* est plus faible, cependant elle est systématique pour tous les QP. Pour l'Intra 4×4 le *bpm* a aussi diminué, mais la réduction est moins importante que pour le profil Baseline. Elle reste tout de même élevée avec 32% de baisse pour le QP 42. La méthode est plus efficace pour l'Intra 4×4 que pour l'Intra 8×8 car le nombre de *bpm* est moins élevé pour l'Intra 8×8. Il est, par conséquent, plus difficile de réduire le débit de ses prédicteurs. La différence de réduction du nombre de *bpm* de l'Intra 4×4 pour le profil Baseline et le profil High s'explique par la sélection de l'Intra 8×8. En théorie, les partitions (Intra) de petite taille sont sélectionnées dans les zones contenant beaucoup de hautes fréquences. A l'inverse, les macroblocs représentant une surface plane seront codés avec des partitions de grande taille. Dans le profil High, l'Intra 4×4 est donc sélectionné pour les régions comprenant beaucoup de hautes fréquences, alors que pour le profil Baseline ce mode sera choisi pour ces mêmes régions mais aussi pour une partie de celles qui auraient été codées avec l'Intra 8×8 pour le profil High. Notre méthode élimine un prédicteur s'il en existe un autre dans l'ensemble qui lui est semblable. C'est précisément dans les zones contenant peu de hautes fréquences que les prédicteurs sont semblables. Par conséquent, la méthode élimine plus de prédicteurs pour ces zones. En conclusion, pour le profil High, l'Intra 4×4 est en majorité sélectionné dans les zones où notre méthode est moins efficace, alors que pour le

	Baseline III		High III			
	H.264	Méthode Proposée	H.264		Méthode Proposée	
QP	Intra 4×4	Intra 4×4	Intra 8×8	Intra 4×4	Intra 8×8	Intra 4×4
17	41,8	<b>38,7</b>	10,8	41,0	10,5	38,9
22	39,2	<b>33,7</b>	9,8	37,5	9,6	34,0
27	36,2	<b>29,1</b>	9,4	33,5	8,9	28,9
32	33,5	<b>23,4</b>	8,2	29,3	7,7	23,6
37	30,8	<b>17,5</b>	7,1	24,2	6,5	18,1
42	27,8	<b>11,2</b>	5,4	18,1	4,8	12,7
<b>Moyenne</b>	34,9	<b>25,6</b>	8,4	30,6	8,0	26,0

TABLE 8.4 – Nombre de bits moyen par macrobloc (*bpm*) nécessaire au codage des prédicteurs, des modes Intra, 4×4 et 8×8, en fonction des QP, pour la référence et la méthode proposée, codées avec le profil Baseline III et le profil High III, calculé pour l'ensemble des séquences.

profil Baseline, la sélection spatiale de l'Intra 4×4 engendre de plus grandes performances.

## 8.2.5 Application à l'Intra 1D

### 8.2.5.1 Gain sur le débit total

Dans le tableau 8.5, nous avons reporté les réductions de débit liées à l'utilisation conjointe de l'Intra 1D (cf. chapitre 6) et de la méthode de suppression des prédicteurs Intra pour le profil Baseline III et le profil High III. De plus, pour mesurer l'efficacité de la méthode de suppression des prédicteurs Intra, nous donnons les gains de cette méthode par rapport à l'Intra 1D. Nous noterons que la suppression des prédicteurs Intra, comme dans la section précédente, est faite sur l'ensemble des modes à l'exception de l'Intra 16×16. C'est-à-dire que la méthode est appliquée pour l'Intra 4×4, Intra 8×8 et l'Intra 1D.

Avec le profil Baseline III, les gains de l'utilisation des deux méthodes que nous avons développées sont de 9,2% pour l'ensemble des QP recommandés par le groupe VCEG et 12,4% à bas débit. La séquence Schumacher CIF obtient le gain maximal de 21,5% pour ce profil. Pour la configuration High III, les résultats sont respectivement de 4,0% et 4,9% pour les moyens et bas débits. Pour ces deux configurations, les gains sont systématiques.

Si l'on regarde plus précisément les résultats de la méthode proposée dans ce chapitre par rapport à l'Intra 1D, on remarque que pour les deux configurations la méthode apporte systématiquement une réduction de débit. Pour la configuration High III, le gain de la méthode dans cette configuration est supérieur à celui obtenu par la méthode comparée à H.264/AVC. Ceci était prévisible puisque les indices de prédicteurs de l'Intra 1D ont un nombre de *bpm* compris entre celui de l'Intra 4×4 et de l'Intra 8×8 (cf. tableau 6.11 du chapitre 6). En effet, plus le nombre de *bpm* est important, plus

la réduction engendrée par notre méthode est grande.

### 8.2.5.2 Analyses

Le tableau 8.6 donne les pourcentages de sélection de chacun des modes Intra pour les deux configurations. La proportion de sélection a été modifiée avec l'introduction de la méthode proposée. Pour le profil Baseline, lorsque l'on compare cette sélection avec celle du tableau 6.4 du chapitre 6, on remarque que la sélection de l'Intra  $4 \times 4$  est la seule qui bénéficie d'une augmentation. Bien que la sélection de l'Intra 1D ait diminué, elle demeure la sélection la plus élevée en moyenne en dépassant celle de l'Intra  $16 \times 16$  pour tous les QP, excepté le QP 42. Dans le tableau 6.4 du chapitre 6, nous avons vu que l'Intra  $16 \times 16$  dominait la sélection pour le QP 42 et le QP 37.

Pour le profil High III, lorsque l'on compare les sélections avec le tableau 6.5 du chapitre 6, on note que tous les modes Intra sur lesquels nous avons appliqué la méthode subissent une hausse. Cette hausse est présente à tous les débits, excepté pour l'Intra  $8 \times 8$  à QP 42. La sélection de l'Intra  $16 \times 16$  est largement en baisse puisqu'elle passe de 21,7% à 13,2%.

Concernant le nombre de *bpm*, nous notons, comme dans la section précédente, une baisse systématique à tous les débits. Le tableau 8.7 nous donne le nombre de *bpm* pour les modes Intra lorsque notre méthode est utilisée, dans le profil Baseline. La baisse du nombre de *bpm* de l'Intra  $4 \times 4$  est encore très élevée comme dans la section précédente. Pour l'Intra 1D la baisse du nombre de *bpm* est inférieure à celle de l'Intra  $4 \times 4$  car, pour ce mode Intra, nous avons fixé un nombre de prédicteurs pour chacun des sous-modes Intra 1D inférieur en moyenne au nombre de prédicteurs de l'Intra  $4 \times 4$ . Les mêmes remarques sur l'Intra  $4 \times 4$  et l'Intra 1D s'appliquent au profil High, comme on peut le voir dans le tableau 8.8. On notera que la baisse du nombre de *bpm* reste faible pour l'Intra  $8 \times 8$ .

## 8.3 Extension de la méthode

Les résultats présentés dans la section précédente sont élevés par rapport à la proportion des indices des prédicteurs Intra dans le débit total. La question suivante découle naturellement de ces résultats : cette méthode est-elle applicable aux prédicteurs temporels ?

La problématique de l'exploitation des redondances temporelles par prédiction est différente de celle de l'exploitation des redondances spatiales par prédiction. En effet, pour le codage spatial prédictif, les prédicteurs sont créés à partir des pixels voisins déjà décodés. Pour le codage prédictif temporel, si l'on considère qu'il n'existe aucune déformation des objets, le signal à coder est contenu dans les images précédentes. Dans ce cas précis, aucun résiduel ne devrait être transmis.

Au fil de nos recherches, nous avons constaté que l'estimation de mouvement sous-pixellique pouvait se révéler inutile dans certains cas. Dans le chapitre 4, nous avons vu que l'outil de l'estimation de mouvement au  $\frac{1}{8}$ -pel engendrait des pertes sur l'ensemble des séquences 720p VCEG. Il est par conséquent possible que les compensations de mouvement au  $\frac{1}{2}$  et au  $\frac{1}{4}$  de pixel produisent, elles aussi, des pertes. Au cours d'une expérience sur un ensemble de séquences CIF, nous avons observé qu'à bas débit, il était parfois plus efficace de ne pas utiliser de compensation de mouvement sous-pixellique et de coder le vecteur en pixel entier. Finalement, dans [LAP<sup>+</sup>08], une estimation de

Profil	Baseline III				High III			
	H.264		Intra 1D		H.264		Intra 1D	
QP	22-37	27-42	22-37	27-42	22-37	27-42	22-37	27-42
	<b>QCIF</b>							
Container	7,0	9,5	1,1	1,9	4,9	6,0	0,5	0,9
Foreman	5,5	8,3	1,4	2,7	3,2	3,9	1,1	1,7
Silent	6,6	10,5	1,6	3,8	4,7	6,1	0,9	1,7
	<b>CIF</b>							
Foreman	5,7	8,7	1,8	3,4	3,1	4,0	0,8	1,7
Mobile	3,6	4,4	0,5	1,1	3,2	3,2	0,4	0,9
Paris	6,3	8,8	1,4	2,6	5,2	5,9	0,8	1,6
Tempete	4,6	6,4	0,7	1,7	3,4	3,2	0,4	0,9
Carphone	10,8	15,1	3,8	6,4	7,3	9,4	2,4	<b>4,1</b>
Crew	8,7	14,3	3,9	7,3	3,5	4,7	1,7	2,8
Ice	7,9	10,5	3,5	5,5	5,2	6,5	2,2	3,3
Schumacher	17,9	<b>21,5</b>	2,7	4,1	17,0	<b>19,5</b>	1,6	2,5
	<b>4CIF</b>							
City	10,8	15,1	1,8	4,4	3,9	4,3	0,6	1,4
Ice	10,5	12,3	4,2	6,1	6,0	6,7	2,3	3,3
Soccer	11,3	14,4	1,9	4,1	2,3	2,7	0,6	1,3
	<b>720p</b>							
Bigships	10,3	15,7	3,1	6,1	3,2	4,6	1,2	2,4
Shuttlestart	9,5	11,1	4,4	6,3	1,8	2,8	1,5	2,6
City	8,7	13,0	1,8	4,4	1,3	1,8	0,5	1,1
Crew	11,9	15,7	5,2	<b>7,7</b>	2,0	3,4	1,4	2,2
Night	8,5	11,7	2,4	4,7	3,5	4,1	0,8	1,8
Raven	5,6	6,3	2,9	4,4	0,4	0,7	0,4	0,7
Stockholm	8,4	12,8	1,7	4,1	4,1	4,7	0,5	1,2
Cyclists	12,9	15,8	4,4	6,5	1,3	1,6	0,8	1,7
Jets	15,6	16,8	4,9	6,1	3,3	3,7	2,1	2,9
Panslow	10,8	15,8	2,2	4,2	8,3	10,4	1,0	2,2
Sailormen	11,1	15,4	2,3	4,8	4,7	4,8	0,9	1,8
Spincalendar	8,7	12,7	1,6	3,3	5,3	6,8	0,9	1,9
	<b>1080p</b>							
Blue sky	9,0	9,8	2,2	3,7	4,5	2,7	0,8	1,2
Intotree	5,5	11,9	2,6	5,7	0,6	3,4	0,1	2,1
Station2	10,5	12,4	3,6	5,7	0,7	1,3	0,7	1,4
Toys and calendar	11,8	15,7	4,2	6,6	3,2	3,4	1,4	2,5
<b>Moyenne</b>	<b>9,2</b>	<b>12,4</b>	<b>2,7</b>	<b>4,7</b>	<b>4,0</b>	<b>4,9</b>	<b>1,0</b>	<b>1,9</b>

TABLE 8.5 – Pourcentage de réduction de débit de l'ensemble des séquences avec le critère  $C_{opt}$  et la méthode Intra 1D, comparée à la méthode Intra 1D seule et la référence, pour le profil Baseline III et le profil High III, pour deux ensembles de QP, lorsque la méthode est activée pour les modes Intra 4×4, 8×8 et 1D.

	Méthode Proposée						
	Baseline III			High III			
QP	Intra 16×16	Intra 4×4	Intra 1D	Intra 16×16	Intra 8×8	Intra 4×4	Intra 1D
22	6,9	35,1	<b>58,0</b>	3,4	<b>41,8</b>	17,3	37,5
27	15,2	29,8	<b>55,0</b>	8,4	33,9	18,2	<b>39,5</b>
32	25,3	29,0	<b>45,7</b>	11,5	<b>36,6</b>	18,1	33,8
37	35,7	27,1	<b>37,2</b>	18,1	<b>38,6</b>	16,5	26,8
42	<b>50,5</b>	23,3	26,2	24,5	<b>46,9</b>	11,9	16,7
<b>Moyenne</b>	26,7	28,9	<b>44,4</b>	13,2	<b>39,6</b>	16,3	30,9

TABLE 8.6 – Pourcentage de sélection, en fonction des QP, des modes Intra 16×16, 4×4, 8×8 et Intra 1D pour la méthode proposée, codée avec le profil Baseline III et le profil High III, calculé pour l'ensemble des séquences.

	Baseline III			
	H.264		Méthode Proposée	
QP	Intra 4×4	Intra 1D	Intra 4×4	Intra 1D
22	41,8	27,2	35,1	24,4
27	38,9	23,0	30,5	20,1
32	36,6	19,8	25,2	16,8
37	33,9	16,9	19,4	13,5
42	30,3	13,8	13,3	10,2
<b>Moyenne</b>	36,3	20,1	24,7	17,0

TABLE 8.7 – Nombre de bits moyen par macrobloc nécessaire au codage des prédicteurs, des modes Intra, 4×4, en fonction des QP, pour la référence et la méthode proposée, codées avec le profil Baseline III, calculé pour l'ensemble des séquences.

High III						
H.264				Méthode Proposée		
QP	Intra 4×4	Intra 8×8	Intra 1D	Intra 4×4	Intra 8×8	Intra 1D
22	39,6	10,1	26,2	34,6	10,0	37,5
27	35,1	9,8	20,9	28,9	9,2	33,5
32	29,9	8,6	16,2	22,8	7,9	29,3
37	23,5	7,4	12,8	17,2	6,6	24,2
42	16,1	5,7	9,3	11,9	4,8	25,5
<b>Moyenne</b>	28,8	8,3	17,1	23,1	7,7	16,0

TABLE 8.8 – Nombre de bits moyen par macrobloc nécessaire au codage des prédicteurs, des modes Intra 4×4 et 8×8, en fonction des QP, pour la référence et la méthode proposée, codées avec le profil High III, calculé pour l'ensemble des séquences.

mouvement au pixel entier a été testée sur plusieurs résolutions. Les auteurs précisent qu'ils n'ont pas modifié le codage des vecteurs mouvement. Les résultats sont très intéressants, puisque cette modification engendre des pertes pour les basses résolutions, mais, plus la résolution augmente, plus les pertes diminuent. Cette diminution se transforme en gain pour les bas débits de la résolution 1080p (la plus grande résolution testée).

Nous avons donc développé une méthode de suppression de résolutions sous-pixelles basée sur le critère  $C_{opt}$ . Pour déterminer la résolution sous-pixelle optimale pour chaque bloc, nous utilisons le bloc collocated de la première image de référence. Pour ce bloc collocated, si les huit blocs de la résolution  $\frac{1}{2}$  pixelle sont trop proches au sens du critère  $C_{opt}$ , seule l'estimation de mouvement au pixel entier est effectuée et le vecteur du bloc courant est codé au pixel entier. Dans le cas contraire, (si au moins un bloc du  $\frac{1}{2}$  pixel est différent au sens du critère  $C_{opt}$ ), la résolution au  $\frac{1}{4}$  de pixel est calculée pour les blocs du  $\frac{1}{2}$  pixel différents du bloc collocated au sens du  $C_{opt}$ . Si au moins un bloc de l'ensemble des blocs de la résolution du  $\frac{1}{4}$  de pixel est différent au sens du  $C_{opt}$ , l'estimation de mouvement est faite pour le bloc courant au  $\frac{1}{4}$  de pixel, sinon elle est effectuée au  $\frac{1}{2}$  pixel et le vecteur du bloc courant est codé au  $\frac{1}{2}$  pixel. Nous ne donnerons pas dans ce manuscrit les résultats complets obtenus car cette étude n'est pas finalisée. En effet, les  $QP_{opt}$  de chacune des tailles de bloc n'ont pas fait l'objet d'une étude complète. Les résultats préliminaires pour  $QP_{opt} = QP - 17$ , sont déjà très intéressants, comme nous pouvons le voir dans le tableau 8.9 puisque la méthode engendre des pertes pour les résolutions QCIF, CIF et 4CIF, ce qui était prévisible. Mais à partir de la résolution 720p, les séquences contenant peu de hautes fréquences obtiennent des gains, et pour la résolution 1080p, les gains sont quasi systématiques et atteignent 9,2% sur la séquence *Station2* 25Hz. Ces résultats sont cohérents avec les observations décrites dans [LAP<sup>+</sup>08].

Ces résultats préliminaires montrent l'efficacité de l'application de la méthode développée dans ce

	<b>Baseline IPP</b>
QP	27-42
<b>QCIF</b>	-2,3
<b>CIF</b>	-2,1
<b>4CIF</b>	-1,5
<b>720p</b>	-0,7
<b>1080p</b>	
Blue sky	-5,0
Intotree	2,5
Station2	9,2
Toys and calendar	1,8

TABLE 8.9 – Pourcentage de réduction de débit calculé sur l'ensemble des séquences avec le critère  $C_{opt}$  pour la suppression de résolutions sous-pixelliques de la compensation de mouvement, pour le profil Baseline IPP l'ensemble des QP 27-42.

chapitre pour les prédicteurs temporels. Notons enfin que la méthode, pour les prédicteurs temporels et les prédicteurs spatiaux est décodable, et n'engendre aucun problème de lecture du flux.

## 8.4 Conclusion et perspectives

Dans ce chapitre, nous avons présenté une méthode de suppression de prédicteurs pour réduire le débit de l'information de compétition. Le critère d'élimination est basé sur la distance dans le domaine transformé entre deux prédicteurs. Cette distance est évaluée avec le processus de quantification dont les paramètres évoluent avec ceux du codage, ce qui permet à la méthode d'être adaptée aux différents débits. Les résultats présentés sur la suppression de prédicteurs des modes Intra montrent une amélioration de l'efficacité de codage par rapport à la norme H.264/AVC et à la méthode de l'Intra 1D du chapitre 6.

Les perspectives d'utilisation et d'amélioration de la méthode sont nombreuses. Des résultats préliminaires sur l'application du critère pour l'élimination de résolutions sous-pixelliques se révèlent déjà très bons pour les hautes résolutions, même en l'absence d'une recherche systématique des meilleurs paramètres. Enfin, les paramètres du critère d'élimination ont été fixés empiriquement et il serait plus intéressant de les sélectionner au fil de l'eau afin que la méthode soit adaptée au contenu de chaque séquence.

Dans le chapitre suivant, nous proposons une méthode basée elle aussi sur l'augmentation de l'intelligence du décodeur pour réduire l'information de compétition. Cette méthode utilise une estimation de mouvement reproductible au décodeur, afin de ne pas transmettre l'information de mouvement.



## Chapitre 9

# Estimation de mouvement conjointe à l'encodeur et au décodeur

*L'information de mouvement a une proportion élevée dans le débit total. Nous avons présenté une méthode de codage par compétition de prédictions de vecteurs mouvement dans le chapitre 4, afin de réduire cette information. Dans ce chapitre, nous proposons aussi de réduire l'information de mouvement mais en utilisant une démarche différente. En effet, nous considérons que l'information de mouvement est une information de compétition. Ainsi, nous proposons un critère qui permet de ne pas transmettre des résiduels de vecteurs mouvement. Ce critère est basé sur une estimation de mouvement à l'encodeur à partir d'informations déjà encodées permettant alors de ne pas transmettre le vecteur utilisé pour la compensation de mouvement du bloc courant. Le décodeur applique la même estimation de mouvement afin de déterminer ce vecteur. Ce nouveau mode de codage est mis en compétition avec les modes Inter classiques de la norme H.264/AVC à l'aide du critère de débit-distorsion. Afin d'augmenter les performances de la méthode, nous utilisons, en plus des partitions bloc de la norme, un partitionnement 1D du macrobloc.*

### 9.1 Description de la méthode

Dans le chapitre 1, nous avons vu que l'encodeur effectue une compensation de mouvement correspondant à une différence entre le bloc courant et un bloc prédicteur. Le résiduel engendré est codé (transformé et quantifié) puis transmis au décodeur (cf. figure 9.1(a)). Ce dernier décode (déquantifie et détransforme) ce résiduel, avant de l'ajouter au bloc prédicteur afin d'obtenir le bloc reconstruit (bloc original dégradé) (cf. figure 9.1(b)). Pour que le décodeur puisse déterminer le bloc prédicteur, l'encodeur transmet avec le résiduel de bloc une information de mouvement, comme on peut le voir dans la figure 9.1(a). Au niveau du décodeur (cf. figure 9.1(b)), le résiduel de

---

1. Publications de l'auteur associées : [Brev7] et [Brev3].

vecteur mouvement et l'indice d'image de référence sont lus dans le train binaire, puis sont décodés. Le meilleur bloc prédicteur provient de l'estimation de mouvement effectuée au codeur, qui est le processus permettant de réaliser une mise en correspondance entre les éléments de deux trames différentes, basée sur un critère de fidélité. Dans le standard H.264/AVC et ses prédécesseurs, seul le codeur applique le processus d'estimation de mouvement.

L'estimation de mouvement au décodeur a été proposée dans le codage vidéo distribué. Cette technique de codage provient des travaux sur la théorie de l'information de Slepian et Wolf [SW73] et Wyner et Ziv [WZ76]. Le but de ce type de codage est d'alléger la complexité de l'encodeur en augmentant celle du décodeur. Dans [PMR07] et [GARRM05], le décodeur effectue une estimation de mouvement. La sélection du meilleur bloc se fait grâce à la transmission d'informations complémentaires de l'encodeur vers le décodeur. L'estimation de mouvement est donc effectuée au niveau du décodeur. Dans la méthode proposée dans ce chapitre, la même estimation de mouvement est utilisée à la fois au codeur et au décodeur, permettant de ne pas transmettre l'information de mouvement. Le but de notre méthode est donc d'augmenter l'efficacité de codage et non de limiter la complexité au niveau du codeur.

La figure 9.2 représente le schéma de notre méthode d'estimation de mouvement conjointe à l'encodeur et au décodeur. Nous noterons EMD cette estimation de mouvement dans la suite du chapitre. Dans le schéma de l'encodeur (cf. figure 9.2(a)), nous pouvons voir que l'estimation EMD ne dépend pas de la partition courante mais des données causales de cette partition courante (partitions encodées) contrairement au schéma classique (cf. figure 9.1(a)). Le vecteur mouvement n'est donc pas codé et inséré dans le train binaire. Au décodeur (cf. figure 9.2(b)), la même estimation de mouvement EMD est effectuée avec les mêmes données en entrée. Le vecteur résultant de cette estimation de mouvement est ensuite utilisé pour la décompensation de mouvement.

Pour améliorer l'efficacité de la méthode, nous proposons aussi un partitionnement 1D du macrobloc afin de minimiser la distance spatiale entre le signal à prédire et le signal de référence utilisé pour l'appariement.

### 9.1.1 Critère d'appariement

L'estimation de mouvement au décodeur (EMD) impose de nouvelles contraintes par rapport à l'estimation de mouvement classique, implémentée généralement dans les codeurs vidéo. Dans une estimation classique, le bloc original est utilisé comme référence pour l'appariement (cf. figure 9.1(a)). Par définition du codage avec pertes, ce bloc original n'est pas disponible au décodeur. Il faut donc définir un signal qui sera utilisé pour l'appariement de l'EMD à l'encodeur et au décodeur.

Les méthodes "d'error concealment" peuvent répondre à cette contrainte du meilleur appariement temporel sans utiliser le signal de référence [TP00]. Pour le codage d'un bloc "perdu", le signal voisin de ce bloc est utilisé pour l'appariement d'une EMD. Dans [LKD04], des informations sont ajoutées dans les images reconstruites afin que l'EMD soit robuste. Dans ces méthodes, on trouve aussi des techniques de reconstruction spatiale [YML96] ou spatio-temporelle [AF95].

Dans le logiciel de référence de la norme, la sélection du meilleur vecteur mouvement (le meilleur bloc prédicteur) est basée sur le critère débit-distorsion de l'équation (2.3) du chapitre 2 ( $J = D + \lambda_{Motion} R_{mv}$ ). Pour l'estimation de mouvement utilisée dans nos configurations du logiciel de

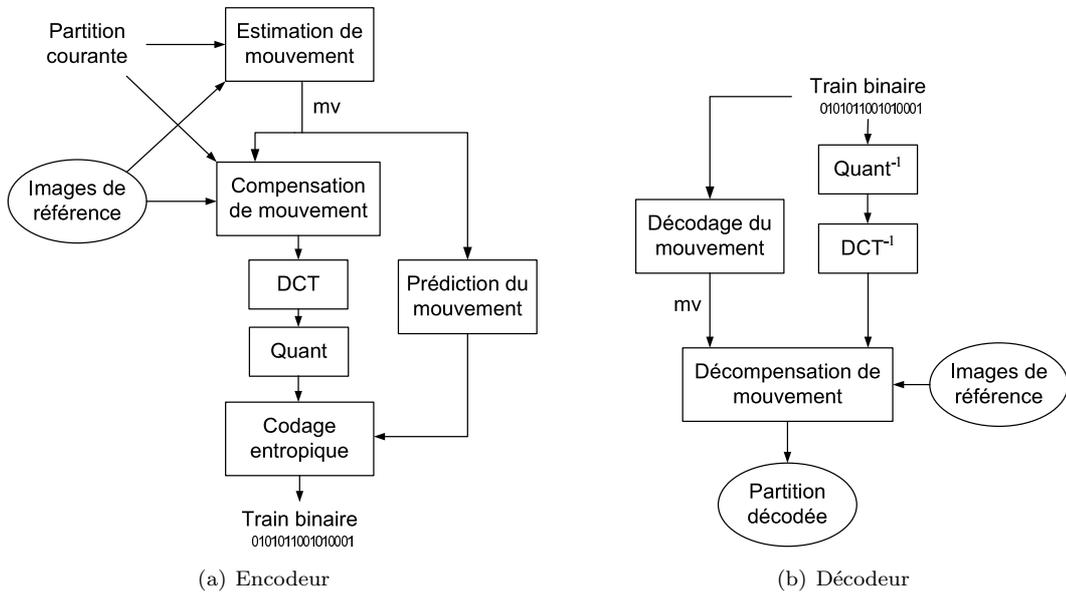


FIGURE 9.1 – Schéma d’encodage et de décodage de la prédiction Inter de type H.264/AVC.

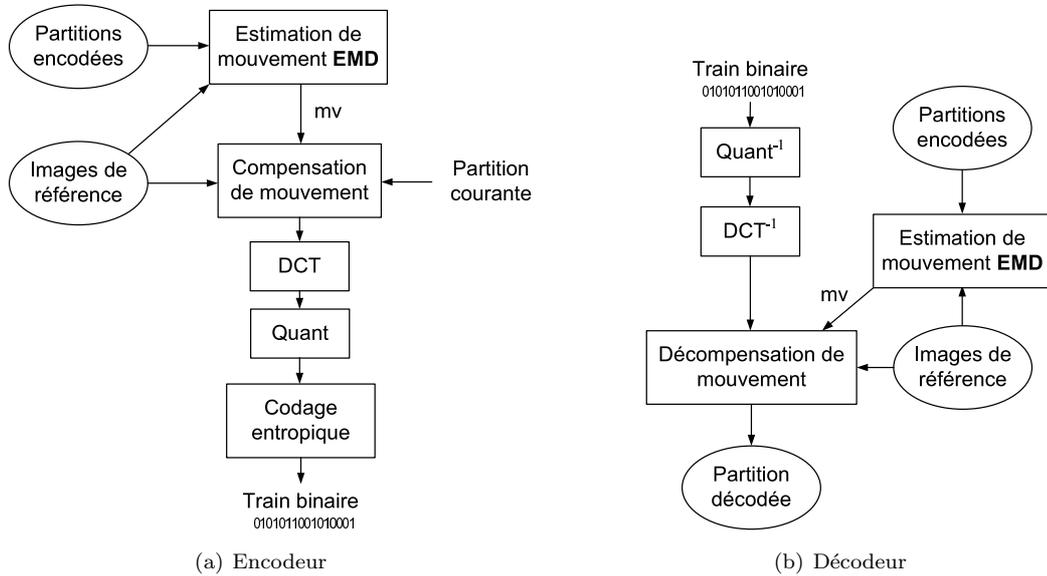


FIGURE 9.2 – Schéma d’encodage et de décodage de la prédiction Inter pour la méthode d’estimation de mouvement au décodeur proposée.

référence, la distorsion  $D$  de cette équation est un SAD entre le bloc courant et le bloc prédicteur pour l'estimation du mouvement au pixel entier, et une transformée d'Hadamard sur le résiduel du bloc courant et du bloc prédicteur pour l'estimation sous-pixellique.

Dans le cas du développement d'une EMD pour la non-transmission de l'information de mouvement, il est nécessaire que l'estimation de mouvement utilise des signaux disponibles au décodeur pour que le mouvement soit décodable. Par conséquent, le bloc courant ne peut pas être utilisé à l'encodeur. Nous utilisons les données causales du bloc courant. Dans notre développement, les lignes au-dessus du bloc courant  $l_a^i, \forall i = 1..N$  et les colonnes à gauche du bloc courant  $c_p^j, \forall j = 1..M$ , représentées dans la figure 9.3, sont utilisées comme signal à apparier. Notons que ces signaux ont déjà été décodés au moment du traitement du bloc courant. Avec ces signaux, le critère utilisé dans l'estimation de mouvement classique de l'équation (2.3) du chapitre 2 est remplacé par le critère  $J_{EMD}$  donné dans la formule suivante :

$$J_{EMD} = |l_a^1 - l_p^1| + .. + |l_a^i - l_p^i| + .. + |l_a^N - l_p^N| + |c_a^1 - c_p^1| + .. + |c_a^j - c_p^j| + |c_a^M - c_p^M| \quad (9.1)$$

où  $l_p^i, \forall i = 1..N$  et  $c_p^j, \forall j = 1..M$  sont respectivement les lignes voisines et les colonnes voisines du bloc prédicteur. Le bloc prédicteur dont les lignes et les colonnes voisines minimisent ce critère sera sélectionné. Ce critère  $J_{EMD}$  ne prend pas en compte le débit du résiduel du vecteur puisque ce dernier n'est pas transmis. L'efficacité de ce critère va dépendre du nombre de lignes  $N$  et du nombre de colonnes  $M$  utilisées. Remarquons que d'autres critères auraient pu être définis pour cette estimation de mouvement. Par exemple le bloc original pourrait être remplacé par un autre bloc, construit à partir des informations spatiales de l'image courante. La construction de ces blocs pourrait provenir de prédicteurs Intra ou d'algorithmes spatiaux d'error concealment. Cependant ces méthodes utilisent les pixels voisins du bloc courant, comme le critère  $J_{EMD}$  proposé. Ces méthodes n'utilisent pas d'information supplémentaire. Par conséquent, l'efficacité d'une telle méthode devrait être semblable au critère  $J_{EMD}$ .

### 9.1.2 Partitionnements 1D

Dans le chapitre 6 décrivant la méthode de l'Intra 1D, nous avons défendu l'idée que la réduction de la distance spatiale entre un pixel à coder et son pixel prédicteur permettait d'augmenter les corrélations spatiales moyennes et, par conséquent, de réduire l'entropie de la prédiction. Les réductions de débits, les sélections élevées et la baisse d'entropie engendrée par les modes Intra 1D ont montré que cette théorie se vérifiait en pratique. Pour l'estimation de mouvement au décodeur, le critère que nous avons sélectionné utilise les lignes et les colonnes voisines du bloc courant comme signal à apparier. Par conséquent, la méthode devrait être moins efficace pour les blocs de grandes tailles, car la distance spatiale moyenne entre les pixels à prédire du bloc original et les pixels utilisés pour l'appariement est plus élevée que pour les blocs de petites tailles. En effet, si l'on réduit la distance moyenne entre les pixels à prédire et ceux utilisés pour l'appariement, l'estimation de mouvement au décodeur devrait être plus efficace. Nous vérifierons cette hypothèse dans les résultats expérimentaux.

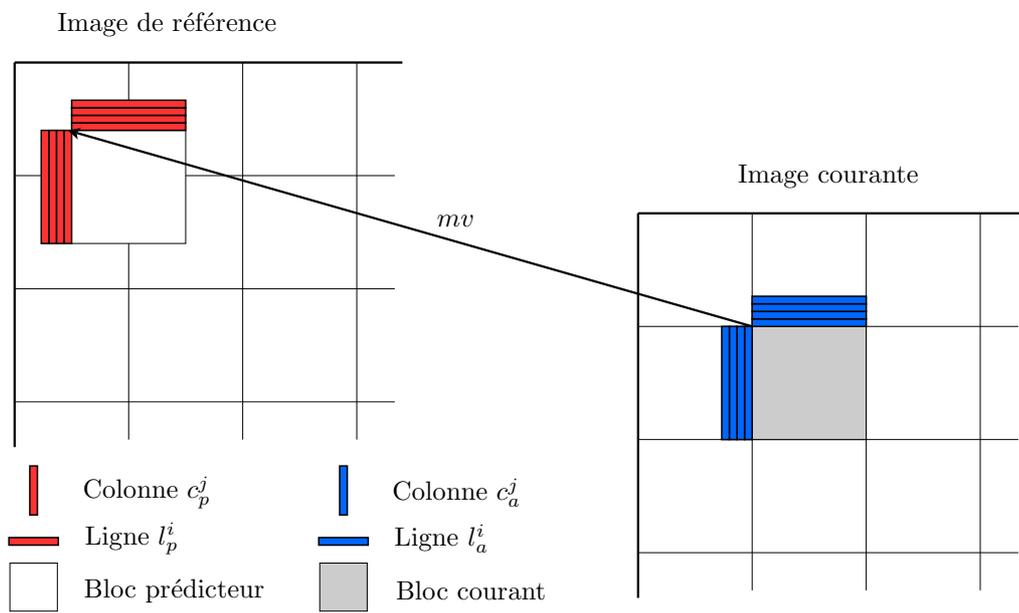


FIGURE 9.3 – Bloc courant et bloc prédicteur avec les lignes et colonnes voisines utilisées pour l'appariement dans la méthode d'estimation de mouvement au décodeur.

### 9.1.2.1 Inter 1D

La création d'un partitionnement pour l'Inter 1D est la même que celle d'un partitionnement pour l'Intra 1D. Le partitionnement 1D, décrit dans le chapitre 6, provient d'une partition initiale de forme linéaire (ligne, colonne, 'L', point, contour) sur laquelle on applique un algorithme de dilatation morphologique engendrant des partitions qui ne se chevauchent pas.

Pour l'Intra 1D, nous avons défini trois ordres de parcours des partitions 1D : le parcours lexicographique, le parcours bidirectionnel et le parcours hiérarchique. Ces parcours peuvent bien entendu être utilisés dans le cas de l'Inter 1D.

La définition de l'Inter 1D nécessite de fixer un prédicteur du vecteur mouvement. Dans le cas de l'ordre de parcours lexicographique, comme pour la définition des prédicteurs de l'Intra 1D, le prédicteur de vecteur mouvement qui est le plus adapté est le vecteur qui a été utilisé pour le codage de la partition  $Par_{i-1}$ . En faisant la même analogie avec les prédicteurs de l'Intra 1D, les partitions qui sont entourées par deux autres partitions, pour les ordres de parcours bidirectionnel et hiérarchique, ont comme prédicteur de vecteur mouvement, la moyenne des vecteurs des partitions qui entourent ces dernières. Pour les autres partitions du parcours bidirectionnel ou la première partition du parcours hiérarchique, nous pouvons prendre le vecteur de la partition précédemment encodée comme prédicteur.

Enfin, pour que l'étape de transformation des résiduels temporels de l'Inter 1D exploite au mieux les redondances spatiales, nous utilisons la transformée DCT 1D définie dans le chapitre 6 adaptée à la forme de ces partitions.

### 9.1.2.2 Inter 1D sans vecteur

L'Inter 1D trouve toute son utilité pour la méthode de suppression de l'information de mouvement au décodeur puisque chaque pixel de la partition courante a un pixel voisin qui appartient aux pixels utilisés pour l'appariement. Ce n'est pas le cas pour les partitions par bloc de la référence. Le partitionnement Inter 1D Ligne a exactement le même nombre de pixels que le mode Inter  $4 \times 4$ . Si le nombre de lignes et de colonnes utilisé pour l'appariement est égal à 1 ( $N=M=1$ ), alors la distance moyenne entre chaque pixel de la partition courante et le pixel le plus proche dans la zone d'appariement pour l'Inter 1D est inférieure à celle de l'Inter  $4 \times 4$ . Les partitions entourées par deux partitions déjà encodées de l'ordre de parcours bidirectionnel présentent d'importants avantages puisque, dans ce cas, le signal utilisé pour l'appariement entoure la partition courante, augmentant ainsi l'efficacité de l'appariement. Notons que le critère  $J_{EMD}$  défini pour les partitionnements de la norme est aussi utilisé pour les partitionnements 1D.

## 9.1.3 Sélection des modes sans vecteur

Dans l'état de l'art d'estimation de mouvement au décodeur, une seule estimation de mouvement est utilisée. La méthode de suppression de l'information de mouvement définie dans ce chapitre n'est pas efficace pour l'ensemble des blocs de l'image, comme nous le démontrons dans les résultats expérimentaux. Dans le but d'améliorer l'efficacité de codage, nous proposons donc de mettre en compétition l'estimation de mouvement proposée ci-dessus avec l'estimation de mouvement classique

à l'aide du critère débit-distorsion. Dans [YR05], le critère de sélection prend en compte à la fois l'efficacité de codage et l'efficacité de décodage en cas de pertes. Dans notre méthode, les deux prédicteurs de la partition courante, obtenus avec chacune des deux estimations de mouvement, sont comparés à l'encodeur selon l'équation suivante :

$$\min\{D_{EMD}, D_{classique} + \lambda_{Motion}R_{mv}\}, \quad (9.2)$$

où  $D_{EMD}$  est la distorsion engendrée par la meilleure partition prédictrice obtenue avec l'EMD définie dans ce chapitre et  $D_{classique}$  est la distorsion engendrée par la meilleure partition prédictrice de l'estimation de mouvement classique. Pour chaque partition, un indice correspondant à l'estimation sélectionnée par l'encodeur est transmis au décodeur. Au décodeur, si l'indice correspondant à l'estimation de mouvement classique est lu, alors le résiduel du vecteur mouvement est extrait du flux. Dans le cas contraire, la même estimation de mouvement utilisant les données voisines du bloc décodé est effectuée, afin de retrouver la même partition prédictrice obtenue par l'encodeur.

## 9.2 Résultats expérimentaux

Dans cette section, nous présentons l'efficacité de l'estimation de mouvement au décodeur (EMD) pour une recherche au pixel entier, ainsi que pour une recherche sous-pixelle, afin d'évaluer indépendamment la problématique engendrée par les sous-résolutions. Ensuite, nous étudions l'efficacité de la méthode pour les partitionnements de la norme, puis pour le partitionnement 1D, avant de fournir les résultats conjoints. Enfin, nous donnons les résultats de la combinaison de la méthode complète proposée dans ce chapitre et de la méthode MVComp proposée dans le chapitre 4. Toutes ces expériences sont menées dans le JM KTA 1.3 avec le profil Baseline IPP, pour trois ensembles de QP, sur les séquences QCIF (*Container, Foreman, Silent*) et CIF (*Paris, Foreman, Mobile, Tempete*) recommandées par le groupe VCEG. Tout d'abord, précisons nos choix d'implémentation pour la méthode proposée.

### 9.2.1 Implémentation

Dans le logiciel de référence de la norme, l'estimation de mouvement préconisée par le groupe VCEG est effectuée en deux étapes : une étape pour le pixel entier et une étape pour le  $\frac{1}{2}$  et le  $\frac{1}{4}$  de pixel. Pour l'estimation de mouvement au pixel entier, l'ensemble des positions dans la fenêtre de recherche sont testées. Pour l'estimation de mouvement sous-pixelle, l'algorithme "Enhanced Predictive Zonal Search" (EPZS) est appliqué pour les positions au  $\frac{1}{2}$ -pixel autour de la position du pixel entier sélectionnée. De même, cette recherche rapide est effectuée pour les positions au  $\frac{1}{4}$ -pixel autour de la position du  $\frac{1}{2}$ -pixel sélectionnée. Dans nos tests, nous avons désactivé l'utilisation de cet algorithme EPZS pour la référence et pour la méthode proposée. Le critère d'appariement  $J_{EMD}$  utilise les données causales du bloc courant et du bloc prédicteur. Par conséquent, nous avons désactivé l'estimation de mouvement au décodeur pour les macroblocs se trouvant en haut et à gauche de l'image, pour toutes les configurations étudiées ci-dessous.

L'implémentation de l'estimation de mouvement (classique) au pixel entier est dite "rapide". En effet, le calcul du SAD se fait dans un premier temps pour chaque bloc  $4 \times 4$ . Puis, pour chaque partition Inter, à chaque position (bloc prédicteur), le SAD est la somme des SAD de ces blocs  $4 \times 4$  pré-calculés. Pour la méthode proposée, il n'est pas possible d'utiliser ce calcul rapide, car le signal utilisé pour l'appariement est différent pour chaque partition. Par conséquent, l'implémentation de la méthode a une complexité de calcul élevée. En effet, la recherche entière (au pixel entier) implique le calcul de tous les SAD pour l'ensemble des partitionnements. L'estimation de mouvement est, dans la configuration du logiciel de référence, le processus le plus coûteux en terme de temps de calcul. La méthode proposée multiplie ce temps de calcul par le nombre de partitions utilisées. Ceci explique en partie l'utilisation d'ensembles réduits de séquences par rapport aux autres chapitres. Nous noterons que dans une implémentation réaliste d'un codec vidéo, l'estimation de mouvement a une complexité de calcul beaucoup moins élevée grâce aux méthodes d'estimation rapide du mouvement. Au niveau du décodeur, ce surcoût en complexité de calcul dépend à la fois du nombre de résiduels de vecteurs mouvement non transmis et de la complexité de l'estimation de mouvement conjointe utilisée. Les algorithmes d'estimation de mouvement rapides doivent aussi être utilisés pour l'estimation de mouvement au décodeur. Dans notre implémentation, nous utilisons une estimation EMD entière afin que le coût, en débit distorsion, de cette dernière ne soit pas pénalisé par rapport à l'estimation de mouvement classique, qui est entière dans les conditions VCEG.

Pour améliorer le compromis entre le débit du codage du résiduel d'un macrobloc (CBP + résiduel de bloc) et la distorsion engendrée par la quantification, dans le logiciel de référence, une mise à zéro des coefficients des blocs transformés quantifiés est utilisée. Cette mise à zéro dépend du nombre de coefficients différents de zéro dans ces blocs. Cette étape intervient après la quantification de chaque macrobloc. Nous avons désactivé ce choix de codage pour la méthode proposée et pour la référence, car il modifie les données causales du bloc courant après l'étape d'estimation de mouvement.

### 9.2.2 Partitionnements H.264/AVC

L'estimation de mouvement EMD est implémentée pour l'ensemble des modes Inter de la norme, excepté le mode Skip. Nous avons restreint les tests présentés ci-dessous au cas d'une seule image de référence, afin de limiter les temps de calcul. L'utilisation d'images de référence multiples augmente l'information de mouvement. En effet, pour chaque vecteur, un indice d'image de référence est transmis au décodeur. Ces indices d'images de référence peuvent aussi se déterminer automatiquement avec l'estimation de mouvement EMD. Dans cette section, nous présentons les résultats de l'estimation de mouvement au décodeur sans compétition avec l'estimation de mouvement classique (EMD seule), afin, d'une part, de déterminer les paramètres optimaux du critère d'appariement (nombre de lignes et de colonnes) et, d'autre part, de mesurer l'efficacité de la méthode sans compétition sur chacune des tailles de partition. Ensuite, nous donnons les résultats de la compétition entre les deux estimations de mouvement, avant de les analyser.

	Inter 4×4		Inter 16×16		Inter 4×4, 8×8, 16×16	
Résolution pixelique	Entière	$\frac{1}{4}$ de pixel	Entière	$\frac{1}{4}$ de pixel	Entière	$\frac{1}{4}$ de pixel
$N = M = 1$	-11,3	-47	-23,6	-81,0	-12,2	-70,8
$N = M = 4$	2,5	-6,13	-7,9	-31,0	-6,9	-20,5
$N = M = 16$	3,3	3,5	-9,1	-18,2	-9,1	-14,6

TABLE 9.1 – Pourcentage de réduction de débit obtenu lorsque seule la méthode EMD est utilisée au pixel entier et au  $\frac{1}{4}$  de pixel, pour l’Inter 4×4, l’Inter 16×16 et la combinaison de 3 modes Inter (4×4, 8×8, 16×16) en fonction du nombre  $N$  de lignes ( $l_a^i, \forall i = 1..N$ ) et  $M$  de colonnes ( $c_a^j, \forall j = 1..M$ ) utilisées pour l’appariement, pour l’ensemble des QP 22-37.

### 9.2.2.1 EMD seule

Le tableau 9.1 donne le pourcentage de réduction de débit lorsque seule la méthode EMD est utilisée. Notons que la méthode proposée et la référence utilisent les mêmes modes Inter (Inter 4×4 ou Inter 16×16 ou Inter 4×4, 8×8 et 16×16), ainsi que l’ensemble des modes Intra et le mode Skip. Pour l’Inter 16×16 et la combinaison de trois modes Inter, les résultats donnent des pertes. Cependant, des gains sont obtenus pour l’Inter 4×4 lorsque 16 lignes ( $l_a^i, \forall i = 1..16$ ) et 16 colonnes ( $c_a^j, \forall j = 1..16$ ) sont utilisées pour l’appariement, mais aussi avec 4 lignes et 4 colonnes pour la compensation de mouvement entière. Les résultats sont corrélés avec les tailles des blocs, ce qui était prévisible, puisque l’information de mouvement est beaucoup plus importante lorsque les partitions de petites tailles sont utilisées. On notera que l’estimation au pixel entier est systématiquement plus efficace que celle au  $\frac{1}{4}$  de pixel. De plus, les résultats sont en relation avec le nombre de lignes appliquées lors de l’appariement (plus le nombre de lignes augmente, plus les gains sont élevés ou plus les pertes diminuent).

### 9.2.2.2 Compétition des estimations de mouvement

**Gain sur le débit total** Le tableau 9.2 donne le pourcentage de réduction de débit de la compétition débit-distorsion entre l’EMD et l’estimation de mouvement classique pour chaque mode Inter. Dans ces configurations, seul le mode Skip et un mode Inter sont utilisés en plus des modes Intra. Au cours d’expériences préliminaires, nous avons remarqué que lorsque ces deux estimations sont en compétition, il est plus efficace de choisir  $N = M = 4$  (4 lignes et 4 colonnes) pour le calcul de  $J_{EMD}$  que de choisir  $N = M = 1$  ou  $N = M = 16$ . On remarque que les gains sont systématiques. Ces gains sont ordonnés en fonction de la taille des blocs, car plus le partitionnement du macrobloc est petit, plus la proportion de l’information de mouvement est élevée. Pour les petites tailles de bloc, l’estimation au  $\frac{1}{4}$  de pixel est plus performante que celle au pixel entier. Ces résultats prouvent l’efficacité de la méthode proposée. Remarquons que les résultats présentés incluent le débit de la compétition entre les deux estimations de mouvement.

Les résultats précédents sont intéressants mais ne sont pas réalistes, puisque seul le mode Skip et un autre mode Inter sont utilisés en plus des modes Intra. Le tableau 9.3 donne les pourcentages de

Mode Inter	Résolution pixellique	
	Entière	$\frac{1}{4}$ de pixel
16×16	0,5	0,3
16×8	1,1	1,2
8×16	1,2	1,4
8×8	1,8	2,2
8×4	3,8	4,2
4×8	3,9	4,6
4×4	4,9	5,7

TABLE 9.2 – Pourcentage de réduction de débit de la compétition entre l'estimation de mouvement au décodeur avec  $N = 4$  et  $M = 4$  et l'estimation de mouvement classique, pour la résolution au pixel entier et au  $\frac{1}{4}$  de pixel, pour chaque mode Inter, en moyenne sur l'ensemble des QP 22-37.

réduction de débit obtenus sur chaque séquence, avec la compétition débit-distorsion des deux estimations de mouvement lorsque l'ensemble des modes Inter est activé. Les gains sont systématiques. Ces gains sont ordonnés en fonction des débits (gains plus élevés à bas débit), car la proportion de l'information de mouvement est plus grande à bas débit qu'à haut débit (cf. chapitre 3). Nous avons fait la même remarque pour la méthode MVComp (cf. chapitre 4). Les gains obtenus avec la compensation de mouvement au pixel entier sont légèrement supérieurs à ceux au  $\frac{1}{4}$  de pixel. Le gain maximal est de 4% pour la séquence *Foreman* CIF à bas débit. Par rapport à la compensation de mouvement au  $\frac{1}{4}$  de pixel, qui est celle préconisée, la méthode proposée obtient en moyenne 1% de gain pour l'ensemble des QP VCEG et 1,5% à bas débit.

**Analyses** Le tableau 9.4 donne le pourcentage de sélection de l'estimation de mouvement classique (avec vecteur mouvement) et la méthode EMD (sans vecteur mouvement), pour chacune des configurations du tableau 9.2 et celle du tableau 9.3 où tous les modes sont activés. Lorsque la compensation de mouvement est au pixel entier, l'estimation de mouvement proposée est la plus sélectionnée pour l'ensemble des configurations. Pour la configuration incluant l'ensemble des modes Inter, cette sélection est de 63%. Cependant, le pourcentage est plus élevé pour les partitions de petites tailles, pour lesquelles la réduction de débit est plus forte (cf. tableau 9.2). Pour la compensation de mouvement au  $\frac{1}{4}$  de pixel, la méthode proposée est encore la plus sélectionnée pour les petites tailles de bloc. Ce pourcentage de sélection s'inverse pour la configuration incluant uniquement l'Inter 16×16. Lorsque tous les modes sont activés, le nombre de sélections de chacune des deux estimations de mouvement est équivalent. Cette proportion moins élevée de notre méthode pour la compensation de mouvement au  $\frac{1}{4}$  de pixel s'explique par la nature de l'outil de compensation de mouvement sous-pixellique. En effet, cet outil permet d'affiner le mouvement pour réduire l'erreur de prédiction. L'estimation de mouvement au décodeur n'utilise pas le bloc courant, donc le

		Résolution pixelique					
		Entière			$\frac{1}{4}$ de pixel		
	QP	17-32	22-37	27-42	17-32	22-37	27-42
Séquence	Fréquence	QCIF					
Container	15Hz	0,4	0,4	0,6	0,5	0,5	1,0
Foreman	15Hz	1,2	1,7	2,5	1,2	1,7	2,2
Silent	15Hz	1,0	1,3	1,7	0,4	0,7	1,0
		CIF					
Foreman	30Hz	1,2	2,1	4,0	1,4	2,5	4,3
Mobile	30Hz	0,4	0,4	0,5	0,4	0,4	0,4
Paris	15Hz	0,9	1,3	1,6	0,4	0,4	0,4
Tempete	30Hz	0,5	0,6	0,9	0,5	0,6	1,0
<b>Moyenne</b>		<b>0,8</b>	<b>1,1</b>	<b>1,7</b>	<b>0,7</b>	<b>1,0</b>	<b>1,5</b>

TABLE 9.3 – Pourcentage de réduction de débit de la compétition entre l'estimation de mouvement au décodeur pour 4 lignes et 4 colonnes utilisées pour l'appariement et l'estimation de mouvement classique, pour la résolution au pixel entier et au  $\frac{1}{4}$  de pixel, pour trois ensembles de QP.

Mode Inter	Résolution pixellique			
	Entière		$\frac{1}{4}$ de pixel	
	Sans MV	Avec MV	Sans MV	Avec MV
16×16	<b>62</b>	38	36	<b>64</b>
16×8	<b>68</b>	32	<b>50</b>	<b>50</b>
8×16	<b>69</b>	31	<b>51</b>	49
8×8	<b>73</b>	27	<b>59</b>	41
8×4	<b>79</b>	21	<b>69</b>	31
4×8	<b>79</b>	21	<b>69</b>	31
4×4	<b>84</b>	16	<b>75</b>	25
Tous	<b>63</b>	37	<b>50</b>	<b>50</b>

TABLE 9.4 – Pourcentage de sélection de l'estimation de mouvement au décodeur (Sans MV) pour 4 lignes et 4 colonnes utilisées pour l'appariement et l'estimation de mouvement classique (Avec MV), pour la résolution au pixel entier et au  $\frac{1}{4}$  de pixel, pour chaque configuration présentée dans le tableau 9.2 et pour la configuration incluant l'ensemble des modes Inter, en moyenne pour l'ensemble des QP 22-37.

mouvement engendré est plus "approximatif". Par conséquent, la méthode est moins efficace pour cette estimation très fine du mouvement.

Le tableau 9.5 donne la proportion de l'information de mouvement pour la méthode proposée et pour la référence en fonction des QP. On note une baisse systématique de la proportion de cette information. On remarquera que la proportion, dans le débit total, de l'indice de signalisation de l'estimation de mouvement utilisée est de 2,7%, ce qui représente plus de la moitié de la baisse de proportion du débit des résiduels de vecteurs mouvement. Pour réduire ce débit nous avons testé deux critères a priori, l'un permettant de sélectionner la meilleure des deux estimations et l'autre pour sélectionner les blocs nécessitant la compétition entre ces deux estimations. Ces critères sont basés sur les valeurs des vecteurs des blocs voisins. Les résultats engendrés sont inférieurs à ceux de la compétition avec le critère  $J_{EMD}$ .

### 9.2.3 Inter 1D

Pour nos tests sur le partitionnement 1D pour le codage Inter, nous avons sélectionné une ligne comme partition initiale  $Par_1$ . Cette partition engendre un partitionnement du macrobloc composé de 16 lignes. Nous utilisons uniquement le parcours lexicographique. Le vecteur prédicteur de la première partition,  $Par_1$ , est le médian de la norme ; pour les autres partitions,  $Par_i$ , le prédicteur est égal à la valeur du vecteur utilisé pour le codage de la partition  $Par_{i-1}$ . Comme pour l'Intra 1D, nous utiliserons des séquences composées uniquement de la composante de luminance. Avant

	H.264	Méthode proposée		
QP	Résiduel MV	Résiduel MV	Indice EM	Total
17	5,0	3,6	0,9	4,5
22	7,7	5,4	1,4	6,8
27	11,4	7,9	2,0	10,0
32	16,7	11,6	3,0	14,6
37	23,2	16,2	4,1	20,3
42	28,5	19,9	4,7	24,6
<b>Moyenne</b>	<b>15,4</b>	<b>10,8</b>	<b>2,7</b>	<b>13,4</b>

TABLE 9.5 – Proportion du débit des résiduels de vecteurs mouvement pour la référence et du débit des résiduels et des indices de signalisation de l'estimation de mouvement (EM) choisie, en fonction des six QP utilisés. Les résultats concernent la compensation de mouvement au  $\frac{1}{4}$  de pixel.

d'illustrer les gains obtenus par ce partitionnement Inter 1D avec la méthode EMD, nous présentons, dans la section suivante, les résultats de l'utilisation d'une estimation de mouvement classique de ce partitionnement.

### 9.2.3.1 Estimation de mouvement classique

L'estimation de mouvement pour notre mode Inter Ligne comporte le même nombre de tests que celui d'une estimation d'un bloc  $4 \times 4$ . Cependant un SAD est utilisé pour l'estimation de mouvement sous-pixellique à la place de la transformée d'Hadamard (pour les partitionnements bloc) car cette transformée est inadaptée aux partitionnements 1D. Pour distinguer le mode Inter Ligne des autres modes, nous ajoutons un drapeau après le codage de l'indice, spécifiant l'utilisation de l'Inter  $16 \times 16$ . C'est-à-dire que si le décodeur lit dans le flux l'indice du mode Inter  $16 \times 16$ , alors il lit le bit suivant qui indique si ce macrobloc est un Inter  $16 \times 16$  ou un Inter Ligne.

Le tableau 9.6 donne le pourcentage de réduction de débit obtenu avec l'utilisation du mode Inter 1D en compétition avec les autres modes Inter sans que la méthode EMD soit activée. On remarque que l'introduction de ce mode a très peu d'impact sur l'efficacité de codage. Cependant, de très légers gains sont obtenus pour l'ensemble des QP représentant les hauts débits alors que de légères pertes sont observées pour les bas débits.

Le tableau 9.7 donne le pourcentage de sélection en fonction des QP de l'Inter 1D. Ce mode est sélectionné dans 1% des cas et atteint 2% de sélection. Cette faible sélection était attendue, car le seul avantage que pourrait fournir ce type de partitions est une meilleure prédiction du résiduel de vecteur mouvement, grâce à la proximité spatiale des pixels de la partition à prédire et ceux de la partition contenant le vecteur prédicteur, comme pour le mode Inter  $4 \times 4$  de la norme. Cependant, ce partitionnement 1D engendre 16 résiduels de vecteur mouvement pour un macrobloc, ce qui correspond au maximum de résiduels possibles pour un macrobloc codé en Inter  $4 \times 4$  dans la norme.

		Résolution pixellique						
		Entière			$\frac{1}{4}$ de pixel			
		QP	17-32	22-37	27-42	17-32	22-37	27-42
Séquence	Fréquence	QCIF						
Container	15Hz	0,5	0,4	0,3	0,2	-0,1	-0,6	
Foreman	15Hz	0,0	-0,1	-0,4	0,4	0,0	-0,6	
Silent	15Hz	0,0	-0,2	-0,6	-0,1	-0,4	-1,2	
		CIF						
Foreman	30Hz	0,5	0,1	-0,3	0,0	0,0	0,0	
Mobile	30Hz	0,4	0,1	-0,6	0,5	0,0	-0,3	
Paris	15Hz	0,0	0,0	-0,2	0,6	0,2	-0,6	
Tempete	30Hz	0,0	0,0	-0,4	0,1	-0,1	-0,5	
<b>Moyenne</b>		<b>0,2</b>	<b>0,0</b>	<b>-0,3</b>	<b>0,0</b>	<b>-0,2</b>	<b>-0,7</b>	

TABLE 9.6 – Pourcentage de réduction de débit de la compétition entre l'Inter 1D et les autres modes Inter de la norme H.264/AVC, pour une compensation de mouvement au pixel entier et au  $\frac{1}{4}$  de pixel, pour chaque séquence, pour trois ensembles de QP.

QP	Résolution pixellique	
	Entière	$\frac{1}{4}$ de pixel
<b>17</b>	2,02	2,03
<b>22</b>	1,69	1,91
<b>27</b>	1,34	1,55
<b>32</b>	0,82	0,78
<b>37</b>	0,31	0,24
<b>42</b>	0,06	0,04
<b>Moyenne</b>	<b>1,04</b>	<b>1,09</b>

TABLE 9.7 – Pourcentage de sélection de l’Inter 1D, pour la résolution au pixel entier et au  $\frac{1}{4}$  de pixel, pour la configuration présentée dans le tableau 9.6 en fonction des QP.

L’entropie moyenne des résiduels de l’Inter 1D est théoriquement plus faible que ceux d’une partition bloc de la norme, mais le nombre de ces résiduels est élevé. L’intérêt de ce partitionnement est son utilisation dans la méthode EMD permettant de réduire la distance spatiale entre les pixels à prédire et les pixels à apparier, ainsi que les pixels prédicteurs et les pixels utilisés pour l’appariement, par rapport au partitionnement de la norme. Par conséquent, le nombre de résiduels supprimés par la méthode EMD devrait être plus élevé.

### 9.2.3.2 Inter 1D et EMD

Le tableau 9.8 donne le pourcentage de réduction de débit de la compétition entre l’estimation de mouvement classique et l’EMD lorsque l’Inter 1D est activé avec le mode SKIP et les modes Intra pour la compensation de mouvement entière et sous-pixellique. Ces résultats sont donnés en fonction du nombre de lignes utilisées pour l’appariement. Le nombre optimal est visiblement 8 lignes au dessus de la partition courante pour la compensation de mouvement au  $\frac{1}{4}$  de pixel, avec 6% de gain moyen et 2 lignes pour celle au pixel entier avec 4,6% de réduction de débit. Ces gains moyens sont du même ordre de grandeur que les résultats obtenus lorsque seul l’Inter  $4 \times 4$ , des modes Inter avec résiduel, est activé (en plus du mode Skip et des modes Intra) (cf. tableau 9.2). En effet, pour ces deux types de partitionnement l’information de mouvement est très élevée puisque 16 résiduels de vecteurs mouvement sont transmis au décodeur lorsque ces partitionnements sont sélectionnés. Par conséquent, la méthode proposée, dédiée à la réduction de cette information de mouvement, donne une efficacité de codage global plus élevée. Enfin, un gain maximum de 10% est obtenu pour la séquence *Foreman* CIF.

Le tableau 9.9 fournit le pourcentage de sélection de l’estimation de mouvement classique et de celui de l’EMD pour la configuration avec 4 lignes ( $N = 4$ ) étudiées dans le tableau 9.8. On remarque que la sélection de la méthode EMD est très élevée, avec une moyenne de 93% pour la compensation de mouvement au pixel entier (95% pour le QP 37), et de 83% de sélection pour la compensation

Nombre de lignes	Résolution pixellique	
	Entière	$\frac{1}{4}$ de pixel
$N = 1$	3,9	2,0
$N = 2$	4,6	4,1
$N = 4$	4,3	5,8
$N = 6$	3,8	5,9
$N = 8$	3,7	6,0

TABLE 9.8 – Pourcentage de réduction de débit de la compétition débit-distorsion entre la méthode EMD lorsque 1, 2, 4, 6 et 8 lignes sont utilisées pour l'appariement, et l'estimation de mouvement classique, pour la résolution au pixel entier et au  $\frac{1}{4}$  de pixel, pour chaque séquence, pour l'ensemble des QP 22-37.

de mouvement au  $\frac{1}{4}$  de pixel (92% pour la QP 37). Cette sélection est nettement supérieure à la sélection moyenne maximale obtenue pour les autres modes Inter de la norme (cf. tableau 9.4). En effet, le mode  $4 \times 4$  maximisait cette sélection avec 63%. Ceci prouve que le partitionnement Inter 1D proposé est plus efficace pour la méthode EMD que les partitions carrées de la norme, car la distance spatiale entre les pixels à prédire et le signal à apparier est en moyenne moins élevée. Ceci est vrai notamment pour le mode Inter  $4 \times$ , qui a le même nombre de pixels à coder et le même nombre de pixels utilisé pour le critère d'appariement que l'Inter 1D Ligne. Rappelons que le choix entre les deux méthodes d'estimation de mouvement se fait à l'aide du critère débit-distorsion.

#### 9.2.4 Inter bloc et Inter 1D

Le tableau 9.10 donne le pourcentage de gains obtenus avec la méthode EMD lorsque l'ensemble des modes de codage est activé en plus de l'Inter 1D et que les deux estimations de mouvement sont en compétition pour l'ensemble des modes Inter (excepté le mode SKIP). Pour les deux compensations de mouvement, les gains sont systématiques. Nous avons fait le même constat pour le partitionnement de la norme H.264/AVC (cf. tableau 9.3). Par rapport à ces résultats, le pourcentage de gain obtenu lorsque l'Inter 1D est activé est plus élevé, excepté pour les bas débits de la compensation de mouvement au  $\frac{1}{4}$  de pixel, pour lesquels on obtient une légère différence de gains (1,4% contre 1,5%). On remarque que l'augmentation d'efficacité est plus élevée pour les hauts débits que les bas débits. Les gains pour la compensation de mouvement au pixel entier sont de 1,6% pour les moyens débits et 1,9% pour les bas débits. Pour la compensation au  $\frac{1}{4}$  de pixel, les gains sont en moyenne de 1,4% pour l'ensemble des QP. Le gain maximum est encore obtenu pour la séquence *Foreman* CIF, avec 4% de réduction de débit.

Cette efficacité de codage plus importante s'accompagne d'une hausse de la sélection de l'Inter 1D par rapport à celle obtenue sans la méthode EMD (cf. tableau 9.7) comme on peut le voir dans le tableau 9.11. En effet, la sélection de mode Inter 1D a plus que doublé avec la méthode EMD

Résolution pixellique				
Entière			$\frac{1}{4}$ de pixel	
QP	Sans MV	Avec MV	Sans MV	Avec MV
<b>22</b>	<b>92</b>	8	<b>74</b>	26
<b>27</b>	<b>92</b>	8	<b>81</b>	19
<b>32</b>	<b>93</b>	7	<b>87</b>	13
<b>37</b>	<b>95</b>	5	<b>92</b>	8
<b>Moyenne</b>	<b>93</b>	7	<b>83</b>	17

TABLE 9.9 – Pourcentage de sélection de l’estimation de mouvement au décodeur (Sans MV) pour 4 lignes utilisées pour l’appariement et l’estimation de mouvement classique (Avec MV), de la configuration Inter 1D du tableau 9.8 pour la résolution au pixel entier et au  $\frac{1}{4}$  de pixel, en moyenne pour l’ensemble des QP 22-37.

		Résolution pixellique					
		Entière			$\frac{1}{4}$ de pixel		
QP		17-32	22-37	27-42	17-32	22-37	27-42
Séquence	Fréquence	QCIF					
Container	15Hz	1,6	1,5	1,9	2,3	2,0	1,5
Foreman	15Hz	1,6	2,2	2,6	1,5	1,7	1,7
Silent	15Hz	1,2	1,6	1,8	0,3	0,6	0,8
		CIF					
Foreman	30Hz	1,9	1,5	1,3	1,2	0,9	0,8
Mobile	30Hz	2,4	3,0	4,1	2,8	3,1	4,0
Paris	15Hz	0,5	0,7	0,9	0,9	0,8	0,2
Tempete	30Hz	0,6	0,8	0,9	0,9	1,0	0,8
<b>Moyenne</b>		<b>1,4</b>	<b>1,6</b>	<b>1,9</b>	<b>1,4</b>	<b>1,4</b>	<b>1,4</b>

TABLE 9.10 – Pourcentage de réduction de débit lorsque l’ensemble des modes Inter H.264/AVC et l’Inter 1D sont activés, pour la compétition entre la méthode EMD ( $N = 4$  et  $M = 4$ ) et l’estimation de mouvement classique, pour la résolution au pixel entier et celle au  $\frac{1}{4}$  de pixel, pour chaque séquence pour trois ensembles de QP.

	Résolution pixellique	
QP	Entière	$\frac{1}{4}$ de pixel
<b>17</b>	4,38	4,47
<b>22</b>	3,82	3,96
<b>27</b>	3,35	3,01
<b>32</b>	2,37	1,67
<b>37</b>	1,13	0,62
<b>42</b>	0,36	0,19
<b>Moyenne</b>	<b>2,57</b>	<b>2,32</b>

TABLE 9.11 – Pourcentage de sélection de l'Inter 1D lorsque l'ensemble des modes Inter H.264/AVC sont activés pour la compétition débit-distorsion entre la méthode EMD ( $N = 4$  et  $M = 4$ ) et l'estimation de mouvement classique, pour la résolution au pixel entier et celle au  $\frac{1}{4}$  de pixel, pour la configuration présentée dans le tableau 9.10 en fonction des QP.

passant de 1% à 2,5% de sélection et atteignant désormais 4,5% de sélection pour le QP 17.

### 9.2.5 MVComp et EMD

La méthode proposée dans ce chapitre permet de réduire le débit de l'information de mouvement. Dans le chapitre 4, nous avons proposé une méthode de codage par compétition de prédicteurs de vecteurs mouvement (MVComp) dans le même but de réduction de débit de l'information de mouvement. Dans cette section, nous présentons les résultats de l'utilisation conjointe de ces deux méthodes.

Pour le partitionnement 1D, nous n'avons pas implémenté de méthode de codage par compétition de prédiction de vecteurs mouvement. Par conséquent, nous proposons les gains de la méthode par rapport à une configuration ne comportant pas d'Inter 1D. La configuration de test utilisée dans cette section est la même que celle utilisée dans la section 9.2.2.

Le tableau 9.12 donne les résultats de l'utilisation conjointe de la méthode MVComp et de la méthode proposée, à savoir la compétition de l'EMD et de l'estimation de mouvement classique pour une compensation de mouvement au  $\frac{1}{4}$  de pixel. Les résultats moyens obtenus avec l'utilisation de chacune des méthodes sont indiqués en bas du tableau. On remarquera que la méthode MVComp engendre une légère perte pour l'ensemble des QP représentant les hauts débits. Notons que la configuration de test est différente de celle utilisée dans le chapitre 4. Notamment, une seule image de référence est utilisée et seules les basses résolutions sont testées pour lesquelles la méthode MVComp est moins performante.

En moyenne, l'utilisation conjointe des deux méthodes est plus efficace que l'utilisation d'une seule des méthodes pour les moyens et bas débits. Pour chaque séquence, pour tous les débits, cette utilisation conjointe des deux méthodes fournit un gain en réduction de débit. Le gain maximum

		RD EMD + MVComp			
		QP	17-32	22-37	27-42
Séquence	Fréquence	QCIF			
Container	15Hz	0,6	0,7	0,6	
Foreman	15Hz	0,9	1,4	2,9	
Silent	15Hz	0,4	0,6	0,7	
		CIF			
Foreman	30Hz	0,3	0,3	0,6	
Mobile	30Hz	1,2	3,8	8,7	
Paris	15Hz	0,6	1,0	2,6	
Tempete	30Hz	0,5	0,9	2,6	
Moyenne		0,6	1,2	2,7	
RD EMD		0,7	1,0	1,5	
MVComp		-0,1	0,3	1,6	

TABLE 9.12 – Pourcentage de réduction de débit de l'utilisation conjointe de la méthode proposée (RD EMD) et de la méthode MVComp, pour trois ensembles de QP.

est de 8,7% pour la séquence *Foreman* CIF à bas débit. Pour l'ensemble de QP recommandés par le groupe VCEG, 1,2% de gain est obtenu contre 0,3% pour MVComp et 1% pour la méthode proposée. Pour les bas débits, l'utilisation conjointe atteint 2,7% de gain en moyenne.

Cette réduction de débit de l'utilisation conjointe des deux méthodes est naturellement liée à une réduction de la proportion de l'information de mouvement dans le débit total, comme on peut le voir dans la figure 9.4. Notons que la proportion de cette information de mouvement est donnée par QP, pour la référence et la méthode proposée dans le tableau 9.5. Les deux méthodes fournissent une proportion de l'information de mouvement équivalente. Cependant, la proportion de débit des résiduels de vecteurs subit une plus forte réduction avec la méthode proposée qu'avec la méthode MVComp. Le coût de l'indice spécifiant l'estimation de mouvement utilisée a une proportion plus élevée dans le débit total que les indices des prédicteurs de vecteurs mouvement. Ceci est notamment lié à la suppression des indices de prédicteurs par égalité des prédicteurs. L'utilisation conjointe des deux méthodes obtient la plus faible des proportions de débit de l'information de mouvement dans le débit total avec 12,3%. La proportion du débit des résiduels de vecteur est de 8,8% contre 15,4% dans la référence.

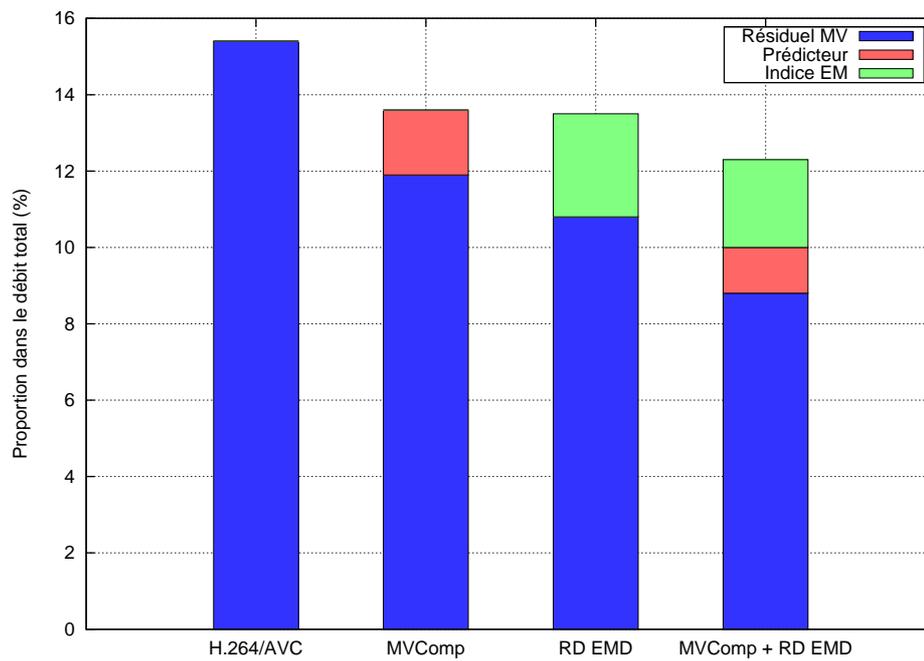


FIGURE 9.4 – Proportion du débit des résiduels de vecteurs mouvement, des résiduels et des indices de l'estimation de mouvement (EM) et de prédicteurs, pour la référence H.264/AVC, la méthode MVComp seule, la méthode proposée (RD EMD) et l'utilisation conjointe de MVComp et la méthode proposée, en moyenne pour l'ensemble des séquences et des six QP utilisés.

### 9.3 Conclusion et perspectives

Dans ce chapitre, nous avons proposé une nouvelle méthode dédiée à la réduction du débit de l'information de mouvement. La démarche adoptée est différente de la compétition de prédictions de vecteurs mouvement proposée dans le chapitre 4 puisque nous considérons l'information de mouvement comme une information de compétition. En effet, nous n'exploitons pas les redondances des champs de vecteurs mouvement, mais les corrélations temporelles entre les images déjà décodées. Dans ce sens, une estimation de mouvement reproductible au décodeur a été mise en place.

Afin d'améliorer cette estimation de mouvement au décodeur, nous avons utilisé un nouveau partitionnement du macrobloc, composé de partitions 1D. L'Inter 1D permet d'obtenir une distance spatiale plus faible entre le signal utilisé pour l'appariement et la partition à prédire et, par conséquent, d'augmenter les corrélations.

Les résultats obtenus sur un ensemble réduit de séquences sont encourageants et dépassent les gains fournis par la méthode MVComp qui a été intégrée dans le JM KTA. L'estimation de mouvement au décodeur engendre un coût supplémentaire en complexité pour l'encodeur et le décodeur. Au niveau du décodeur, ce surcoût en complexité de calcul dépend à la fois du nombre de résiduels de vecteurs mouvement non transmis et de la complexité de l'estimation de mouvement conjointe utilisée. Pour que la méthode puisse être utilisée dans un décodeur temps réel il est donc nécessaire de considérer une estimation de mouvement rapide pour l'estimation EMD. L'augmentation des puissances de calcul et les algorithmes rapides d'estimation de mouvement devraient permettre d'obtenir un encodage et un décodage en temps réel de la méthode.

L'introduction de l'Inter 1D Ligne dans ce schéma de codage a montré son efficacité. Une première perspective pour améliorer les résultats de la méthode est d'implémenter d'autres partitionnements du macrobloc et d'utiliser les autres ordres de parcours des partitions. L'Inter 1D pourrait constituer, à lui seul, un schéma d'amélioration du codage de l'information de mouvement. Nous avons noté que la seule implémentation du mode Inter Ligne avec l'ordre de parcours lexicographique compense déjà le coût de l'indice nécessaire au codage de ce nouveau mode Inter à haut débit. Une implémentation complète de ce mode avec une mise en place de regroupements ou d'éliminations de résiduels (par estimation de mouvement au décodeur par exemple) devrait améliorer le schéma de la norme H.264/AVC. Cette perspective est en cours d'étude par notre équipe.

La méthode d'estimation conjointe à l'encodeur et au décodeur peut facilement être utilisée pour la suppression des indices d'images de référence. Ceci permettrait de réduire encore l'information de mouvement. Nous pouvons aussi prévoir d'utiliser cette estimation de mouvement pour le mode Skip.

D'autres perspectives sont envisageables. On pourrait appliquer la méthode sur une seule des composantes (X ou Y), afin d'augmenter l'efficacité de codage et de réduire la complexité du décodeur. On peut aussi envisager d'appliquer l'estimation conjointe uniquement sur la composante entière du mouvement ou uniquement sur la composante sous-pixellique. On pourrait utiliser le vecteur mouvement généré par cette estimation de mouvement comme prédicteur du vecteur mouvement obtenu par l'estimation de mouvement classique. Ces deux dernières perspectives ont déjà été investiguées et sont décrites dans l'annexe C.

Pour un futur standard, on peut imaginer que le mouvement translationnel soit remplacé par

un mouvement affine. Dans ce cas, l'estimation de mouvement conjointe pourrait être utilisée pour prédire ce mouvement affine ou uniquement la composante de rotation ou uniquement la composante de translation. Dans un codec scalable, ce type de méthode pourrait être utilisé pour coder une seule couche : soit le mouvement de la couche de base, soit celui d'une couche d'amélioration. De même, dans un codec multivues, la méthode pourrait permettre de déterminer le mouvement d'une vue à partir de données causales de la vue courante, ou image courante et celles des autres vues. Enfin, on peut considérer plusieurs critères d'appariement pour l'estimation de mouvement conjointe, qui seraient mis en compétition. On transmettrait alors au décodeur un indice correspondant au meilleur critère d'appariement au lieu de transmettre une information de mouvement.

Le transfert ou, plus précisément, l'équilibre d'intelligence entre l'encodeur et le décodeur est une voie de recherche intéressante, à la frontière entre l'amélioration incrémentale et la rupture technologique du schéma de codage de référence. Les résultats obtenus dans ces trois derniers chapitres ont permis de réduire de façon systématique et significative l'information de compétition, offrant ainsi des perspectives de recherches encourageantes.

Quatrième partie

Résultats globaux



## Chapitre 10

# Intra 1D, MVComp et suppression de prédicteurs

*Dans ce chapitre, nous proposons de mesurer l'efficacité de codage de l'utilisation conjointe de certaines des méthodes étudiées dans ce manuscrit. Le but est de fournir un résultat global de nos contributions remplissant l'objectif principal de la thèse qui est, rappelons-le, de fournir des méthodes normalisables, suffisamment efficaces et réalistes pour être intégrées dans une éventuelle future norme, notamment le successeur d'H.264/AVC. Dans cette combinaison, nous utiliserons l'Intra 1D, la suppression d'indices de prédicteurs Intra et la méthode de compétition de vecteurs mouvement, ainsi que la méthode de suppression d'indices de prédicteurs liés à cette compétition.*

### 10.1 Configurations de tests

Les expériences ont été menées dans le JM KTA 1.4 avec les profils Baseline IPP et High IBBP. La méthode Intra 1D et celle de la suppression des indices de prédicteurs Intra sont aussi activées pour les slices P et B car le codage Intra est autorisé pour ces slices Inter dans les recommandations VCEG. Pour ces deux méthodes, le nombre de prédicteurs et les paramètres sont identiques à ceux définis dans le chapitre 6 et le chapitre 8. De même, les combinaisons de prédicteurs qui ont fourni les meilleures configurations dans le chapitre 4 sont utilisées pour la méthode de compétition de prédiction de vecteurs mouvement et celle de la suppression des indices de ces prédicteurs de vecteurs mouvement. Remarquons que les résultats de la combinaison de l'Intra 1D et de la méthode de suppression d'indices de prédicteurs Intra sont présentés dans le chapitre 8. De même, dans le chapitre 7, nous avons constaté que la méthode de suppression d'indices de prédicteurs pour la méthode de codage par compétition de vecteurs mouvement fournit des gains systématiques par rapport à cette dernière. Dans ce chapitre, nous présentons la combinaison de nos travaux sur le codage Intra et sur le codage Inter. Cependant, pour des raisons d'implémentation, nous présentons uniquement la combinaison de ces quatre méthodes et non des six décrites précédemment.

Pour les résultats présentés ci-dessous, trois ensembles de QP ont été testés et les séquences

sont codées uniquement sur la composante de luminance. Les expériences ont été effectuées pour un GOP de 13 images afin, d'une part, de tester une configuration réaliste de GOP et, d'autre part, d'équilibrer l'impact de nos améliorations du codage Intra, sur le débit total, face aux modifications du codage Inter. Pour le profil Baseline IPP, la première image est une image I et les 12 autres images sont des images P. Le profil High IBBP est composé pour ce GOP 13, d'une image I, de quatre images P et de huit images B. Dans ce chapitre, l'ensemble des options recommandées par le groupe VCEG (cf. annexe A) est utilisé et les gains sont donnés en pourcentage de réduction de débit calculés avec la métrique de Bjøntegaard [Bjø01a].

## 10.2 Gains séquence par séquence

Le tableau 10.1 donne le pourcentage de réduction de débit pour les deux profils testés, séquence par séquence, pour les trois ensembles de QP. Les performances sont plus importantes pour le profil Baseline que pour le profil High. Nous avons fait le même constat pour chacune des méthodes indépendamment, en précisant que le profil High était plus performant que le profil Baseline et qu'il était donc plus difficile de fournir une amélioration incrémentale du profil High. Remarquons aussi que le nombre de prédicteurs de la méthode Intra 1D a été fixé dans une configuration Baseline, ce nombre n'est donc pas optimal pour le profil High.

Pour l'ensemble des séquences, les gains sont systématiques pour les deux profils. Pour le profil High les gains sont respectivement de 3,2%, 5,4%, et 8% pour les hauts, moyens et bas débits. Pour le profil Baseline, ces gains sont de 5%, 8,4% et 13,1%. Les gains sont plus élevés à bas débit, comme dans le cas de l'utilisation indépendante de ces méthodes, car elles (excepté l'Intra 1D) sont dédiées à l'amélioration du codage de l'information de compétition et de l'information de mouvement, et celles-ci ont une plus forte représentation à bas débit. Le gain minimum de 1,1% est observé pour la séquence *Foreman* QCIF 15hz à haut débit pour le profil High. A bas débit pour ce profil, la séquence qui obtient la plus forte réduction de débit avec 17,8% est la séquence *Schumacher* CIF 30 Hz. Cette séquence a obtenu la meilleure efficacité de codage pour l'Intra 1D. Pour toutes les autres configurations, la séquence *Jets* 720p 60Hz obtient la meilleure réduction de débit, atteignant 25,4% pour le profil Baseline à bas débit. Rappelons qu'il s'agit d'une moyenne pour quatre QP. Enfin, notons que la réduction minimale est de 4,6% pour le profil Baseline à bas débit.

## 10.3 Gains par résolution

L'efficacité de codage de la combinaison de ces méthodes est corrélée avec la résolution. La figure 10.1 illustre cette remarque. Le gain sur les séquences hautes définitions est supérieur aux gains sur les basses résolutions. Bien entendu, comme nous avons pu l'expliquer dans les chapitres décrivant chacune de ces méthodes, ce n'est pas forcément l'augmentation de la résolution spatiale qui est corrélée avec les gains mais le contenu des séquences qui est différent pour chaque résolution et la distance temporelle qui augmente généralement avec la résolution spatiale.

Pour les résolutions 1080p, le gain moyen à bas débit est de 10% pour le profil High et de 19,3% pour le profil Baseline. On peut imaginer que la normalisation d'un futur standard vidéo, ou

		Baseline			High		
		17-32	22-37	27-42	17-32	22-37	27-42
Séquence	Fréquence	QCIF					
Container	15Hz	1,8	3,2	4,7	2,5	4,5	5,4
Foreman	15Hz	2,6	4,4	7,9	1,1	2,6	4,9
Silent	15Hz	2,6	3,6	5,2	2,3	3,8	5,2
		CIF					
Foreman	30Hz	3,6	6,8	11,1	2,2	3,7	5,6
Mobile	30Hz	2,0	3,0	4,6	2,7	3,8	5,0
Paris	15Hz	2,8	4,1	6,1	2,8	3,7	5,1
Tempete	30Hz	1,3	2,0	4,7	2,2	3,5	4,8
Carphone	30Hz	2,8	5,2	9,6	2,3	4,1	6,2
Crew	30Hz	4,8	7,6	11,4	3,5	5,7	7,8
Ice	30Hz	5,1	6,6	8,7	3,1	4,1	5,6
Schumacher	30Hz	9,8	14,5	19,8	8,7	12,8	17,8
		4CIF					
City	30Hz	3,3	5,6	10,6	1,5	1,9	3,9
Ice	30Hz	4,8	6,6	9,3	3,0	4,2	5,7
Soccer	60Hz	3,3	5,2	8,9	1,4	2,5	4,4
		720p					
Bigships	60Hz	4,5	9,5	19,2	3,9	5,6	8,5
Shuttlestart	60Hz	4,9	7,5	11,0	2,2	4,4	7,4
City_corr	60Hz	4,3	6,9	13,3	2,0	3,4	6,4
Crew	60Hz	6,9	10,6	14,2	2,8	5,8	8,9
Night	60Hz	3,1	4,8	8,1	2,2	3,3	4,7
Raven	60Hz	6,7	12,1	22,5	3,1	5,7	9,2
Stockholm	60Hz	3,0	6,9	12,0	1,1	5,0	8,8
Cyclists	60Hz	4,2	9,3	17,5	2,1	5,0	8,8
Jets	60Hz	15,7	23,0	25,4	9,1	12,9	16,1
Panslow	60Hz	3,8	11,5	17,3	3,8	9,4	13,1
Sailormen	60Hz	5,3	9,0	15,3	2,8	4,9	9,1
Spincalendar	60Hz	5,8	10,1	15,4	4,7	9,3	11,2
		1080p					
Blue sky	25Hz	10,3	13,5	16,1	4,9	5,6	7,2
Intotree	50Hz	2,5	8,2	20,6	1,8	3,9	11,3
Station2	25Hz	10,0	14,8	21,1	5,4	7,5	10,4
Toys and calendar	60Hz	7,1	14,3	21,4	3,7	8,0	10,9
Moyenne		5,0	8,4	13,1	3,2	5,4	8,0

TABLE 10.1 – Pourcentage de réduction de débit de chaque séquence pour la configuration Baseline GOP 13 IPP et la configuration High GOP 13 IBBP, pour trois ensembles de QP.

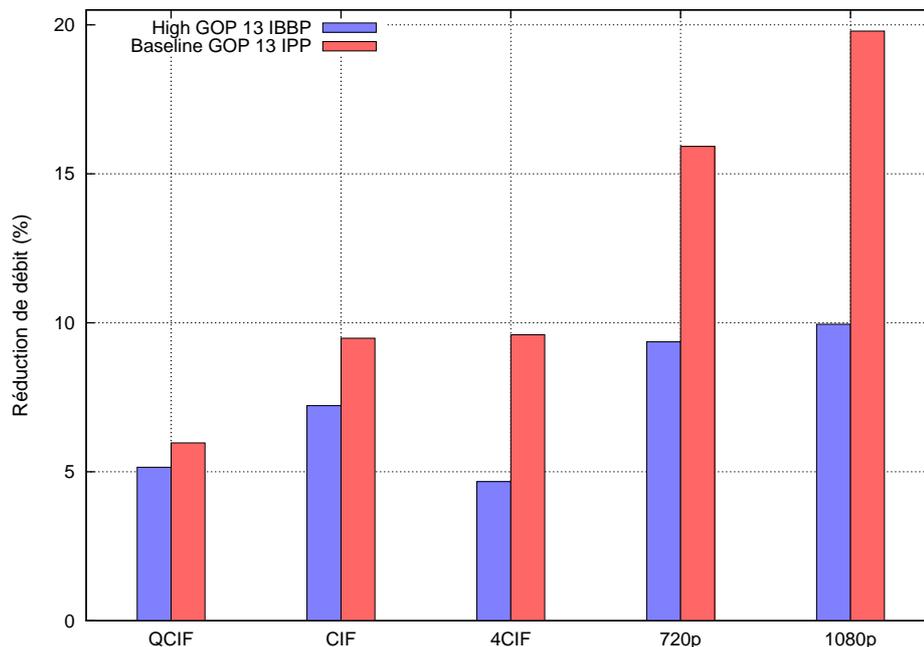


FIGURE 10.1 – Pourcentage de réduction de débit en fonction des résolutions pour les deux configurations testées pour l'ensemble des QP 27-42.

du moins un appel à contributions, s'accompagnera d'un besoin d'augmentation de l'efficacité de codage engendrée par H.264/AVC. Aujourd'hui, les écrans capables d'afficher les résolutions 1080p représentent la majorité de l'offre des téléviseurs traditionnels. La commercialisation des écrans 2060p est déjà prévue par les industriels. Par conséquent, les recommandations du remplaçant de la norme H.264/AVC imposeront sans doute d'améliorer l'efficacité de codage sur ces hautes et très hautes résolutions. La combinaison de ces méthodes répond alors à ce besoin.

## 10.4 Efficacité de codage par slices

Dans le tableau 10.2, nous avons reporté les pourcentages de réduction de débit moyen de chacun des types de slices pour les deux configurations. On remarque que, à bas débit pour le profil Baseline, la réduction de débit pour les deux types de slices est inférieure à la réduction de débit moyenne obtenue pour l'ensemble des séquences. Il ne s'agit pas d'une erreur. En effet, d'une séquence à l'autre la proportion de débit d'une image I par rapport aux images P peut varier fortement en fonction de la texture et du mouvement contenus dans ces séquences. Par conséquent, le ratio débit image I sur débit image P est très variable.

Pour le profil Baseline, la réduction de débit est plus élevée pour les slices I que pour les slices P. Pour le profil High, les slices B obtiennent une meilleure efficacité de codage que les autres types de slices, avec 10,7% pour les bas débits. Cette réduction est plus élevée que celle obtenue pour les slices

				<b>QP</b>		
				17-32	22-37	27-42
<b>Type de slice</b>	<b>Baseline</b>					
I	6,6	9,5	12,7			
P	4,1	6,8	12,0			
				<b>High</b>		
I	3,3	4,5	5,7			
P	2,5	4,7	8,4			
B	5,1	8,0	10,7			

TABLE 10.2 – Pourcentage de réduction de débit de chaque type de slice pour la configuration Baseline GOP 13 IPP et la configuration High GOP 13 IBBP, pour trois ensembles de QP.

P du profil Baseline. On remarque que les gains sont deux fois plus grands pour les slices I du profil Baseline que pour ceux du profil High. Ceci est lié, entre autres, à l'efficacité de l'Intra  $8 \times 8$  qui n'est pas utilisé dans le profil Baseline. Nous avons remarqué que ce mode limite les performances de nos méthodes. De plus, pour l'Intra 1D, la recherche des meilleurs paramètres a été effectuée pour le profil Baseline. Une sélection des prédicteurs optimaux pour le profil High devrait augmenter les performances de cette méthode. Pour le profil High, l'amélioration des slices P à moyen et bas débits est plus grande que celle des slices I. Cependant, à haut débit, les slices P obtiennent la plus faible augmentation d'efficacité de codage avec 2,5%. Ce résultat provient de la faible proportion de l'information de mouvement pour ces débits dans ce profil. En effet, la distance temporelle élevée entre les images P engendre des erreurs de prédiction de blocs plus fortes, augmentant ainsi la proportion de ces dernières dans le débit total.

## 10.5 Courbe de débit-distorsion

La figure 10.2 représente les courbes débit-distorsion de six séquences à bas débit, où la combinaison des méthodes proposée est la plus efficace. On remarque que l'efficacité de codage de la combinaison des méthodes en profil Baseline, pour la séquence Schumacher CIF 30Hz, est supérieure à l'efficacité du profil High de la norme. De même, les courbes de ces deux profils sont confondues pour la séquence *Ice* CIF 30 Hz. On voit nettement sur ces courbes que l'amélioration de l'efficacité de codage est plus élevée pour le profil Baseline. De plus cette efficacité augmente à très bas débit (pour le QP 42), comme on peut le remarquer pour la séquence *Raven* 720p ou les deux séquences 1080p.

## 10.6 Conclusion

Dans ce chapitre nous avons mesuré l'efficacité de codage de l'utilisation conjointe d'une partie des méthodes détaillées dans ce manuscrit. Les méthodes utilisées sont celles qui atteignent un stade d'étude suffisamment avancé pour être proposées au groupe VCEG. D'ailleurs, l'une d'entre elle

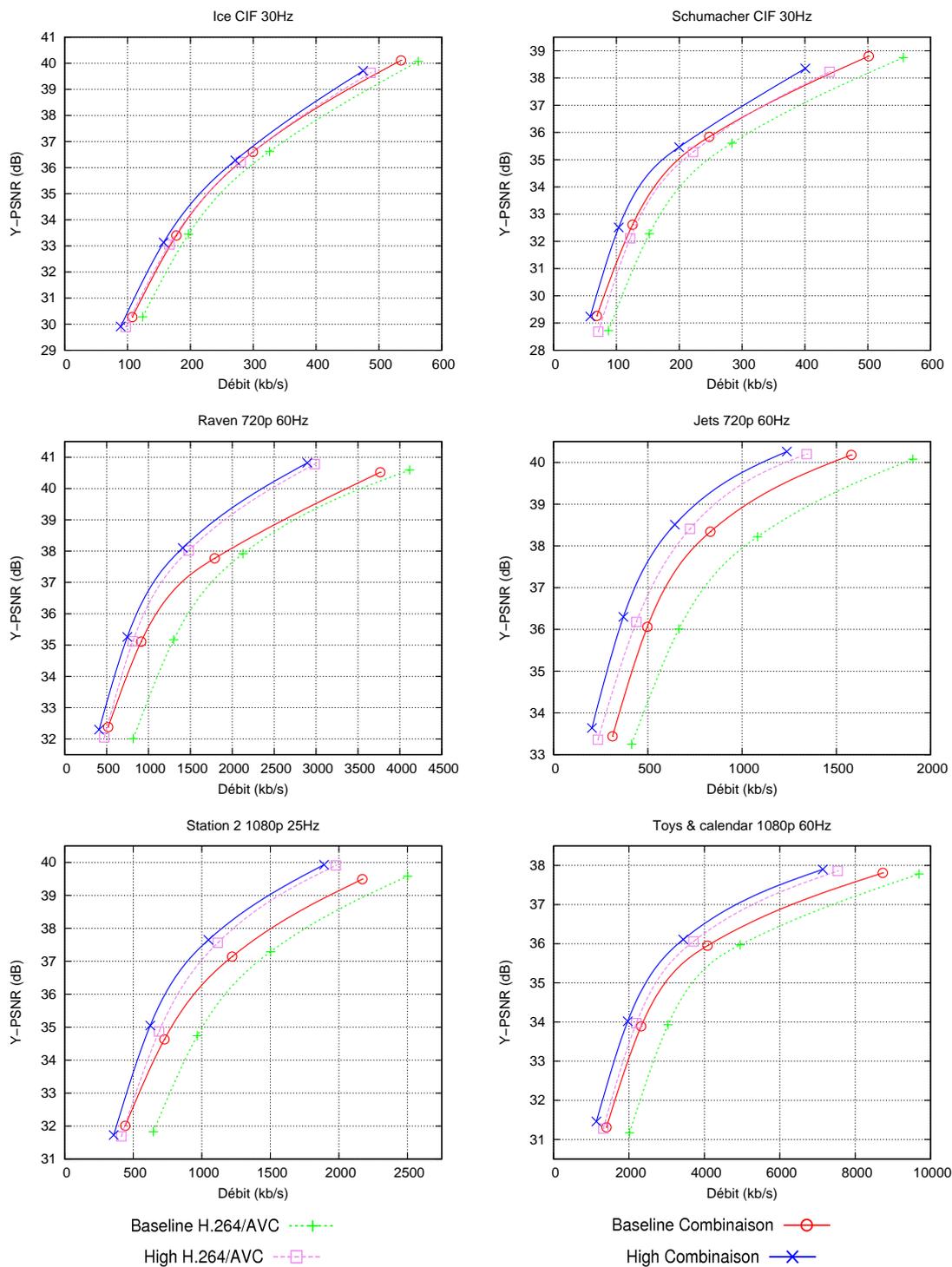


FIGURE 10.2 – Courbes débit-distorsion pour la référence H.264/AVC Baseline GOP 13 IPP et High GOP 13 IBBP et pour la combinaison des méthodes Baseline GOP 13 IPP et High GOP 13 IBBP pour six séquences pour les QP 27, 32, 37, 42.

est déjà intégrée dans le JM KTA. Les gains engendrés par cette utilisation conjointe sont élevés, avec 13,1% de réduction de débit pour le profil Baseline à bas débit et 8% pour le profil High avec l'ensemble des options recommandées par le groupe VCEG. Ces résultats prouvent, entre autres, la complémentarité des méthodes proposées. De plus, les gains obtenus sur les séquences haute définition sont supérieurs aux gains moyens, atteignant près de 20% pour le profil Baseline. Cette performance pour ces grandes résolutions augmente encore les chances d'adoption des méthodes dans un éventuel remplaçant de la norme H.264/AVC.



# Conclusion

L'objectif principal de cette thèse est de fournir des méthodes de codage suffisamment efficaces et réalistes pour être intégrées dans un successeur du standard de référence H.264/AVC. L'orientation choisie est le codage par compétition. Grâce à l'étude comportementale du logiciel de référence de la norme, nous avons identifié les déficiences de celle-ci et déterminé deux grands axes de recherche, qui sont la mise en place de nouveaux modules de codage par compétition et la réduction de l'information de compétition par augmentation de l'intelligence du décodeur.

## Synthèse des travaux

### Modules de codage par compétition

La forte proportion de l'information de mouvement à bas débit provient, d'une part, de la réduction de la proportion des résiduels de texture due à la forte quantification et, d'autre part, de l'augmentation de l'entropie des résiduels de vecteurs mouvement liée à une distance spatiale moyenne plus élevée entre les vecteurs à coder et les vecteurs utilisés pour le calcul de leurs prédicteurs. Afin de réduire l'entropie de ces résiduels de vecteurs, nous avons proposé d'exploiter à la fois les corrélations spatiales et les corrélations temporelles entre les champs de vecteurs mouvement, à l'aide d'une compétition de prédiction des vecteurs mouvement basée sur le critère débit-distorsion. De plus, une compétition de vecteurs prédicteurs pour le mode Skip, basée elle aussi sur le critère débit-distorsion, a été développée, permettant ainsi d'augmenter le nombre de sélections de ce mode de codage très efficace. Enfin, pour réduire l'information de compétition liée à cette méthode, les indices de prédicteurs sont transmis uniquement lorsque les vecteurs sont différents. L'impact de l'exploitation des corrélations temporelles est remarquable à bas débit, puisque la distance entre le vecteur courant et les prédicteurs temporels n'évolue pas en fonction du débit visé. La méthode améliore donc significativement la référence, notamment à bas débit où la proportion de l'information de mouvement est élevée. Le groupe VCEG a jugé cette méthode suffisamment efficace et peu complexe pour être intégrée dans le JM KTA. De plus, la méthode offre des gains systématiquement plus élevés lorsqu'elle est comparée aux autres outils déjà intégrés dans ce logiciel, ce qui augmente ses chances d'être adoptée dans un futur standard.

L'ensemble des prédicteurs pour le mode Skip qui maximise les gains en moyenne n'est pas nécessairement le meilleur pour chaque séquence. Nous avons donc proposé une adaptation de l'ensemble des prédicteurs pour le mode Skip image par image. Le critère a priori d'ajout d'un prédicteur

dans l'ensemble utilisé est basé sur le nombre de fois où ce prédicteur a minimisé le critère débit-distorsion pour l'ensemble des macroblocs de l'image précédente. En remarquant que le mode Skip est fortement sélectionné pour les zones comportant peu ou pas de mouvement, nous avons proposé une adaptation orientée contenu des ensembles de prédicteurs pour ce mode. Dans cette méthode, deux ensembles sont utilisés au lieu d'un seul et chacun évolue indépendamment avec des paramètres différents. Le but est d'utiliser un nombre réduit de prédicteurs dans les zones où le mode Skip a une forte probabilité de sélection, afin de réduire le débit des indices de prédicteurs et d'utiliser un nombre élevé de prédicteurs dans les zones où le mode Skip a une faible probabilité de sélection, pour augmenter l'occurrence de ce mode pour ces zones. Les résultats ont montré qu'effectivement le nombre de sélections du mode Skip augmente avec cette adaptation et que le débit des indices est réduit pour l'ensemble contenant peu de prédicteurs. Le nombre et le type de prédicteurs sélectionnés dans chacun des ensembles dépendent fortement de la séquence codée. Enfin, l'efficacité de codage est accrue par rapport à la méthode classique de compétition.

L'étude comportementale du logiciel de référence de la norme a montré que la proportion des images Intra en termes de débit-distorsion était élevée dans un flux correspondant à une application de diffusion. Une amélioration du codage Intra est alors nécessaire pour améliorer l'efficacité de codage de la norme H.264/AVC. Dans le codage prédictif par bloc, l'erreur de prédiction des pixels en bas à droite d'un bloc est généralement plus grande que l'erreur des pixels en haut à gauche, car la distance spatiale entre les pixels à prédire et les pixels de référence est plus grande. Pour résoudre ce problème nous avons proposé un partitionnement 1D des macroblocs, pour lequel chaque pixel à prédire a au moins un pixel voisin dans le signal de référence. Le partitionnement du macrobloc est obtenu par dilatation morphologique d'une partition linéaire initiale. Afin d'augmenter les performances de la méthode, nous avons utilisé, en plus du parcours lexicographique, un parcours bidirectionnel et un parcours hiérarchique des partitions. Ces deux derniers sont basés sur le fait qu'il est parfois plus efficace, au sens du critère débit-distorsion, de coder certaines partitions avec un signal de référence plus lointain, puis de traiter les autres avec des corrélations spatiales plus grandes. Ces parcours ont montré leur efficacité dans l'exploitation des redondances temporelles entre les images. Des prédicteurs adaptés à ces partitions 1D et à ces ordres de parcours ont été spécialement définis. Le nombre optimal de prédicteurs à utiliser pour chacun des sous-modes a été déterminé empiriquement dans une étude préalable. Enfin, les résiduels de ces partitions sont codés à l'aide d'une transformée DCT 1D afin que ce processus soit adapté à ces signaux 1D. Le mode Intra 1D obtient une forte sélection par rapport aux modes Intra de la norme. L'entropie des résiduels est en dessous de celle obtenue par le codage prédictif des blocs. Par conséquent, la méthode engendre des gains systématiques et élevés par rapport à la référence.

Nous avons donc montré que l'ajout de nouveaux modules de codage par compétition est une voie de recherche qui apporte des gains par rapport à la norme de référence H.264/AVC dont l'efficacité provient entre autres d'un nombre très élevé de possibilités de codage mis en compétition.

## Réduction de l'information de compétition

Le nombre pléthorique de possibilités de codage de la référence engendre une forte proportion de l'information de compétition, représentant plus de la moitié d'un flux à moyen et bas débits. Il est

évident que la réduction de cette information va devenir nécessaire dans le cadre d'une amélioration incrémentale de la référence. Nous avons proposé d'augmenter l'intelligence du décodeur afin qu'il puisse reproduire certains des choix de l'encodeur. Les choix reproduits par le décodeur n'ont pas besoin d'être transmis par l'encodeur, ce qui permet de réduire le débit lié à la compétition.

Dans un premier temps, nous avons exposé une méthode permettant d'éliminer les indices des prédicteurs de vecteurs mouvement. En effet, dans certains cas, le décodeur est capable de "deviner" le prédicteur utilisé, s'il a connaissance à la fois du résiduel du vecteur courant, des prédicteurs testés et du critère de choix. Le nombre d'indices non transmis dépend de ces trois paramètres. La méthode permet de réduire significativement le débit des indices de prédicteurs. Cette méthode simple est cependant applicable uniquement aux données codées sans perte. Pour les données dont le critère de sélection dépend d'une perte non connue au décodeur, d'autres stratégies doivent être mises en place.

Nous avons donc développé un critère d'élimination de prédicteurs de texture. Ce critère est basé sur la distance entre les prédicteurs dans le domaine transformé. L'évaluation de cette distance se fait à l'aide du processus de quantification et dépend du pas de quantification utilisé pour le codage. L'idée est d'éliminer un prédicteur si un autre prédicteur lui ressemble trop. C'est-à-dire que l'utilisation conjointe des deux prédicteurs n'est pas pertinente pour le codage. Cette pertinence dans le cas du codage prédictif de la texture dépend à la fois du signal à prédire et de la qualité visée. Le critère est entièrement reproductible au décodeur. Dans le cadre du codage Intra, pour chaque partition, le processus d'élimination est appliqué sur l'ensemble des prédicteurs prédéfinis. A la fin de ce processus, le nouvel ensemble, composé d'au moins un prédicteur, est utilisé pour le processus de sélection de la meilleure prédiction au niveau de l'encodeur ou pour la lecture du prédicteur au niveau du décodeur. La méthode permet de réduire significativement le nombre de bits utilisés pour le codage de l'indice des prédicteurs sélectionnés, notamment à bas débit. Cette réduction permet d'obtenir une amélioration de l'efficacité de codage de la référence. De plus, l'utilisation de la méthode pour la suppression des résolutions sous-pixelliques du codage Inter présente des gains préliminaires intéressants pour les hautes définitions.

Afin de réduire l'information de mouvement, nous avons adopté une démarche différente de celle étudiée pour la compétition des prédicteurs de vecteurs. Cette démarche s'inscrit dans celle du transfert d'intelligence au profit du décodeur. Le but n'est donc pas de réduire l'entropie des résiduels en exploitant des corrélations des champs de vecteurs, mais de supprimer les résiduels en exploitant les corrélations temporelles entre les images. Pour cela, une méthode d'estimation de mouvement conjointe à l'encodeur et au décodeur a été mise en place. Cette estimation est basée sur les données causales du bloc courant. De plus, un partitionnement 1D du macrobloc a aussi été utilisé pour réduire la distance spatiale entre la partition à prédire et les données causales. La méthode permet de supprimer près de la moitié des résiduels. Par conséquent, des gains sont observés pour la méthode et ils dépassent même l'efficacité de la méthode de compétition des prédicteurs de vecteurs pour la configuration testée. De plus, l'utilisation conjointe de ces deux méthodes accroît encore les gains, prouvant ainsi leur complémentarité.

Enfin, nous avons mesuré l'utilisation conjointe des méthodes qui remplissent l'objectif de la thèse de proposer des améliorations de la norme H.264/AVC suffisamment efficaces et normalisables pour

être intégrées dans un prochain standard. Les gains obtenus dans une configuration de type diffusion sont remarquables, notamment à bas débit. Les méthodes sont donc compatibles et complémentaires, produisant ainsi des résultats supérieurs aux résultats individuels de chaque méthode.

Les deux grands axes développés dans cette thèse, qui étaient la mise en place de modules de compétition et la suppression de l'information de compétition, nous ont conduit à plusieurs conclusions. Tout d'abord l'introduction de nouvelles possibilités de codage mises en compétition avec les possibilités existantes doit engendrer un gain supérieur, en terme de débit-distorsion, au débit lié à cette nouvelle compétition. La deuxième conclusion, qui découle de la première, est que la compétition n'est pas toujours utile et notamment dans la référence. Le meilleur exemple est sans doute la compensation de mouvement sous-pixelique qui est un outil redoutable pour les basses résolutions et qui devient contraignant pour les hautes résolutions. Il est clair qu'une amélioration de la référence réside dans une adaptation ou une sélection des compétitions utiles pour le signal à coder et les qualités visées. L'augmentation de l'intelligence du décodeur, pour la sélection automatique des compétitions utiles, semble être une voie de recherche intéressante et efficace.

## Perspectives

A l'issue des travaux étudiés dans ce manuscrit, des perspectives intéressantes apparaissent. Les trois principales sont le développement complet de la méthode Intra 1D définie théoriquement, l'extension de la méthode de suppression de prédicteurs de textures et la définition de l'estimation de mouvement au décodeur pour différentes granularités du mouvement. Cependant, des études plus succinctes, comme l'adaptation des ensembles de prédicteurs pour les vecteurs mouvement de tous les modes Inter sont aussi envisageables.

Dans l'explication théorique du partitionnement 1D des macroblocs, nous avons défini plusieurs partitions initiales  $Par_1$ . Le test de l'ensemble de ces partitions initiales devrait augmenter l'efficacité de la méthode. Le nombre de prédicteurs pour chaque partitionnement et ordre de parcours a été fixé empiriquement dans une configuration Baseline. Il serait judicieux de refaire cette étude afin de fixer un nombre optimal de prédicteurs lorsque l'Intra  $8 \times 8$  est activé. De même, un partitionnement 1D sur des blocs de taille supérieure à celle d'un macrobloc permettant d'augmenter le nombre de pixels par partition devrait rendre plus compétitif le mode Intra 1D face à l'Intra  $8 \times 8$ . Les ordres de parcours hiérarchique et bidirectionnel engendrent une entropie différente des partitions. Une adaptation de la quantification en fonction de la place de la partition dans l'ordre de parcours devrait augmenter les gains. Finalement, l'extension de la méthode aux composantes de chrominance est indispensable pour une éventuelle contribution en normalisation. Ces perspectives sont actuellement étudiées par notre équipe.

L'estimation de mouvement conjointe à l'encodeur et au décodeur s'est révélée efficace. La contrepartie est, bien entendu, une augmentation de la complexité du décodeur. Cependant, la mise en place d'algorithmes de recherches rapides d'estimation de mouvement permettrait de réduire efficacement cette complexité et rendrait cette méthode réaliste pour une intégration dans un standard. Cette méthode pourrait s'appliquer à différentes granularités du mouvement. On pourrait ainsi l'utiliser pour déterminer la granularité du mouvement au pixel entier ou au sous-pixel. De même, pour

un codage scalable, cette méthode pourrait être utilisée pour affiner le mouvement d'une couche d'amélioration de qualité ou de résolution. Enfin, on pourrait utiliser cette méthode pour donner une composante de rotation au mouvement, sans transmettre de nouvelle information. L'efficacité de la méthode augmente lorsque l'on utilise l'Inter 1D ligne. Le développement des autres partitionnements 1D définis dans le chapitre Intra 1D devrait encore accroître l'efficacité de codage.

L'application du critère d'élimination des prédicteurs du codage Intra a montré sa capacité à réduire le nombre de bits alloués au codage de l'indice du meilleur prédicteur. Les résultats préliminaires obtenus avec ce critère pour l'élimination des résolutions sous-pixelles sont déjà élevés. Le prolongement de cette étude offre donc une perspective intéressante. Ce critère pourrait aussi être utilisé pour l'élimination des prédicteurs de vecteurs mouvement. Dans une telle méthode, les données fournies à l'algorithme d'élimination seraient alors les blocs pointés par les prédicteurs. Ceci permettrait, entre autres, d'augmenter le nombre de prédicteurs de vecteurs dans l'ensemble, avec un débit des indices qui pourrait être réduit grâce au test d'égalité des prédicteurs, mais aussi par élimination des indices de prédicteurs implicites et par cette méthode. Dans l'état de l'art, nous avons vu que le principal axe de recherche pour l'amélioration du codage Intra prédictif était l'augmentation du nombre de prédicteurs. La méthode permettrait donc d'augmenter l'efficacité de ces nouvelles compétitions en sélectionnant un nombre adapté de prédicteurs. Enfin, un codage Intra des résiduels de bloc provenant de la compensation de mouvement, fournissant des gains préliminaires significatifs, a été récemment présenté au groupe VCEG. L'utilisation de l'algorithme devrait accroître les performances de cette technique.



## Annexe A

# Configuration de Test

*Afin de garantir l'efficacité d'une nouvelle technique de codage par rapport à la norme H.264/AVC, le groupe VCEG a défini des conditions de test. Historiquement, ces conditions de test sont celles utilisées pour le choix des outils intégrés dans la norme [Bjø01b]. Suite à l'objectif d'expérimentation de nouvelles techniques de codages, ces conditions ont été mises à jour en 2005 [TSW05]. Depuis, elles sont régulièrement discutées lors des réunions du groupe et actualisées. L'utilisation de ces conditions est impérative pour une éventuelle adoption par le groupe VCEG d'une nouvelle technique de codage. Dans ce manuscrit, nous avons généré nos résultats en fonction des recommandations de janvier 2008 [TSW08]. Ces conditions ont été légèrement modifiées en juillet et en octobre. Les nouvelles conditions intègrent, notamment, de nouvelles séquences et l'utilisation d'outils non normatifs. Dans cette annexe, nous décrirons rapidement ces conditions de test pour les différents profils, avant de récapituler les séquences utilisées dans nos tests.*

### A.1 Conditions VCEG

Les conditions de test VCEG concernent des options normatives ou non normatives qui permettent d'obtenir un codage optimal pour chaque profil ou configuration. L'ensemble des options normatives obligatoires pour chaque profil est implicite.

#### A.1.1 Baseline

Pour le profil Baseline, deux configurations sont recommandées. La configuration Baseline III qui contient uniquement des slices Intra et la configuration Baseline IPP composée d'une seule image Intra suivie d'images P. Les options préconisées pour ces deux configurations (dont certaines sont communes au profil High) sont :

- Profil IDC 66 et Level 40.
- Le codage entropique CAVLC.
- Un slice par image.

- “L’Adaptive rounding offset”. Cet outil non normatif ajuste “l’offset” de la quantification en fonction des blocs de la séquence déjà encodés.
- La sélection “RDopt” des modes de codage. Avec cette option, pour chaque mode Intra et Inter, le débit réel est calculé au bit près.
- Fenêtre de recherche  $32 \times 32$  pour l’estimation de mouvement pour les séquences QCIF et CIF et  $64 \times 64$  pour les séquences de plus hautes résolutions.
- Quatre images de référence pour les images P.

En plus des deux configurations définies par le groupe VCEG, nous avons utilisé une configuration (Baseline GOP 12) composée d’une image I pour 11 images P. Chaque séquence a le même nombre d’images que celui préconisé dans les conditions de test. De plus, dans certains de nos tests, nous avons codé uniquement la composante de luminance au lieu des trois composantes recommandées.

### A.1.2 High

Pour le profil High, quatre configurations sont recommandées. Les configurations High III et High IPP ont les mêmes nombre et type d’images que les configurations Baseline III et Baseline IPP. La configuration High IBBP est composée d’une seule image I, d’images P et de deux images B entre deux images P. Enfin, la configuration High BH est composée, elle aussi d’une image I et d’images P et de sept images B hiérarchiques entre deux images P. Pour ces configurations, l’option de sélection “RDopt” est aussi utilisée. De même, les fenêtres de recherche  $32 \times 32$  ou  $64 \times 64$  et les quatre images de référence sont également préconisées pour l’estimation de mouvement. En plus de ces options, d’autres options spécifiques à ces configurations sont recommandées, dont :

- Profil IDC 100 et Level 40.
- Le codage entropique CABAC.
- Transformée DCT  $8 \times 8$  en plus de la transformée DCT  $4 \times 4$ .
- Sélection “RD Picture Decision” pour chaque slice. Cette option permet la sélection optimale du type d’image et du QP à utiliser pour chaque image.
- Deux images de référence pour les slices B dans la direction passée et une pour la direction future pour la configuration IBBP et deux pour la configuration BH.
- Utilisation des prédictions pondérées pour les images B.
- Ordre pyramidale pour la configuration BH.

En plus de ces quatre configurations, nous avons défini une configuration High GOP 12 composée d’une image I pour trois images P et 8 images B.

### A.1.3 Liste des séquences

Pour les séquences recommandées par le groupe VCEG, le nombre d’images et la fréquence d’encodage sont spécifiées. Dans nos tests, nous avons utilisé un ensemble élargi de séquences. Dans le tableau suivant, nous indiquons pour chacune des séquences utilisées la fréquence, le nombre d’images, l’appartenance à l’ensemble des séquences VCEG et leurs caractéristiques.

Nom	Séquence	Fréquence	Images	VCEG	Caractéristiques
<b>QCIF</b>					
Container		15Hz	150	Oui	Caméra fixe - Faible mouvement
Foreman		15Hz	150	Oui	Caméra mobile - Mouvement fort
Silent		15Hz	150	Oui	Caméra fixe - Mouvement localisé
<b>CIF</b>					
Foreman		30Hz	300	Oui	Même séquence que Forman QCIF 15Hz avec une résolution et une fréquence plus élevées
Mobile		30Hz	300	Oui	Caméra fixe - Mouvements : translation, rotation, zoom etc. - Couleurs vives
Paris		15Hz	150	Oui	Caméra fixe - Faible mouvement - Fond texturé - Séquence type "visio"
Tempete		30Hz	260	Oui	Caméra fixe - Mouvement élevé - Zoom arrière - Tourbillons de feuilles
Carphone		30Hz	150	Non	Caméra fixe - Mouvement élevé - Fond en mouvement
Crew		30Hz	100	Non	Caméra mobile - Mouvement élevé - Flashes

Ice		30Hz	100	Non	Caméra fixe - Mouvement élevé - Fond uniforme
Schumacher		30Hz	100	Non	Caméra mobile - Mouvement fort - Textes
<b>4CIF</b>					
City		30Hz	150	Non	Caméra mobile - Mouvement moyen - Fond texturé - Vue aérienne
Ice		30Hz	150	Non	Même séquence que Ice CIF 30Hz avec une résolution plus élevée
Soccer		60Hz	150	Non	Caméra mobile - Fort Mouve- ment - Fond texturé
<b>720p</b>					
Bigships		60Hz	150	Oui	Caméra fixe - Faible mouvement - Zoom arrière
Shuttlestart		60Hz	150	Oui	Caméra fixe - Mouvement fort - Bruit - Changement de lumino- sité
City_corr		60Hz	150	Oui	Même séquence que City 4CIF 30Hz avec une résolution et une fréquence plus élevées
Crew		60Hz	150	Oui	Même séquence que Crew CIF 30Hz avec une résolution et une fréquence plus élevées

Night		60Hz	150	Oui	Caméra fixe - Mouvement fort - Plusieurs objets en mouvement - Fond texturé
Raven		60Hz	150	Non	Caméra mobile - Mouvement fort - Fond flouté
Stockholm		60Hz	150	Non	Caméra mobile - Mouvement moyen - Bande noire à gauche - Vue aérienne
Cyclists		60Hz	150	Non	Caméra mobile - Mouvement moyen
Jets		60Hz	150	Non	Caméra fixe - Mouvement moyen - Zoom arrière - Bruit - Pluie
Panslow		60Hz	150	Non	Caméra mobile - Faible mouve- ment - Bruit - Objets texturés
Sailormen		60Hz	150	Non	Caméra fixe - Mouvement fort - Bruit - Fond texturé
Spincalendar		60Hz	150	Non	Caméra fixe - Mouvement fort - Rotation - Bruit - Fond texturé

1080p					
Blue sky		25Hz	100	Non	Caméra mobile - Mouvement fort - Rotation
Intotree		50Hz	100	Non	Caméra mobile - Mouvement moyen - Bruit élevé - Fond tex- turé - Vue aérienne
Station2		25Hz	100	Non	Caméra fixe - Mouvement fort - Zoom arrière - Bruit - Fond tex- turé
Toys and calendar		60Hz	100	Non	Caméra mobile - Mouvement moyen - Objets en mouvement

## Annexe B

# Sélection des prédicteurs de l'Intra 1D

*Dans cette annexe, nous donnons les résultats de l'étude qui nous a permis de fixer le nombre et les prédicteurs de chaque sous-modes Intra 1D. C'est-à-dire, pour chaque ordre de parcours de chaque partitionnement testé. Pour chacune des configurations, nous avons réordonné les prédicteurs. Par exemple, pour le parcours lexicographique du mode Ligne, les 17 prédicteurs ont été ordonnés en fonction de leur nombre de sélections jusqu'à ce que les prédicteurs soient ordonnés dans l'ensemble en fonction de ce nombre de sélection. En effet, la position d'un prédicteur dans l'ensemble peut modifier son nombre de sélections. Les 9 prédicteurs les plus sélectionnés ont ensuite été testés et réordonnés suivant le même processus. Nous avons testé 17, 9, 5, 3, 2, 1 prédicteur(s) pour chacun des sous-ensembles. Ces nombres de prédicteurs correspondent au nombre optimal pour le codage des indices de prédicteurs Intra.*

*Pour chacun des tests reportés ci-dessous, pour chaque première image de chaque séquence, nous avons utilisé le profil Baseline. Chacun des sous-modes Intra 1D est donc en compétition avec l'Intra 16×16 et l'Intra 4×4 de la norme H.264/AVC. Les résultats présentés ci-dessous concernent les séquences CIF Foreman, Mobile, Paris, Tempete, Akiyo, Carphone, Crew, Modo, Schumacher. Les résultats sont donnés en pourcentage de réduction de débit calculé avec la métrique de Bjøntegaard [Bjø01a] pour l'ensemble des QP 27, 32, 37, 42.*

## B.1 Partitionnement Ligne

### B.1.1 Ordre lexicographique

Le tableau B.1 donne le pourcentage de réduction de débit obtenu avec l'ordre de parcours lexicographique du mode Ligne en fonction du nombre de prédicteurs. Nous pouvons remarquer que le gain est majoré par l'utilisation de 9 prédicteurs avec un gain de 3,5%. Les résultats obtenus pour 2, 3, 5 et 17 prédicteurs sont cependant très proches. Seule la configuration avec un seul prédicteur donne un résultat réellement plus faible avec 1,5%. Ceci s'explique par les caractéristiques des prédicteurs :

		<b>Ligne Lexicographique</b>					
Nombre de prédicteurs		1	2	3	5	9	17
Moyenne		1,5	2,9	3,1	3,2	<b>3,5</b>	3,0

TABLE B.1 – Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF pour 1, 2, 3, 5, 9, 17 prédicteur(s) dans l'ensemble des prédicteurs du mode Intra 1D Ligne lexicographique.

		<b>Ligne Bidirectionnel</b>					
Nombre de prédicteurs		$Par^P$					
		1	2	3	5	9	17
$Par^B$	1	0,9	<b>3,5</b>	3,3	2,9	3,1	3,0
	2	1,4	2,6	<b>2,7</b>	2,4	2,6	2,4
	3	2,3	3,4	3,3	3,2	<b>3,5</b>	3,2
	5	2,6	3,3	<b>3,3</b>	3,2	3,2	3,1
	9	2,5	3,1	<b>3,2</b>	3,0	3,0	2,9

TABLE B.2 – Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF en fonction du nombre de prédicteurs de chacun des ensembles utilisés pour le codage des partitions  $Par^P$  et  $Par^B$  du mode Intra 1D Ligne bidirectionnel.

le premier est la ligne précédente et le deuxième est la copie du pixel de gauche. Par conséquent, l'utilisation d'un seul prédicteur exploite uniquement les redondances spatiales provenant du haut de la ligne alors que l'utilisation d'au moins deux prédicteurs permet d'exploiter les redondances à la fois du haut de la ligne et celles à gauche de la ligne.

### B.1.2 Ordre bidirectionnel

Le tableau B.2 donne le pourcentage de réduction de débit en fonction des combinaisons du nombre de prédicteurs de chacun des deux ensembles de prédicteurs utilisés pour le codage des partitions  $Par^P$  et  $Par^B$  de l'ordre de parcours bidirectionnel. La meilleure configuration est obtenue avec deux prédicteurs pour l'ensemble de prédicteurs dédié aux partitions  $Par^P$  et un prédicteur pour l'ensemble dédié aux partitions  $Par^B$ . Les pourcentages de gain pour l'ensemble des combinaisons sont cependant équivalents à ce meilleur résultat. Remarquons enfin que ces résultats sont proches de ceux obtenus pour l'ordre de parcours lexicographique de ce mode Ligne.

### B.1.3 Ordre hiérarchique

Le tableau B.3 donne le pourcentage de gain obtenu avec l'ordre de parcours hiérarchique du partitionnement Ligne en fonction du nombre de prédicteurs dans l'ensemble des prédicteurs. La meilleure configuration est obtenue pour cinq prédicteurs. Les gains sont légèrement inférieurs à ceux obtenus avec l'ordre lexicographique.

		Ligne Hiérarchique				
Nombre de prédicteurs		1	2	3	5	9
Moyenne		0,4	2,3	2,0	<b>3,0</b>	2,4

TABLE B.3 – Pourcentage de réduction de débit moyen de la configuration Baseline I pour l’ensemble des QP 27-42, pour neuf séquences CIF pour 1, 2, 3, 5, 9 prédicteur(s) dans l’ensemble des prédicteurs du mode Intra 1D Ligne hiérarchique.

		Ligne Lexicographique					
Nombre de prédicteurs		1 <sup>er</sup> ensemble					
		1	2	3	5	9	17
2 <sup>ème</sup> ensemble	1		3,4	3,6	3,8	<b>3,9</b>	3,7
	2			3,3	3,5	<b>4,1</b>	3,6
	3				3,5	<b>4,0</b>	3,7
	5					<b>3,9</b>	3,6
	9						3,4

TABLE B.4 – Pourcentage de réduction de débit moyen de la configuration Baseline I pour l’ensemble des QP 27-42, pour neuf séquences CIF en fonction du nombre de prédicteurs des deux ensembles du mode Intra 1D Ligne lexicographique.

### B.1.4 Deux ensembles, ordre lexicographique

Dans le cas d’une méthode de codage par compétition, il est parfois intéressant d’utiliser un nombre différent de prédicteurs pour un même mode de codage. Comme nous avons pu le constater tout au long de ce manuscrit, un nombre peu élevé de prédicteurs est efficace à bas débit où le coût de la compétition entre les différents prédicteurs a une proportion élevée dans le débit total. Par contre, à haut débit, un nombre de prédicteurs élevé a peu d’impact sur le débit mais permet de réduire efficacement l’entropie de l’information codée. Nous avons donc utilisé deux ensembles de prédicteurs pour un même sous-mode Intra 1D. Le tableau B.4 donne le pourcentage de gain obtenu avec l’utilisation de deux ensembles de prédicteurs pour l’ordre de parcours lexicographique du mode Ligne, en fonction du nombre de prédicteurs de chacun des deux ensembles. Les résultats du tableau montrent que l’utilisation de deux ensembles de prédicteurs est efficace pour la totalité des configurations testées. De plus, les gains maximisent ceux obtenus avec un seul ensemble de prédicteurs. La meilleure réduction de débit est obtenue avec neuf prédicteurs pour le premier ensemble et deux pour le deuxième.

Nombre de prédicteurs	Colonne Lexicographique					
	1	2	3	5	9	17
Moyenne	0,4	1,1	1,2	<b>1,6</b>	1,4	1,1

TABLE B.5 – Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF pour 1, 2, 3, 5, 9, 17 prédicteur(s) dans l'ensemble des prédicteurs du mode Intra 1D Colonne lexicographique.

## B.2 Partitionnement Colonne

### B.2.1 Ordre lexicographique

Le tableau B.5 donne le pourcentage de gain obtenu avec l'ordre de parcours lexicographique du mode Ligne, en fonction du nombre de prédicteurs. En utilisant cinq prédicteurs, le gain est majoré de 1,6%. Ces gains sont beaucoup moins élevés que ceux du même ordre de parcours pour le mode Ligne. La seule explication cohérente est tout simplement que les corrélations spatiales dans les séquences utilisées sont plus élevées horizontalement que verticalement. Pour ce mode, on peut faire la même remarque sur le gain faible obtenu par l'utilisation d'un seul prédicteur. En effet, dans ce cas, seules les corrélations spatiales provenant des pixels de gauche sont exploitées alors que l'utilisation d'au moins deux prédicteurs permet d'exploiter en plus les corrélations spatiales provenant des pixels en haut de la colonne courante.

### B.2.2 Ordre bidirectionnel

Le tableau B.6 donne le pourcentage de réduction de débit en fonction des combinaisons du nombre de prédicteurs de chacun des deux ensembles de prédicteurs utilisés pour le codage des partitions  $Par^P$  et  $Par^B$  de l'ordre de parcours bidirectionnel. La meilleure configuration est obtenue avec neuf prédicteurs pour l'ensemble de prédicteurs dédié aux partitions  $Par^P$  et avec deux prédicteurs pour l'ensemble dédié aux partitions  $Par^B$ . Les gains sont proches de ceux obtenus pour l'ordre de parcours lexicographique de ce mode Colonne.

### B.2.3 Ordre hiérarchique

Le tableau B.3 donne le pourcentage de gain obtenu avec l'ordre de parcours hiérarchique du partitionnement Colonne en fonction du nombre de prédicteurs dans l'ensemble des prédicteurs. La meilleure configuration est obtenue avec deux prédicteurs. Les résultats ne dépassent pas 1%. Les gains sont inférieurs à ceux obtenus avec l'ordre lexicographique.

### B.2.4 Deux ensembles, ordre lexicographique

Le tableau B.8 donne le pourcentage de gain obtenu avec l'utilisation de deux ensembles de prédicteurs pour l'ordre de parcours lexicographique du mode Colonne, en fonction du nombre de prédicteurs de chacun des deux ensembles. La configuration contenant neuf prédicteurs pour le

		Colonne Bidirectionnel					
		$Par^P$					
Nombre de prédicteurs		1	2	3	5	9	17
$Par^B$	1	0,1	1,6	1,6	1,2	1,2	<b>1,7</b>
	2	0,3	0,8	0,9	0,5	<b>2,0</b>	1,1
	3	0,7	1,1	1,1	0,9	0,8	<b>1,4</b>
	5	0,9	1,0	1,4	1,0	0,6	<b>1,5</b>
	9	0,9	0,9	1,0	0,6	0,8	<b>1,4</b>

TABLE B.6 – Pourcentage de réduction de débit moyen de la configuration Baseline I pour l’ensemble des QP 27-42, pour neuf séquences CIF en fonction du nombre de prédicteurs de chacun des ensembles utilisés pour le codage des partitions  $Par^P$  et  $Par^B$  du mode Intra 1D Colonne bidirectionnel.

		Colonne Hiérarchique				
Nombre de prédicteurs		1	2	3	5	9
Moyenne		0,1	<b>0,8</b>	0,7	0,6	0,7

TABLE B.7 – Pourcentage de réduction de débit moyen de la configuration Baseline I pour l’ensemble des QP 27-42, pour neuf séquences CIF pour 1, 2, 3, 5, 9 prédicteur(s) dans l’ensemble des prédicteurs du mode Intra 1D Colonne hiérarchique.

premier ensemble et 1 prédicteur pour le second donne, en moyenne, le meilleur résultat. Comme pour le partitionnement Ligne, les gains obtenus sont supérieurs à ceux obtenus avec l’utilisation d’un seul ensemble de prédicteurs.

		Colonne Lexicographique					
		1 <sup>ier</sup> ensemble					
Nombre de prédicteurs		1	2	3	5	9	17
2 <sup>ieme</sup> ensemble	1		1,4	1,9	1,9	<b>2,1</b>	1,8
	2			1,6	2,0	<b>2,0</b>	1,8
	3				1,9	<b>2,0</b>	2,0
	5					2,0	<b>2,1</b>
	9						1,8

TABLE B.8 – Pourcentage de réduction de débit moyen de la configuration Baseline I pour l'ensemble des QP 27-42, pour neuf séquences CIF en fonction du nombre de prédicteurs des deux ensembles du mode Intra 1D Colonne lexicographique.

## Annexe C

# Autres méthodes et perspectives étudiées

*Dans cette annexe, nous proposons de décrire les méthodes que nous avons développées dont les résultats préliminaires se sont révélés peu intéressants pour une étude plus approfondie. Nous décrivons donc une méthode permettant de générer de nouvelles images de référence pour l'estimation et la compensation de mouvement, et une méthode d'estimation de mouvement récursive conjointe à l'encodeur et au décodeur. De plus, nous reportons des perspectives qui se sont révélées infructueuses pour la méthode de codage par compétition des vecteurs mouvement et pour la méthode d'estimation de mouvement conjointe à l'encodeur et au décodeur.*

### C.1 Demi images de référence

Afin d'améliorer l'efficacité de la compensation de mouvement, nous avons effectué, durant cette thèse, une étude préliminaire nommée "demi- images de référence". Dans notre analyse de la sélection des images de référence, nous avons constaté que la majorité des blocs codés en Inter ont un prédicteur qui provient de l'image de référence ayant la distance temporelle la plus faible par rapport à l'image courante. De plus, la sélection des images de référence plus éloignées temporellement augmente avec le nombre d'images par seconde. Par conséquent, une liste d'images de référence composée d'images temporellement plus "proches" de l'image courante devrait augmenter l'efficacité de codage. Dans cette méthode, nous proposons de fusionner des images précédemment encodées/décodées dans le but de créer de nouvelles images de référence qui seront insérées dans la liste des images de référence utilisées pour l'estimation et la compensation de mouvement de l'image courante. Les images ainsi créées ont un instant temporel plus "proche" de l'image courante que les images de référence précédemment décodées, comme on peut le constater dans la figure C.1. L'objectif de la méthode est donc de diminuer l'entropie des résiduels de texture et d'améliorer la qualité des prédicteurs de blocs Inter.

Dans cette section, nous décrivons les tests que nous avons effectués pour valider nos hypothèses,

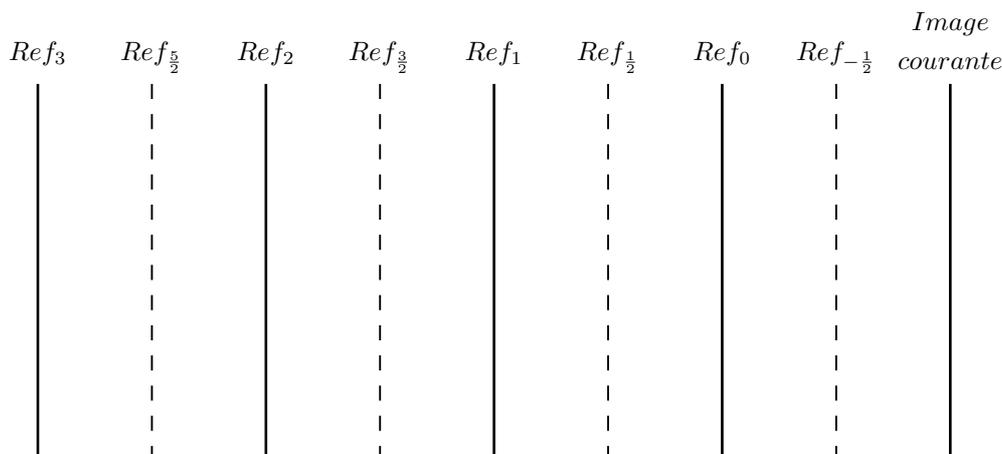


FIGURE C.1 – Représentation des demi-images de référence et des images de référence classiques.

avant de présenter les résultats obtenus.

### C.1.1 Implémentation

Pour évaluer le potentiel de la méthode et valider nos hypothèses, nous avons implémenté trois versions. Pour chacune de ces versions, l'image de référence  $Ref_1$  est remplacée par une nouvelle image de référence. Dans la première version, l'image  $Ref_1$  est remplacée par l'image  $Ref_{\frac{1}{2}}^{V1}$  représentée dans la figure C.1. Cette image est la moyenne, pixel à pixel, des deux images  $Ref_0$  et  $Ref_1$ .

Dans la seconde implémentation, le mouvement de l'image  $Ref_0$  vers  $Ref_1$  est pris en compte pour la construction de l'image  $Ref_{\frac{1}{2}}$ . C'est-à-dire que chaque pixel de l'image  $Ref_{\frac{1}{2}}^{V2}$  est égal à la moyenne entre le pixel de l'image  $Ref_0$  de même position et le pixel de l'image  $Ref_1$  déplacé par le vecteur mouvement associé au pixel de l'image  $Ref_0$ . Si le pixel de l'image  $Ref_0$  est codé en Intra, le pixel de l'image  $Ref_{\frac{1}{2}}^{V2}$  est égal au pixel de l'image  $Ref_{\frac{1}{2}}^{V1}$ . Remarquons que dans cette implémentation chaque pixel de l'image  $Ref_{\frac{1}{2}}^{V2}$  est la moyenne de deux pixels.

Enfin, dans la troisième implémentation, l'image  $Ref_1$  est remplacée par l'image  $Ref_{-1}^{Estimated}$  qui est une projection de l'image  $Ref_0$  à l'instant temporel de l'image courante. Cette image est construite par extrapolation des pixels de l'image  $Ref_0$ , en fonction des vecteurs mouvement inversés de l'image  $Ref_0$  vers l'image  $Ref_1$ . Dans cette implémentation, les zones de recouvrement et de découvrement ne sont pas prises en compte. Lorsqu'aucune valeur n'est obtenue pour le pixel de l'image  $Ref_{-1}^{Estimated}$ , ce pixel est égal au pixel de  $Ref_0$  de même position. Ceci génère nécessairement des artefacts.

### C.1.2 Résultats

Les tests ont été effectués avec le profil Baseline de la norme H.264/AVC pour deux images de référence. Les séquences utilisées sont quatre séquences CIF 30Hz : Akyo, Irene, Marc et Silent. 10

images sont codées pour chacune de ces séquences. Ces séquences ont été sélectionnées spécialement pour leur fond fixe, afin de limiter l'impact des zones de recouvrement et dé-couvrement qui n'est pas géré dans la troisième implémentation.

L'utilisation de l'image  $Ref_{\frac{1}{2}}^{V1}$  de la première version de l'implémentation fournit seulement 0.1% de gain moyen avec un maximum de 0.8%. Pour la seconde implémentation, le gain moyen est nul, mais le maximum atteint 0.9% de réduction de débit. Enfin, l'utilisation de l'image  $Ref_{-1}^{Estimated}$  ne fournit malheureusement aucun gain positif. Toutefois, pour toutes ces implémentations, on constate un nombre de sélections plus élevé des nouvelles images par rapport à la sélection de l'image  $Ref_1$ .

Les interpolations créées ( $Ref_{\frac{1}{2}}^{V1}$ ,  $Ref_{\frac{1}{2}}^{V2}$ ,  $Ref_{-1}^{Estimated}$ ), génèrent des défauts dans les zones de mouvement, comme on peut le constater dans la figure C.2, dans laquelle sont représentées ces trois types d'interpolations pour la séquence Irene. L'image  $Ref_{\frac{1}{2}}^{V1}$  (cf. figure C.2(b)) contient plus de flou que l'image  $Ref_1$  (cf. figure C.2(a)), notamment dans la zone de la main en mouvement. Pour l'image  $Ref_{\frac{1}{2}}^{V2}$  (cf. figure C.2(c)), le flou est encore présent dans la zone de la main en mouvement, même si celui-ci est moins important que pour l'image  $Ref_{\frac{1}{2}}^{V1}$ . Enfin, l'image  $Ref_{-1}^{Estimated}$  (cf. figure C.2(d)) ne contient pas de flou lié à l'extrapolation de mouvement. Cependant, on constate des artefacts importants autour de la zone de la main en mouvement, liés à la non prise en compte des zones de recouvrement et de dé-couvrement. Les artefacts de ces images permettent d'expliquer les mauvais résultats obtenus en moyenne, et les pertes constatées pour les séquences contenant du mouvement.

### C.1.3 Conclusion

Les résultats préliminaires obtenus dans cette étude ne nous ont pas paru suffisamment conséquents pour envisager une étude approfondie. Toutefois, la sélection plus élevée des nouvelles images de référence par rapport à l'image de référence  $Ref_1$  nous a permis de valider notre hypothèse. Pour obtenir des gains significatifs, il aurait fallu à la fois proposer des interpolations temporelles complexes et considérer un encodage plus efficace des indices d'images de référence et des vecteurs mouvement.

## C.2 Estimation de mouvement récursive conjointe à l'encodeur et au décodeur

Dans cette section, nous reportons les résultats d'une étude sur une estimation de mouvement récursive effectuée avec Marco Cagnazzo, Maître de Conférences à l'Ecole Nationale Supérieure des Télécommunications de Paris. Dans cette étude, nous avons mis en place une estimation de mouvement basée sur l'algorithme de Cafforio-Rocca [CR83]. Cette estimation de mouvement récursive utilise l'information des résiduels de texture à chaque itération pour affiner le vecteur mouvement de chaque pixel. Le but est d'utiliser cet algorithme conjointement à l'encodeur et au décodeur pour améliorer la compensation de mouvement. En effet, cette méthode fournit un vecteur mouvement par pixel qui peut être de résolution sous-pixellique inférieure au  $\frac{1}{4}$  de pixel utilisé dans la norme H.264. Nous décrirons brièvement cet algorithme dans la section C.2.1. Nous avons implémenté deux algorithmes d'amélioration de la compensation de mouvement. Dans le premier, décrit dans la

(a) Image  $Ref_1$ (b) Image  $Ref_{\frac{1}{2}}^{V1}$ (c) Image  $Ref_{\frac{1}{2}}^{V2}$ (d) Image  $Ref_{-1}^{Estimated}$ 

FIGURE C.2 – Image  $Ref_1$  et les interpolations  $Ref_{\frac{1}{2}}^{V1}$ ,  $Ref_{\frac{1}{2}}^{V2}$  et  $Ref_{-1}^{Estimated}$  obtenues dans la méthode des demi images de référence pour la séquence Irene CIF 30Hz.

section C.2.2, l'estimation de mouvement récursive est utilisée pour fournir un vecteur mouvement par pixel afin d'améliorer le prédicteur du bloc courant. Dans le deuxième algorithme, décrit dans la section C.2.3, un second résiduel de texture est transmis au décodeur. Ce résiduel contient la différence entre le prédicteur de texture de l'estimation de mouvement classique et celle obtenue par l'estimation de mouvement récursive.

### C.2.1 Estimation de mouvement récursive de type Cafforio-Rocca

L'algorithme de Cafforio-Rocca [CR83] est un algorithme d'estimation de mouvement récursif. Cet algorithme fournit un vecteur mouvement par pixel. Si l'on considère le vecteur mouvement  $\overrightarrow{mv}_{n,m}^{i-1}$  le vecteur mouvement à l'itération  $i-1$  du pixel de position  $(n, m)$  dans l'image courante, le vecteur mouvement de l'itération  $i$ ,  $\overrightarrow{mv}_{n,m}^i$ , est donné par la formule suivante :

$$\overrightarrow{mv}_{n,m}^i = \overrightarrow{mv}_{n,m}^{i-1} - \frac{e_{n,m}}{\lambda + |\nabla_i|^2} \overrightarrow{\nabla}_i \quad (\text{C.1})$$

où  $e_{n,m}$  est la différence entre le pixel de position  $(n, m)$  de l'image courante et le pixel pointé par le vecteur mouvement  $\overrightarrow{mv}_{n,m}^{i-1}$  de l'image de référence. Pour la première itération ( $i=0$ ), on considère que le vecteur  $\overrightarrow{mv}_{n,m}^0 = (0, 0)$ .  $\nabla_i$  est le gradient de luminance du pixel de l'image de référence pointée par le vecteur mouvement  $\overrightarrow{mv}_{n,m}^{i-1}$ .  $\lambda$  est un paramètre dit d'"amorti". Ce paramètre, qui dépend des implémentations, permet de fixer le compromis entre le nombre d'itérations et la précision de l'algorithme.

### C.2.2 Amélioration du prédicteur de bloc

Dans notre première utilisation de l'estimation de mouvement récursive de Cafforio-Rocca, le but est d'améliorer le bloc prédicteur. Dans cette implémentation, pour le bloc courant, le schéma classique de codage est appliqué. C'est-à-dire que pour ce bloc courant, l'estimation de mouvement et la compensation de mouvement classique sont effectuées, puis l'ensemble des autres opérations (transformée, quantification, transformation inverse, quantification inverse, etc..) sont appliquées. A la fin de ces opérations, l'encodeur a un bloc prédicteur  $PredBlock_0$  et un bloc reconstruit  $DecodedBlock_0$ . De plus, on considère que chaque pixel de position  $(n, m)$  a un vecteur mouvement  $\overrightarrow{mv}_{n,m}^0$  égal au vecteur utilisé pour la compensation de mouvement de ce bloc.

Afin d'améliorer le bloc reconstruit, nous appliquons ensuite une itération de l'algorithme de Cafforio-Rocca. Pour chaque pixel de position  $(n, m)$ , nous déterminons un vecteur mouvement  $\overrightarrow{mv}_{n,m}^1$  à l'aide de l'équation (C.1). Dans ce cas, pour le pixel de position  $(n, m)$ ,  $e_{n,m}$  est la différence entre le pixel de même position dans le bloc reconstruit  $DecodedBlock_0$  et le pixel  $(n, m)$  dans le bloc prédicteur  $PredBlock_0$ . Le gradient  $\overrightarrow{\nabla}_i$  est calculé sur le bloc reconstruit  $DecodedBlock_0$ . A la fin de l'estimation de mouvement de type Cafforio-Rocca, nous avons un vecteur mouvement  $\overrightarrow{mv}_{n,m}^1$  pour chaque pixel du bloc courant. Ces vecteurs peuvent être de résolution sous-pixellique inférieure au  $\frac{1}{4}$  de pixel utilisé dans la norme H.264/AVC. Ces vecteurs permettent de construire un nouveau prédicteur de bloc  $PredBlock_1$ . Ce nouveau prédicteur est utilisé pour calculer le nouveau bloc reconstruit  $DecodedBlock_1$  qui est donné par la formule suivante :

$$DecodedBlock_1 = DecodedBlock_0 - PredBlock_0 + PredBlock_1 \quad (C.2)$$

Le bloc reconstruit  $DecodedBlock_1$  ne fournit pas nécessairement une amélioration de qualité du bloc  $DecodedBlock_0$ . Nous avons donc mis en place une compétition entre ces deux blocs reconstruits. Pour cette compétition, nous utilisons le critère de débit distorsion. Cette compétition est effectuée au niveau macrobloc. Par conséquent, un indice indiquant l'utilisation ou non de la méthode est inséré dans le flux. Notons que cet indice est la seule information supplémentaire transmise au décodeur par rapport à la référence, la norme H.264/AVC.

Nous avons implémenté cette méthode dans le JM KTA. Le gain (en réduction de débit) apporté par la méthode ne compense pas le débit nécessaire à la transmission de l'indice pour différencier les deux macroblocs reconstruits ( $DecodedBlock_0$  et  $DecodedBlock_1$ ). De plus, une large majorité des erreurs  $e_{n,m}$  est égale à zéro avec l'estimation de mouvement complète, la quantification des résiduels et l'utilisation du mode Skip. Par conséquent, peu de  $DecodedBlock_1$  sont différents de  $DecodedBlock_0$ . Notons que l'indice n'est pas transmis lorsque le résiduel d'un macrobloc est nul. La méthode est donc applicable uniquement sur un ensemble réduit de macroblocs (ceux qui ont un résiduel de bloc). Par conséquent, la méthode est très peu sélectionnée à bas débit. Pour les macroblocs ayant un résiduel non nul, la méthode est tout de même sélectionnée dans près d'un tiers des cas. Cependant, aucun gain significatif (0.1%) et aucune perte significative n'ont été observés.

Enfin, remarquons que nous avons testé différents paramètres  $\lambda$  afin de trouver une valeur optimale. Pour améliorer la méthode, nous avons aussi essayé différentes initialisations du vecteur  $\overrightarrow{mv_{n,m}^0}$  en prenant, par exemple, les vecteurs déjà calculés  $\overrightarrow{mv_{n,m}^1}$  des pixels voisins.

### C.2.3 Second résiduel de texture

Dans la deuxième implémentation, nous utilisons une seconde itération de l'algorithme. La même estimation présentée dans la section précédente est effectuée. Cependant, le bloc reconstruit  $DecodedBlock_1$  est différent puisqu'il est l'addition du bloc prédicteur  $PredBlock_1$  et d'un second résiduel  $ResidualBlock_1$  qui est transmis au décodeur. Ce second résiduel de texture est obtenu par l'équation suivante :

$$ResidualBlock_1 = DCT^{-1} (Q^{-1}Q (DCT(OriginalBlock - PredBlock_1))) \quad (C.3)$$

où  $OriginalBlock$  est le bloc original. Dans cette version, le premier résiduel transmis est uniquement utilisé pour la seconde estimation de mouvement. Cette implémentation engendre des pertes en termes de débit distorsion. Elle est donc moins efficace que la précédente. En effet, la transmission de ce second résiduel augmente le débit des blocs pour lesquels ce second résiduel est transmis.

### C.2.4 Conclusion

Notre implémentation d'une seconde estimation de mouvement, basée sur l'algorithme d'estimation de mouvement de Cafforio-Rocca, ne présente pas de gains suffisamment intéressants. Ces résultats sont sans aucun doute liés à une forte proportion de résiduels de texture nuls qui ne permet pas une utilisation de la méthode sur une proportion élevée de blocs des images Inter. La méthode

pourrait se révéler plus efficace dans d'autres contextes, comme un codeur vidéo qui n'utiliserait pas d'estimation de mouvement sous-pixellique. Ceci engendrerait une plus grande quantité de résiduels non nuls et, par conséquent, une plus forte sélection de la méthode proposée.

### C.3 Adaptation d'ensembles de prédicteurs de vecteurs mouvement

Dans le chapitre 5, nous avons proposé une adaptation d'ensembles de prédicteurs pour le mode Skip basée sur nos observations. Nous avons aussi tenté, durant cette thèse, de définir des adaptations d'ensembles de prédicteurs pour les autres modes Inter. Dans cette section, nous présentons les tests que nous avons mis en oeuvre pour déterminer d'éventuelles corrélations entre les ensembles de prédicteurs. Puis, nous donnerons les principales conclusions de cette étude.

Dans notre recherche du meilleur ensemble de prédicteurs pour le mode Skip et pour les modes Inter (cf. chapitre 4), nous n'avons pas pu mettre en évidence, pour une ou plusieurs séquences, un ensemble de prédicteurs offrant des gains significativement supérieurs à ceux obtenus pour le couple  $\{mv_{H.264}, mv_{Scol_{R_0}}\}$ . Cependant, certains autres couples spatio-temporel composés du médian ( $mv_{H.264}$ ) et du prédicteur  $mv_{trans}$  ou  $mv_{trans_{sum_j}}$  ou  $mv_{col_{sum_j}}$  obtiennent des gains légèrement supérieurs au couple  $\{mv_{H.264}, mv_{Scol_{R_0}}\}$  pour certaines séquences. La sélection automatique séquence par séquence du meilleur ensemble ne dépasse pas une amélioration de 0,2% de réduction de débit. Par conséquent, contrairement au mode Skip, nous n'avons pas pu définir de critère évident pour une sélection automatique du meilleur ensemble de prédicteurs pour les autres modes Inter.

Toutefois, nous avons essayé de déterminer des corrélations entre les différentes données disponibles au décodeur, permettant une sélection automatique du meilleur prédicteur, vecteur par vecteur. Ces tests ont été effectués sur plusieurs séquences pour plusieurs QP. Parmi les données évaluées, nous avons testé les valeurs et la variance des vecteurs voisins et du vecteur collocated. Nous avons aussi considéré les indices d'image de référence utilisés pour le codage des blocs voisins, ainsi que la taille des partitions. Les conclusions de ces tests sont que le prédicteur collocated est plus sélectionné pour les vecteurs dont les vecteurs spatio-temporels voisins ont une forte variance. Nous avons aussi remarqué que la sélection de ce prédicteur est plus élevée pour les grandes tailles de partitions, comme nous l'avons expliqué dans le chapitre 4. Cependant, aucune de ces corrélations ne se révèle efficace pour une sélection automatique entre le prédicteur médian et le prédicteur collocated. Finalement, la compétition débit distorsion entre ces deux prédicteurs est la plus efficace, en moyenne, pour l'ensemble des modes Inter (non Skip).

### C.4 Perspectives de l'estimation de mouvement conjointe à l'encodeur et au décodeur

Dans le chapitre 9, nous avons présenté une méthode d'estimation de mouvement conjointe à l'encodeur et au décodeur dans le but de supprimer une partie de l'information de mouvement.

Les réductions de débit obtenues avec cette méthode dépassent celles de la méthode de codage par compétition de prédiction de vecteurs mouvement (cf. chapitre 4). Les perspectives offertes par la méthode sont nombreuses. Nous avons implémenté une partie de ces perspectives que nous décrivons dans cette section.

La première est l'utilisation du vecteur mouvement généré par l'estimation de mouvement EMD comme prédicteur du vecteur mouvement obtenu par l'estimation de mouvement classique. Dans ce cas, l'estimation EMD n'est pas utilisée pour supprimer une partie des résiduels de vecteurs mouvement comme pour la méthode originale proposée dans le chapitre 9. Une augmentation de débit moyenne de 0.5% est observée, pour l'ensemble des séquences testées (cf. chapitre 9), lorsque ce prédicteur est utilisé à la place du prédicteur médian. Lorsqu'il est en compétition avec le prédicteur médian ou le prédicteur collocated, les gains ne sont pas supérieurs à ceux obtenus par la meilleure configuration de MVComp (cf. chapitre 4). Afin de tenter d'expliquer ces résultats, nous avons comparé la valeur de ce prédicteur avec le prédicteur médian et le prédicteur collocated. Le vecteur obtenu par l'estimation EMD est proche du prédicteur médian. Le prédicteur EMD a la même valeur que le prédicteur médian dans la quasi totalité des cas, lorsque les blocs voisins n'ont pas de résiduel et que ces vecteurs sont égaux. Ces cas sont les plus fréquents, comme nous l'avons constaté dans la section C.2. De plus, comme le vecteur obtenu par l'estimation EMD n'utilise pas le bloc original, il ne peut pas être plus précis que le vecteur médian.

Dans le chapitre 9, nous avons mis en évidence que l'estimation EMD est plus efficace pour la partie entière de l'estimation de mouvement que pour l'estimation sous-pixellique. Nous avons implémenté une alternative à la méthode proposée dans le chapitre 9, en utilisant l'estimation EMD pour l'estimation de mouvement entière et l'estimation de mouvement classique pour l'estimation sous-pixellique. Dans cette estimation de mouvement hybride, seule la partie sous-pixellique du résiduel de vecteur mouvement est transmise au décodeur. Cette nouvelle estimation de mouvement n'est pas en compétition avec l'estimation de mouvement classique. Cette méthode génère des pertes par rapport à la référence, qui sont supérieures à celle de l'estimation EMD originale seule. Nous avons aussi évalué l'utilisation de l'estimation de mouvement classique pour la partie entière de l'estimation de mouvement et l'estimation EMD pour la partie sous-pixellique. Dans ce cas, le résiduel transmis au décodeur correspond uniquement à la partie entière. Cette implémentation obtient aussi des pertes par rapport à la référence mais elles sont inférieures à celles de l'implémentation précédente. Toutefois, ces pertes sont supérieures à l'estimation EMD originale.

Finalement, nous avons implémenté la méthode pour l'estimation sous-pixellique au  $\frac{1}{8}$  de pixel qui est intégrée dans le JM KTA 1.3. Lorsque l'estimation EMD est en compétition avec l'estimation de mouvement classique, les gains obtenus par rapport à une référence utilisant une compensation de mouvement au  $\frac{1}{8}$  de pixel, sont inférieurs à ceux obtenus (cf. chapitre 9) par la méthode par rapport à la norme H.264.

# Bibliographie

- [AAAB05] M.A. Agostini, T. Andre, M. Antonini, et M. Barlaud. Scalable motion coding for video coders with lifted MCWT. In *Proc. International Workshop on Very Low Bit-rate Video-coding*, Costa Rei, Sardaigne - Italie, Septembre 2005.
- [AAB06] M.A. Agostini, M. Antonini, et M. Barlaud. Model-based bit allocation between wavelet subbands and motion information in MCWT video coders. In *Proc. European Signal Processing Conference (EUSIPCO)*, Florence, Italie, Septembre 2006.
- [ABMD92] M. Antonini, M. Barlaud, P. Mathieu, et I. Daubechies. Image coding using wavelet transforms. *IEEE Trans. on Image Processing*, 12 :205–220, Avril 1992.
- [AF95] S. Aign et K. Fazel. Temporal and spatial error concealment techniques for hierarchical MPEG-2 video codec. volume 3, pages 1778–1783, Juin 1995.
- [ANR74] N. Ahmed, T. Natarajan, et K.R. Rao. Discrete Cosine Transform. *IEEE Trans. on Computer*, 23(1) :90–94, Janvier 1974.
- [ARCEP] Arcep. <http://www.arcep.fr/>.
- [ARK<sup>+</sup>04] A. Al, B.P. Rao, S.S. Kudva, S. Babu, D. Sumam, et A.V. Rao. Quality and complexity comparison of H.264 Intra mode with JPEG2000 and JPEG. In *Image Processing, ICIP, IEEE International Conference on*, Octobre 2004.
- [BB00] W. Butera et Jr.V.M. Bove. The coding ecology : image coding via competition among experts. *IEEE Trans. on Circuits and System for Video Technology*, 10(7) :1049–1058, Octobre 2000.
- [BDB05] S. Boltz, E. Debreuve, et M. Barlaud. A joint motion computation and segmentation algorithm for video coding. In *Proc. European Signal Processing Conference (EUSIPCO)*, Antalya, Turquie, Septembre 2005.
- [Ben48] W. Bennett. Spectra of quantized signals. *Bell System Technical Journal*, 27 :446–472, Juillet 1948.
- [Ber71] T. Berger. *Rate Distortion Theory : A Mathematical Basis for Data Compression*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
- [Bie88] M. Bierling. Displacement estimation by hierarchical block matching. In *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, volume 1001, Cambridge, Massachusetts, USA., Novembre 1988.

- [Bjø01a] G. Bjøntegaard. Calculation of average PSNR differences between RD-curves. Proposition VCEG-M33, ITU-T VCEG, Texas, USA, Avril 2001.
- [Bjø01b] G. Bjøntegaard. Recommended Simulation Conditions for H.26L. Input / Discussion VCEG-M75, ITU-T VCEG, Santa Barbara, USA, Septembre 2001.
- [Bru90] H. Brusewitz. Motion compensation with triangles. In *Proc. 3rd Int Conf. on 64kbit Coding of Moving Video*, Rotterdam, Pays-Bas, Septembre 1990.
- [Bye03] J. Byeungwoo. Fast mode decision for H.264. Proposition JVT-J033, JVT, Waikoloa, Hawaii, USA, Décembre 2003.
- [CAA<sup>+</sup>09] M. Cagnazzo, M.A. Agostini, M. Antonini, G. Laroche, et J. Jung. Motion vector quantization for efficient low bit-rate video coding. In *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, San jose, Californie, USA, Janvier 2009.
- [Cam04] N. Cammas. *Codage vidéo scalable par maillages et ondelettes t+2D*. Thèse de Doctorat, Université de Rennes 1, France, Novembre 2004.
- [CC00] T. Chinen et A. Chien. Visual evaluation of JPEG2000 color image compression performance. Document JVT-N1583, JVT, Tokyo, Japon, Mars 2000.
- [Che89] C.T. Chen. Adaptive transform coding via quadtree-based variable blocksize DCT. *Acoustics, Speech, and Signal Processing, ICASSP, IEEE International Conference on*, 3 :1854–1857, Mai 1989.
- [CN07] T. Chujoh et R. Noda. Internal bit depth increase for coding efficiency. Proposition VCEG-AE13, ITU-T VCEG, Marrakech, Maroc, Janvier 2007.
- [CPN04] M. Chaumont, S. Pateux, et H. Nicolas. Object-based video coding using a dynamic coding approach. In *Image Processing, ICIP, IEEE International Conference on*, volume 2, pages 1105–1108, Singapour, 2004.
- [CR83] C. Cafforio et F. Rocca. The Differential Method for Image Motion Estimation. *Springer-Verlag Image sequence processing and dynamic scene analysis*, 1983.
- [CW97] C.M. Chen et A.N. Willson. A spatial and temporal motion vector coding algorithm for low-bit-rate video coding. In *Image Processing, ICIP, IEEE International Conference on*, volume 2, pages 791–794, Santa Barbara, Californie, USA, Octobre 1997.
- [CYC90] M.H. Chan, B.Y. Yu, et A.G. Constantinides. Variable size block matching motion compensation with applications to video coding. *Proc. Inst. Elec. Eng.*, 137 :205–212, Août 1990.
- [DEY<sup>+</sup>07] C. Dai, O.D. Escoda, P. Yin, X. Li, et C. Gomila. Geometry-Adaptive Block Partitioning for Intra Prediction in Image & Video Coding. In *Image Processing, ICIP, IEEE International Conference on*, volume 6, San Antonio, Texas, USA, Octobre 2007.
- [DH98] A. Deever et S. Hemami. Dense motion field reduction for motion estimation. In *Signals, Syst. and Comput.*, volume 2, pages 944–948, Novembre 1998.

- [DK92] F. Dufaux et M. Kunt. Multigrid block matching motion estimation with an adaptive local mesh refinement. In *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, volume 1818, pages 97–109, 1992.
- [DKBR99] S. Deuk Kim et J. Beom Ra. An efficient motion vector coding scheme based on minimum bitrate prediction. *IEEE Trans. on Image Processing*, 8(8) :1117–1120, 1999.
- [DNG01] F.G.B. De Natale et F. Granelli. Low-complexity context-based motion compensation for VLBR video encoding. In *Multimedia & Expo, ICME, IEEE International Conference on*, pages 439–442, Tokyo, Japon, Août 2001.
- [DSCW00] A.L. Da Silva Cruz et J.W. Woods. Adaptive motion vector quantization for video coding. In *Image Processing, ICIP, IEEE International Conference on*, volume 2, pages 867–870, Vancouver, Canada, Octobre 2000.
- [EB08] H. El Bahi. Codage vidéo H.264. Rapport de Stage, UFR Mathématiques - Informatique, Université Paris Descartes, Juin 2008.
- [Ebr94] T. Ebrahimi. A new technique for motion field segmentation and coding for very low bitrate video coding applications. In *Image Processing, ICIP, IEEE International Conference on*, volume 2, pages 433–437, Austin, Texas, USA, Novembre 1994.
- [Eno07] D. Enoe. Compression vidéo l’après H.264. Rapport de Stage, ENSICAEN, Octobre 2007.
- [Eve63] H. Everett. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operation Research*, 11 :399–417, 1963.
- [EWK95] O. Egger, L. Wei, et M. Kunt. High Compression Image Coding Using an Adaptive Morphological Subband Decomposition. 83(2) :272–287, Février 1995.
- [FPF09] Festival Pocket Films. <http://www.festivalpocketfilms.fr/>.
- [MFF09] Mobile Film Festival. <http://fr.mobilefilmfestival.com/>.
- [FG03] M. Flierl et B. Girod. Generalized B pictures and the draft H.264/AVC video-compression standard. *IEEE Trans. on Circuits and System for Video Technology*, 13(7) :587–597, Juillet 2003.
- [Fle99] P. Fleury. *Dynamic Scheme Selection In Image Coding*. Thèse de Doctorat, Ecole Polytechnique Fédérale de Lausanne, Suisse, 1999.
- [FMW] L. Fan, S. Ma, et F. Wu. Overview of avs video standard. In *Multimedia & Expo, ICME, IEEE International Conference on*, volume 1, pages 423–426, Taipei, Taiwan.
- [FPR95] E. Feig, H. Peterson, et V. Ratnakar. Image compression using spatial prediction. In *Acoustics, Speech, and Signal Processing, ICASSP, IEEE International Conference on*, volume 4, pages 2339–2342, Detroit, Michigan, USA, 1995.
- [GARRM05] B. Girod, A. Aaron, S. Rane, et D. Rebollo-Monedero. Distributed video coding. *Proceedings of the IEEE, Special Issue on Video Coding and Delivery Papier Invité*, 93(1) :71–83, Janvier 2005.

- [Gir94] B. Girod. Rate-constrained Motion Estimation. In *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, volume 2, page 2308, Chicago, Illinois, USA, Septembre 1994.
- [GLH04] Z. Gang, G. Li, et Y. He. The Intra prediction based on sub block. In *Signal Processing, ICSP International Conference on*, volume 1, pages 467–469, Istanbul, Turquie, Septembre 2004.
- [GMW04] V. George, D. Marpe, et T. Wiegand. Performance Comparison of Intra-only H.264/AVC High Profile and JPEG2000 for a Set of Monochrome. Proposition JVT-M014, ISO/IEC Test Images, Joint Video Team, Palma de Majorque, Espagne, Octobre 2004.
- [H.264/AVC] Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC). ITU-T and ISO/IEC JTC 1, Version 1 : Mai 2003, Version 2 : Mai 2004, Version 3 : Mars 2005, Version 4 : Septembre. 2005, Version 5 and Version 6 : Juin 2006, Version 7 : Avril. 2007.
- [H.264/SVC] Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC). ITU-T and ISO/IEC JTC 1, Version 8 (SVC extension) : Juillet 2007.
- [H.263] Video coding for low bit rate communication. ITU-T Recommendation H.263, (Version 2), 1998.
- [H.264/MVC] Call for Proposals on Multi-view Video Coding. ISO/IEC JTC1/SC29/WG11 N7327, 2005.
- [JM] H.264/AVC software coordination. <http://iphome.hhi.de/suehring/tml/download/>.
- [HC97] H.M. Hang et J.J. Chen. Source model for transform video coder and its application. I. Fundamental theory. *IEEE Trans. on Circuits and System for Video Technology*, 7(2) :287–298, Avril 1997.
- [HK01] A. Hallapuro et M. Karczewicz. Low complexity transform and quantization Part 1 : Basic Implementation. Document JVT-B038, JVT, Février 2001.
- [HNEY<sup>+</sup>07] S. Har-Noy, O.D. Escoda, P. Yin, C. Gomila, et T.Q. Nguyen. Adaptive in-loop prediction refinement for video coding. In *Multimedia Signal Processing, MMSP, IEEE Workshop on*, pages 171–174, Crète, Grèce, Octobre 2007.
- [Hor08] M. Horowitz. A coding efficiency-computational complexity analysis of KTA 1.8 coding tools. Information VCEG-C409, ITU-T VCEG, Genève, Suisse, Avril 2008.
- [HP01] S.C. Han et C.I. Podilchuk. Video compression with dense motion fields. *IEEE Trans. on Image Processing*, 10(11) :1605–1612, Novembre 2001.
- [HS81] B.K.P. Horn et B.G. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17(1-3) :185–203, Août 1981.
- [IY09] T. Ikai et T. Yamamoto. Very large block transform for Intra coding. Proposition COM 16 C 89 E, ITU-T VCEG, Genève, Suisse, Janvier 2009.

- [Jac92] A.E. Jacquin. Image coding based on a fractal theory of iterated contractive image transformations. *Image Processing, IEEE Transactions on*, 1(1) :18–30, Janvier 1992.
- [Jasper] Jasper. <http://www.ece.uvic.ca/mdadams/jasper/>.
- [JFB95] R.L. Joshi, T.R. Fischer, et R.H. Bamberger. Lossy encoding of motion vectors using entropy-constrained vector quantization. 3 :109–112, Octobre 1995.
- [JJS93] N. Jayant, J. Johnston, et R. Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10) :1385–1422, Octobre 1993.
- [JL06a] J. Jung et G. Laroche. Competition-Based Motion Vector Prediction - new results. Information VCEG-C79, ITU-T VCEG, Genève, Suisse, Novembre 2006.
- [JL06b] J. Jung et G. Laroche. Competition-Based Scheme for Motion Vector Selection and Coding. Information VCEG-AC06, ITU-T VCEG, Klagenfurt, Autriche, Juillet 2006.
- [JL07] J. Jung et G. Laroche. Performance evaluation of the KTA 1.2 software. Information VCEG-AE09, ITU-T VCEG, Marrakech, Maroc, Janvier 2007.
- [JLPP07] J. Jung, G. Laroche, et B. Pesquet-Popescu. RD optimized competition scheme for efficient motion prediction. In *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, San Jose, Californie, USA, Février 2007. Papier invité.
- [JP07] J. Jung et S. Pateux. An excel add-in for computing Bjøntegaard metric and its evolution. Information VCEG-AE07, ITU-T VCEG, Marrakech, Maroc, Janvier 2007.
- [JPEG] JPEG standard. ISO/IEC 10918-1 ITU-T Recommendation T.81, 1993.
- [JPEG2000] JPEG 2000 Part I : Final Draft International Standard (ISO/IEC FDIS15444-1). ISO/IEC JTC1/SC29/WG1 N1855, 2000.
- [JZG<sup>+</sup>04] X. Ji, D. Zhao, W. Gao, Q. Huang, S. Ma, et Y. Lu. New bi-prediction techniques for B pictures coding. In *Multimedia & Expo, ICME, IEEE International Conference on*, volume 1, pages 101–104, Taipei, Taiwan, 2004.
- [KAKADU] Kakadu. <http://www.kakadusoftware.com/>.
- [Kar07] M. Karczewicz. Improvement Intra Coding. Proposition VCEG-AF15, ITU-T VCEG, San Jose, Californie, USA, Avril 2007.
- [KCL92] C.W. Kok, S.C. Chan, et S.H. Leung. Image coding using DCT of wavelet coefficients. In *Time-Frequency and Time-Scale Analysis, 1992., Proceedings of the IEEE-SP International Symposium*, pages 475–478, Victoria, Colombie Britannique, Canada, 1992.
- [KD92] J. Konrad et E. Dubois. Bayesian Estimation of Motion Vector Fields. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(9) :910–927, Septembre 1992.
- [Kri97] R. Krishnamurthy. *Compactly-Encoded Optical Flow Fields for Motion-Compensated Video Compression and Processing*. Thèse de Doctorat, Rensselaer Polytechnic Institute, Troy, New York, USA, 1997.
- [KSK04] S. Kondo, H. Sasai, et S. Kadono. Tree structured hybrid Intra prediction. In *Image Processing, ICIP, IEEE International Conference on*, volume 1, pages 473–476, Singapour, Octobre 2004.

- [KTA] KTA software coordination. <http://iphome.hhi.de/suehring/tml/download/KTA/>.
- [KTCS06] T. Keng Tan, S.B. Choong, et Y. Suzuki. Intra Prediction by Template Matching. In *Image Processing, ICIP, IEEE International Conference on*, pages 1693–1696, Atlanta, Géorgie, USA, Octobre 2006.
- [KYC08a] M. Karczewicz, Y. Ye, et I. Chong. Rate Distortion Optimized Quantization. Proposition VCEG-AH21, ITU-T VCEG, Antalya, Turquie, Janvier 2008.
- [KYC08b] M. Karczewicz, Y. Ye, et I. Chong. RDO quantization chroma improvement and mode-dependent directional transform. Proposition VCEG-AI36, ITU-T VCEG, Antalya, Turquie, Juillet 2008.
- [LAP<sup>+</sup>08] K. Lee, E. Alshina, J. Park, W. Han, et J. Min. Analysis for Ad Hoc Group on New Challenges in Video Coding Standardization. Proposition MPEG2008/M15579, ISO/IEC JTC1/SC29/WG11, Hanovre, Allemagne, Juillet 2008.
- [LH97] A.Y. Lan et J.N. Hwang. Scene context dependent reference frame placement for MPEG video coding. In *Acoustics, Speech, and Signal Processing, ICASSP, IEEE International Conference on*, volume 4, pages 2997–3000, Munich, Allemagne, 1997.
- [LHS06] Y.L. Lee, K.H. Han, et G.J. Sullivan. Improved lossless Intra coding for H.264/MPEG-4 AVC. volume 15, pages 2610–2615, Atlanta, Géorgie, USA, Septembre 2006.
- [LJL<sup>+</sup>03] P. List, A. Joch, J. Lainema, G. Bjøntegaard, et M. Karczewicz. Adaptive deblocking filter. 13(7) :614–619, Juillet 2003.
- [LJO98] L. Liang-Jin et A. Ortega. Bit-rate control using piecewise approximated rate-distortion characteristics. *IEEE Trans. on Circuits and System for Video Technology*, 8(4) :446–459, Août 1998.
- [LJPP08] G. Laroche, J. Jung, et B. Pesquet-Popescu. RD Optimized Coding for Motion Vector Predictor Selection. *IEEE Trans. on Circuits and System for Video Technology*, 18(9) :1247–1257, Septembre 2008.
- [LKD04] C.M. Lee, M. Kieffer, et P. Duhamel. Robust reconstruction of motion vectors using frame expansion. In *Acoustics, Speech, and Signal Processing, ICASSP, IEEE International Conference on*, volume 4, Mai 2004.
- [LKF96] L. Lurng-Kuo et E. Feig. A block-based gradient descent search algorithm for block motion estimation in video coding. *IEEE Trans. on Circuits and System for Video Technology*, 6(4) :419–422, Août 1996.
- [MC97] D. Marpe et H.L. Cycon. Efficient Pre-Coding Techniques for Wavelet-Based Image Compression. In *Picture Coding Symposium (PCS)*, pages 45–50, Berlin, Allemagne, 1997.
- [MGC03] D. Marpe, V. George, et H.L. Cycon. Performance evaluation of Motion-JPEG2000 in comparison with H.264/AVC operated in pure Intra coding mode. In *Wavelet Applications in Industrial Processing, Proceedings of the SPIE*, Rhode Island, USA, Octobre 2003.

- [MJPEG] Digital Compression and Coding of Continuous-Tone Still Images (version 3). ISO/IEC 10918 JPEG, (version 3), 1994.
- [MJPEG2000] Motion-JPEG2000, JPEG 2000 part 3. ISO/IEC 15444-3, JPEG2000 Part 3, 2002.
- [MK02] G. Melnikov et A.K. Katsaggelos. A jointly optimal fractal/DCT compression scheme. *Multimedia, IEEE Transactions on*, 4(4) :413–422, Décembre 2002.
- [MPEG2] Generic coding of moving pictures and associated audio. ITU-T Recommendation H.262 ISO/IEC 13818-2 MPEG-2 video, Part2 : Video, 1994.
- [MPEG4] Overview of the MPEG-4 Version 1 Standard. ISO/IEC JTC1/SC29/WG11 N2196, 1998.
- [MSW03] D. Marpe, H. Schwarz, et T. Wiegand. Context-based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard. *IEEE Trans. on Circuits and System for Video Technology*, 13(7) :620–636, Juillet 2003.
- [MT05] L. Merello et D. Taubman. Improvements to the Intra-coding Modes Offered by H.264. In *Multimedia Signal Processing, MMSP, IEEE Workshop on*, pages 1–4, Shanghai, Chine, Octobre 2005.
- [MTI00] S.A. Morley, K.S. Thyagarajan, et A.C. Irvine. Balancing technology in digital cinema systems. In *SMPTE Technical Conference*, Pasadena, Californie, USA, Octobre 2000.
- [MTKY07a] S. Matsuo, S. Takamura, K. Kamikura, et Y. Yashima. Extension of Intra Prediction using Reference Lines. Proposition VCEG-AF05, ITU-T VCEG, San Jose, Californie, USA, Avril 2007.
- [MTKY07b] S. Matsuo, S. Takamura, K. Kamikura, et Y. Yashima. Weighted Intra Prediction. Proposition VCEG-AG18, ITU-T VCEG, Shenzhen, Chine, Octobre 2007.
- [MWG05] D. Marpe, T. Wiegand, et S. Gordon. H.264/MPEG4-AVC fidelity range extensions : tools, profiles, performance, and application areas. In *Image Processing, ICIP, IEEE International Conference on*, volume 1, pages I–593–6, Gènes, Italie, Septembre 2005.
- [NH91] Y. Nakaya et H. Harashima. An iterative motion estimation method using triangular patches for motion compensation. In *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, volume 1605, pages 546–567, Boston, Massachusetts, USA, 1991.
- [NM06] M. Narroschke et H.G. Musmann. Adaptive prediction error coding in spatial and frequency domain with a fixed scan in the spatial domain. Proposition VCEG-AD07, ITU-T VCEG, Hangzhou, Chine, Octobre 2006.
- [NNMB<sup>+</sup>03] P. Ndjiki-Nya, B. Makai, G. Blattermann, A. Smolic, H. Schwarz, et T. Wiegand. Improved H.264/AVC coding using texture analysis and synthesis. In *Image Processing, ICIP, IEEE International Conference on*, volume 3, pages 849–852, Barcelone, Espagne, 2003.
- [ODE06] M. Ouaret, F. Dufaux, et T. Ebrahimi. On Comparing JPEG 2000 and Intraframe AVC. *SPIE Applications of Digital Image Processing XXIX*, 2006.

- [ODE07] M. Ouaret, F. Dufaux, et T. Ebrahimi. On Comparing Image and Video Compression Algorithms. In *Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM*, Scottsdale, Arizona, USA, 2007.
- [OR98] A. Ortega et K. Ramchandran. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine*, 15(6) :23–50, Novembre 1998.
- [PMCM01] S. Pateux, G. Marquant, et D. Chavira-Martinez. Object mosaicking via meshes and crack-lines technique. application to low bit-rate video coding. In *Picture Coding Symposium, PCS*, Séoul, Corée du Sud, Avril 2001.
- [PMR07] R. Puri, A. Majumdar, et K. Ramchandran. PRISM : A Video Coding Paradigm With Motion Estimation at the Decoder. *IEEE Trans. on Image Processing*, 16(10) :2436–2448, Octobre 2007.
- [PPJ07] J.Y. Park, S.W. Park, et B.M. Jeon. Intra prediction with subpel samples. Information VCEG-AG09, ITU-T VCEG, Shenzhen, Chine, Octobre 2007.
- [RCLB<sup>+</sup>96] E. Reusens, R. Castagno, C. Le Buhan, L. Piron, T. Ebrahimi, et M. Kunt. Dynamic video coding-an overview. In *Image Processing, ICIP, IEEE International Conference on*, volume 2, pages 377–380, Lausanne, Suisse, 1996.
- [REK97] E. Reusens, T. Ebrahimi, et M. Kunt. Dynamic coding of visual information. *IEEE Trans. on Circuits and System for Video Technology*, 7(3) :489–500, Juin 1997.
- [RELB<sup>+</sup>97] E. Reusens, T. Ebrahimi, C. Le Buhan, R. Castagno, V. Vaerman, L. Piron, C.D.S. Fabregas, S. Bhattacharjee, F. Bossen, et M. Kunt. Dynamic approach to visual data compression. *IEEE Trans. on Circuits and System for Video Technology*, 7(1) :197–211, Février 1997.
- [Rob08] A. Robert. *Transformées orientées par blocs pour le codage vidéo hybride*. Thèse de Doctorat, Télécom ParisTech, France, Janvier 2008.
- [RU07] D. Rusanovskyy et K. Ugur. Fast Selection of Adaptive Interpolation Filter. Contribution COM 16 - C 150 - E, ITU-T VCEG, Genève, Suisse, Juin 2007.
- [RUHL08] D. Rusanovskyy, K. Ugur, A. Hallapuro, et J. Lainema. Implementation of Directional Interpolation Filter with 16-bit Integer Arithmetic. Proposition VCEG-AH18, ITU-T VCEG, Antalya, Turquie, Janvier 2008.
- [RUL07] D. Rusanovskyy, K. Ugur, et J. Lainema. Adaptive Interpolation with Directional Filters. Proposition VCEG-AG21, ITU-T VCEG, Shenzhen, Chine, Octobre 2007.
- [SFK09] Y. Shishikui, Y. Fujita, et K. Kubota. Super Hi-Vision : the star of the show ! Technical report, EBU Technical Review, Janvier 2009.
- [SG88] Y. Shoham et A. Gersho. Efficient bit allocation for an arbitrary set of quantizers. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 36(9) :1445–1453, Septembre 1988.
- [Sha48] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 2,7 :379– 423, 1948.

- [SLS05] G.J. Sullivan, G. Lou, et M.T.J. Sun. Adaptive quantization yielding 10 dB gain. Information JVT-P015, JVT, Poznan, Pologne, Juillet 2005.
- [SMW05] H. Schwarz, D. Marpe, et T. Wiegand. Hierarchical B pictures. Information JVT-P014, JVT of ISO/IEC MPEG and ITU-T VCEG, Poznan, Pologne, Juillet 2005.
- [STC07] T. Shiodera, A. Tanizawa, et T. Chujoh. Block Based Extra/Inter-Polating Prediction for Intra Coding. In *Image Processing, ICIP, IEEE International Conference on*, volume 6, San Antonio, Texas, USA, Octobre 2007.
- [Sug08] M. Sugawara. Super Hi-Vision : research on a future ultra-HDTV system. Technical Report Q2, EBU Technical Review, 2008.
- [SV04] M. Smith et J. Villasenor. Intra-frame JPEG2000 vs. Inter-frame Compression Comparison. In *SMPTE Tech. Conf.*, Pasadena, Californie USA, Octobre 2004.
- [SW73] D. Slepian et J.K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. on Information Theory*, IT-19(4) :471–480, Juillet 1973.
- [SW98] G.J. Sullivan et T. Wiegand. Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine*, 15(6) :74–90, Novembre 1998.
- [TC06] A. Tanizawa et T. Chujoh. Adaptive Quantization Matrix Selection on KTA Software. Information VCEG-AD06, ITU-T VCEG, Hangzhou, Chine, Octobre 2006.
- [TCY08] A. Tanizawa, T. Chujoh, et T. Yamakage. Encoding complexity reduction of AQMS. Proposition VCEG-C.403, ITU-T VCEG, Genève, Suisse, Avril 2008.
- [TG96] K.H. Tan et M. Ghanbari. Block classification for an adaptive 1-D/2-D DCT video coding. *Vision, Image and Signal Processing, IEEE Proceedings*, 143(3) :155–163, Juin 1996.
- [Thi08] J.M. Thiesse. Méthode de compression d’images et vidéos par codage Intra 1D. Rapport de Stage, ISIMA, Clermont-Ferrand, France, Septembre 2008.
- [Tom07] M. Tomokazu. Adaptive Picture Flipping Coding on KTA Software. Information VCEG-AE17, ITU-T VCEG, Marrakech, Maroc, Janvier 2007.
- [Top05] P. Topiwala. Comparative study of JPEG2000 and H.264/AVC FRExt I-frame coding on high-definition video sequences. *SPIE Applications of Digital Image Processing XXVIII*, 5909(1) :284–292, Août 2005.
- [TP00] S. Tsekeridou et I. Pitas. MPEG-2 error concealment based on block-matching principles. *IEEE Trans. on Circuits and System for Video Technology*, 10(4) :646–658, Juin 2000.
- [TSW05] T.K. Tan, G.J. Sullivan, et T. Wedi. Recommended Simulation Common Conditions for Coding Efficiency Experiments. Input / Discussion VCEG-AA10, ITU-T VCEG, Nice, France, Octobre 2005.
- [TSW08] T.K. Tan, G.J. Sullivan, et T. Wedi. Recommended Simulation Common Conditions for Coding Efficiency Experiments Revision 2. Input / Discussion VCEG-AH10r3, ITU-T VCEG, Antalya, Turquie, Janvier 2008.

- [TWL05] A. Tourapis, F. Wu, et S. Li. Direct mode coding for bipredictive slices in the H.264 standard. *IEEE Trans. on Circuits and System for Video Technology*, 15(1) :119–126, Janvier 2005.
- [TYTA07] T. Tsukuba, T. Yamamoto, Y. Tokumo, et T. Aono. Adaptive Multidirectional Intra Prediction. Proposition VCEG-AG05, ITU-T VCEG, Shenzhen, Chine, Octobre 2007.
- [Use01] B.E. Usevitch. A tutorial on modern lossy wavelet image compression : Foundations of JPEG 2000. *IEEE Signal Processing Magazine*, 18 :22–35, Septembre 2001.
- [Val02] S. Valente. Rate/distortion-based combination of multiple DCT transforms for video coding. In *Image Processing, ICIP, IEEE International Conference on*, volume 3, pages 665–668, Rochester, New York, USA, 2002.
- [VC106] VC-1 Technical Overview. document, Microsoft Corporation, Octobre 2006.
- [VEW<sup>+</sup>05a] Y. Vatis, B. Edler, I. Wassermann, D.T. Nguyen, et J. Ostermann. Coding of Coefficients of two-dimensional Adaptive Wiener Interpolation Filter. In *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, Pékin, Chine, Juillet 2005.
- [VEW<sup>+</sup>05b] Y. Vatis, B. Edler, I. Wassermann, D.T. Nguyen, et J. Ostermann. Two-dimensional non-separable Adaptive Wiener Interpolation Filter for H.264/AVC. Proposition VCEG-Z17, ITU-T VCEG, Busan, Corée du Sud, Avril 2005.
- [VO05] Y. Vatis et J. Ostermann. Comparison of complexity between two-dimensional non-separable adaptive interpolation filter and standard Wiener filter. Proposition VCEG-AA11, ITU-T VCEG, Nice, France, Octobre 2005.
- [VO06] Y. Vatis et J. Ostermann. Prediction of P- and B-Frames Using a Two-dimensional Non-separable Adaptive Wiener Interpolation Filter for H.264/AVC. Proposition VCEG-AD08, ITU-T VCEG, Hangzhou, Chine, Octobre 2006.
- [VO07] Y. Vatis et J. Ostermann. Rate-distortion optimised coder control for adaptive interpolation filter in the KTA reference model. Proposition VCEG-AE16, ITU-T VCEG, Marrakech, Maroc, Janvier 2007.
- [Wed00] T. Wedi.  $\frac{1}{8}$ -pel motion vector resolution for H.26L. Proposition Q15-K-21, ITU-T VCEG, Portland, Oregon, USA, Août 2000.
- [Wed01] T. Wedi. Complexity Reduced Motion Compensated Prediction with  $\frac{1}{8}$ -pel Displacement Vector Resolution. Proposition VCEG-L20, ITU-T VCEG, Eibsee, Allemagne, Janvier 2001.
- [Wer96] O. Werner. Drift analysis and drift reduction for multiresolution hybrid video coding. *Signal Processing : Image Commun.*, 8(5), Juillet 1996.
- [WG91] S.W. Wu et A. Gersho. Rate-constrained optimal block-adaptive coding for digital tape recording of HDTV. *IEEE Trans. on Circuits and System for Video Technology*, 1(1) :100–112, Mars 1991.
- [Wie01] M. Wien. Intra coding using Variable Block Sizes. Proposition VCEG-O31, ITU-T VCEG, Pattaya, Thaïlande, Décembre 2001.

- [Wie02] T. Wiegand. Text of committee draft of joint video specification (ITU-T Rec. H.264-ISO/IEC 1449610 AVC). Technical Report JVT-C167, JVT of ISO/IEC MPEG and ITU-T VCEG, Fairfax, Virginie, USA, Mai 2002.
- [Wie03] M. Wien. Variable block-size transforms for H.264/AVC. *IEEE Trans. on Circuits and System for Video Technology*, 13(7) :604–613, Juillet 2003.
- [WK97] Z. Wu et T. Kanamaru. Block-based DCT and wavelet selective coding for arbitrarily shaped images. In *Visual Commun. Image Processing*, pages 658–665, San Jose, Californie, USA, Janvier 1997.
- [WLM<sup>+</sup>96] T. Wiegand, M. Lightstone, D. Mukherjee, T.G. Campbell, et S.K. Mitra. Rate-distortion optimized mode selection for very low bit rate video coding and the emerging H.263 standard. *IEEE Trans. on Circuits and System for Video Technology*, 6(2) :182–190, Avril 1996.
- [WLN07] Z. Wei, H. Li, et K.N. Ngan. An Efficient Intra Mode Selection Algorithm For H.264 Based On Fast Edge Classification. *Circuits and Systems, ISCAS, IEEE International Symposium on*, pages 3630–3633, Mai 2007.
- [Wol94] M. Wollborn. Prototype prediction for colour update in object-based analysis-synthesis. *IEEE Trans. on Circuits and System for Video Technology*, 4(3) :236–245, Juin 1994.
- [WT08] T. Wedi et T.K. Tan. AHG report - Coding Efficiency. report VCEG-AI03, ITU-T VCEG, Berlin, Allemagne, Juillet 2008.
- [WW07] S. Wittman et T. Wedi. Separable adaptive interpolation filter. Proposition C-219, ITU-T VCEG, Genève, Suisse, Juillet 2007.
- [WZ76] A.D. Wyner et J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. on Information Theory*, IT-22(1) :1–10, Janvier 1976.
- [WZG99] T. Wiegand, X. Zhang, et B. Girod. Long-term memory motion-compensated prediction. *IEEE Trans. Circuits and System for Video Technology*, 9(1) :70–84, Février 1999.
- [Yam07] T. Yamamoto. A new scheme for motion vector predictor encoding. Proposition VCEG-AF13, ITU-T VCEG, San Jose, Californie, USA, Avril 2007.
- [YC88] Y.B. Yu et A.G. Constantinides. Variable block size and position transform coding. In *Proc. European Signal Processing Conference (EUSIPCO)*, Grenoble, France, Septembre 1988.
- [YC02] S.L. Yu et C. Chrysafis. New Intra Prediction using Intra-Macroblock Motion Compensation. Proposition JVT-C151, JVT, Fairfax, Virginie, USA, Mai 2002.
- [YCC90] Y.B. Yu, M.H. Chan, et A.G. Constantinides. Low bit rate video coding using variable block size model. In *Acoustics, Speech, and Signal Processing, ICASSP, IEEE International Conference on*, volume 4, pages 2329–2332, Albuquerque, Nouveau Mexique, USA, 1990.

- [YK07] Y. Ye et M. Karczewicz. Improved Intra Coding. Proposition VCEG-AG11, ITU-T VCEG, Shenzhen, Chine, Octobre 2007.
- [YML96] Gongsan Yu, M.W. Marcellin, et M.M.K. Liu. Recovery of video in the presence of packet loss using interleaving and spatial redundancy. In *Image Processing, ICIP, IEEE International Conference on*, volume 2, pages 105–108, Septembre 1996.
- [YNK07] T. Yoshino, S. Naito, et A. Koike. An extended Intra prediction scheme based on frequency component representation. Proposition VCEG-AG06, ITU-T VCEG, Shenzhen, Chine, Octobre 2007.
- [YNY99] A. Yoneyama, Y. Nakajima, H. Yanagihara, et M. Sugano. MPEG encoding algorithm with scene adaptive dynamic GOP structure. In *Multimedia Signal Processing, IEEE 3rd Workshop on*, pages 297–302, Copenhagen, Danemark, 1999.
- [YPL04] C.L. Yang, L.M. Po, et W.H. Lam. A fast H.264 Intra prediction algorithm using macroblock properties. In *Image Processing, ICIP, IEEE International Conference on*, volume 1, pages 461–464, Singapour, Octobre 2004.
- [YR05] H. Yang et K. Rose. Rate-Distortion optimized motion estimation for error resilient video coding. In *Acoustics, Speech, and Signal Processing, ICASSP, IEEE International Conference on*, volume 2, pages 173–176, Mars 2005.
- [YVK95] J. Yeh, M. Vetterli, et M. Khansari. Motion compensation of motion vectors. In *Image Processing, ICIP, IEEE International Conference on*, volume 1, pages 574–577, Washington, District de Columbia, USA, Octobre 1995.
- [YYW95] L. Yoon Yung et J.W. Woods. Motion vector quantization for video coding. *Image Processing, IEEE Transactions on*, 4(3) :378–382, Mars 1995.
- [ZM05] H. Zhihai et S.K. Mitra. From rate-distortion analysis to resource-distortion analysis. *IEEE Circuits Syst. Magazine*, 5 :6–18, 2005.
- [ZZ91] Y.Q. Zhang et S. Zafar. Predictive block-matching motion estimation for TV coding. II. Inter-frame prediction. *Broadcasting, IEEE Transactions on*, 37(3) :102–105, Septembre 1991.
- [ZZB91] S. Zafar, Y.Q. Zhang, et J.S. Baras. Predictive block-matching motion estimation for TV coding. I. Inter-block prediction. *IEEE Trans. on Broadcasting*, 37(3) :97–101, Septembre 1991.
- [ZZM<sup>+</sup>04] P. Zhang, D. Zhao, S. Ma, Y. Lu, et W. Gao. Multiple modes Intra-prediction in Intra coding. In *Multimedia & Expo, ICME, IEEE International Conference on*, volume 1, Taipei, Taiwan, Juin 2004.

# Publications de l'auteur

## Articles de revue internationale

- [Revue1] G. Laroche, J. Jung, et B. Pesquet-Popescu. Intra Coding with Prediction Mode Information Inference. *IEEE Trans. on Circuits and System for Video Technology, Soumis*.
- [Revue2] G. Laroche, J. Jung, et B. Pesquet-Popescu. RD Optimized Coding for Motion Vector Predictor Selection. *IEEE Trans. on Circuits and System for Video Technology*, volume 18(9), pages 1247–1257, Septembre 2008.

## Articles de conférences internationales

- [Conf1] S. Corrado, M.A. Agostini, M. Cagnazzo, G. Laroche, et M. Jung, J. and Antonini. Improving H.264 performances by quantization of motion vectors. In *Picture Coding Symposium (PCS)*, Chicago, Illinois, USA, mai 2009.
- [Conf2] G. Laroche, J. Jung, et B. Pesquet-Popescu. Intra Prediction with 1D Macroblock Partitioning for Image and Video Coding. *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, San jose, Californie, USA, Janvier 2009.
- [Conf3] M. Cagnazzo, M.A. Agostini, M. Antonini, G. Laroche, et J. Jung. Motion vector quantization for efficient low bit-rate video coding. *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, San jose, Californie, USA, Janvier 2009.
- [Conf4] G. Laroche, J. Jung, et B. Pesquet-Popescu. Motion Vector Competition based on Macroblock Classification for the H.264 JM KTA software. *Proc. Advances Concepts for Intelligent Vision Systems (ACIVS)*, Delft, Hollande, Août 2007.
- [Conf5] J. Jung, G. Laroche, et B. Pesquet-Popescu. RD optimized competition scheme for efficient motion prediction. *Proc. SPIE Conf. on Visual Communications and Image Processing (VCIP)*, San Jose, Californie, USA, Janvier 2007. **Papier invité.**
- [Conf6] G. Laroche, J. Jung, et B. Pesquet-Popescu. A spatio-temporal competing scheme for the rate-distortion optimized selection and coding of motion vectors. *Proc. European Signal Processing Conference (EUSIPCO)*, Florence, Italie, Septembre 2006.

## Article de revue nationale

- [RevueFR1] G. Laroche, J. Jung, et B. Pesquet-Popescu. Suppression d'indices de prédicteurs implicites pour la compétition de prédictions de vecteurs mouvement. *Traitement du Signal, Soumis*.

## 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.
- [ConfFR2] G. Laroche, J. Jung, et B. Pesquet-Popescu. Adaptation Orientée Contenu pour le Codage par Compétition de Prédicteurs de Mouvement. *Proc. Compression et Représentation des Signaux Audiovisuels (CORESA)*, Montpellier, France, Novembre 2007.
- [ConfFR3] G. Laroche, J. Jung, et B. Pesquet-Popescu. Codage de vecteurs mouvement par compétition de prédicteurs spatio-temporels dans le standard H.264. *Proc. Compression et Représentation des Signaux Audiovisuels (CORESA)*, Caen, France, Novembre 2006.

## Contributions VCEG

- [Contrib1] J. Jung et G. Laroche. Performance evaluation of the KTA 1.2 software. Information VCEG-AE09, ITU-T VCEG, Marrakech, Maroc, Janvier 2007.
- [Contrib2] J. Jung et G. Laroche. Competition-Based Motion Vector Prediction - new results. Information VCEG-C79, ITU-T VCEG, Genève, Suisse, Novembre 2006.
- [Contrib3] J. Jung et G. Laroche. Competition-Based Scheme for Motion Vector Selection and Coding. Information VCEG-AC06, ITU-T VCEG, Klagenfurt, Autriche, Juillet 2006.

## Brevets

- [Brev1] J. Jung, G. Laroche et Jean-Marc Thiesse Calcul de prédicteurs de vecteurs mouvement. *Dépôt FR* : Juillet 2009. *Non divulgué à ce jour*.
- [Brev2] J. Jung, G. Laroche et Jean-Marc Thiesse Calcul de prédicteurs temporels de vecteurs mouvement. *Dépôt FR* : Juillet 2009. *Non divulgué à ce jour*.
- [Brev3] G. Laroche et J. Jung. Prédicteurs estimés. *Dépôt FR* : Décembre 2008. *Non divulgué à ce jour*.
- [Brev4] G. Laroche et J. Jung. Codage et décodage de prédicteurs prédéterminés. *Dépôt FR* : Octobre 2008. *Non divulgué à ce jour*.

- [Brev5] J. Jung et G. Laroche. Codage et décodage de partitions de pixels de forme linéaire. *Dépôt FR* : Avril 2008. *Non divulgué à ce jour.*
- [Brev6] J. Jung et G. Laroche. Prédicteurs de partitions Intra. *Dépôt FR* : Avril 2008. *Non divulgué à ce jour.*
- [Brev7] J. Jung et G. Laroche. Codage et décodage de mouvement. *Dépôt FR* : Novembre 2007. *Non divulgué à ce jour.*
- [Brev8] J. Jung et G. Laroche. Sélection de fonctions de décodage distribuées au décodeur. *Brevet n°* : FR 07 04710. Juin 2007. **Extension** : internationale.
- [Brev9] J. Jung, G. Laroche et M. Baillavoine. Codage et décodage par compétition prédictifs et adaptatifs. *Brevet n°* : FR 06 08658. Octobre 2006.
- [Brev10] J. Jung, G. Laroche, et M. Baillavoine. Codage et décodage adaptatifs. *Brevet n°* : FR 06 00273. Janvier 2006. **Extension** : Europe, USA, Japon, Chine, Corée, Inde.

## Exposés publics

- [Expo1] G. Laroche, J. Jung, et B. Pesquet-Popescu. Compétition de prédictions de vecteurs mouvement pour les futurs codeurs vidéo. *Présentation 15+15 MPIS*, Orange-Labs, Décembre 2007.
- [Expo2] G. Laroche, J. Jung, et B. Pesquet-Popescu. Méthode de codage par compétition de prédiction de vecteur mouvement dans le JM KTA. *GDR-ISIS Thème B - Image et Vision*, Paris, France, Mai 2007.
- [Expo3] G. Laroche, J. Jung, et B. Pesquet-Popescu. Etude d'un moteur de compétition pour le codage de séquences vidéo. *Journées doctorants du Pôle "Image et Multimédia"-FTR&D*, Rennes, France, Novembre 2006.
- [Expo4] G. Laroche. Introduction à la théorie de l'information. Cours-TD : *Licence, troisième année, Mathématiques, Informatique et Applications, Paris 5*. Paris, France, 2007.
- [Expo5] G. Laroche. Introduction au codage vidéo. Cours : *Master recherche, deuxième année, Mathématiques et Informatique spécialité Image, Paris 5*. Paris, France, 2005, 2006, 2007.



**Résumé** Cette thèse est dédiée à l'amélioration de la compression de séquences vidéo. Le but est de concevoir des méthodes suffisamment efficaces et réalistes pour être proposées aux organismes de normalisation de standard vidéo. Les approches choisies sont l'ajout de nouveaux modules de codage par compétition et la suppression de l'information de compétition. Les méthodes développées pour ces deux approches concernent à la fois le codage Intra et le codage Inter.

Dans un souci d'amélioration du codage de l'information de mouvement, un module de codage par compétition de prédicteurs de vecteurs mouvement a été intégré dans le KTA, permettant d'exploiter, de manière optimale au sens débit-distorsion, les redondances spatiales et temporelles des champs de vecteurs mouvement. Afin de prédéterminer l'ensemble optimal de prédicteurs, une méthode de sélection automatique d'ensembles de prédicteurs orientée contenu est aussi proposée. Enfin, un nouveau mode de codage Intra, basé sur un partitionnement 1D du macrobloc est présenté. Ce mode permet de réduire la distance spatiale entre le signal de référence et la partition courante, par rapport au codage par bloc de la norme. Ce mode de codage Intra est mis en compétition avec les modes bloc afin d'augmenter le nombre de compétitions disponibles pour l'exploitation des redondances spatiales entre les pixels.

Le dernier standard de compression vidéo, H.264/AVC, est basé sur un codage prédictif de l'information. Ce standard offre un nombre de compétitions beaucoup plus élevé que celui de ses prédécesseurs, à la fois pour le codage Intra et pour le codage Inter. Par conséquent, l'information de compétition est fortement représentée dans le débit total, notamment à bas débit. Afin de réduire ce débit lié à l'information de compétition, une partie de l'intelligence du codeur a été transférée au décodeur. Les indices des prédicteurs de vecteurs mouvement implicites, liés à la méthode de compétition de vecteurs mouvement, ont pu être ôtés du flux. De même, une partie des prédicteurs Intra a été supprimée en prenant en compte les paramètres du processus de quantification et le signal de référence utilisé dans la construction de ces prédicteurs. Enfin, en considérant que l'information de mouvement est une information de compétition, une estimation de mouvement au décodeur a ainsi été mise en place. Les résultats expérimentaux montrent que ces voies de recherches sont efficaces, particulièrement à bas débit.

L'ensemble des méthodes présentées dans ce manuscrit offre des réductions de débit significatives par rapport à la référence. La combinaison d'une partie de ces méthodes permet d'obtenir un gain moyen de 20% par rapport au standard H.264/AVC pour un ensemble de séquences 1080p.

**Mots clés** H.264/AVC, KTA, Codage par compétition, Critère RD, Codage Intra, Codage Inter, Vecteurs mouvement, Partitionnement 1D, Elimination de prédicteurs.

**Abstract** This thesis deals with the improvement of video compression algorithms. The goal is to create methods that are efficient and realistic enough to be submitted to video standardization committees, in order to be integrated in a future video standard. The approaches selected are the use of competition based scheme and the suppression of competition information. The proposed schemes described in this thesis are related to the Intra and Inter coding.

In order to improve the coding of the motion information, a competing framework for better motion vector coding was integrated in the KTA. This method takes advantage from the use of the spatial and temporal redundancies in the motion vector field where the selection is based on the rate distortion criterion. Moreover, to select the optimal motion vectors predictors set, an automatic adaptation of the predictors sets according to the video content is also proposed. To improve the Intra prediction, a new Intra coding mode, based on a 1D macroblock partitioning, is described. This mode enables to reduce the spatial distance between the reference signal and the current partition compared to the block modes of the norm. This Intra 1D mode is competing with the other existing Intra modes to increase the amount of Intra coding possibilities.

The recent video compression standard H.264/AVC is based on a predictive coding. This standard offers a much higher number of competitions than his predecessors for both Intra and Inter coding. Consequently, the competition information is highly represented in the global bit rate, especially at low bit rate. In order to reduce that bit rate linked to the competition information, a part of the encoder intelligence is transferred to the decoder. This transfer enables to eliminate some motion vectors predictors indexes related to the motion vectors competition scheme. Moreover, it also enables to erase some Intra predictors according to the quantization parameters and to the reference signal. Eventually, in the same way, by considering the motion information as a competition information, we propose a motion estimation at the decoder side. The experimental results show the efficiency of this scheme, especially at low bit rate.

All proposed schemes, in this manuscript, improve the coding efficiency compared to the reference. The combination of some of these methods offers an average percentage bit rate reduction of 20% compared to the H.264/AVC standard for a set of 1080p sequences.

**Keywords** H.264/AVC, KTA, Competition based scheme, RD criterion, Intra coding, Inter coding, Motion vector, 1D partitioning, Elimination of predictors.