Université Mohamed V-Agdal & Université de Poitiers

Faculté Des Sciences Fondamentales et Appliquées (SP2MI) de Poitiers & Faculté des Sciences de Rabat

SIC, UMR CNRS 6172 & GSCM-LRIT UFR Informatique & Télécommunications École doctorale Sciences et Ingénierie pour l'Information S2I, Limoges & Centre d'études doctorales en Sciences et Techniques, Rabat

THÈSE EN CO-TUTELLE

Pour obtenir le titre de **Docteur en Informatique et applications**

Discipline : Sciences de l'ingénieur Spécialité : Informatique & Télécommunications

Présentée par

Miryem HRARTI

OPTIMISATION DU CONTROLE DE DEBIT DE H.264/AVC BASEE SUR UNE NOUVELLE MODELISATION DEBIT-QUANTIFICATION ET UNE ALLOCATION SELECTIVE DE BITS

Dirigée par :

Driss ABOUTAJDINE Ahmed TAMTAOUI Abdelhakim SAADANE Mohamed-Chaker LARABI

Soutenue le 10 Décembre 2011 devant le jury composé de :

E. MOUADDIB	Professeur (UPJV, Amiens, France)	(Président)
M. DAOUDI	Professeur (Télécom Lille 1, France)	(Rapporteur)
A. HAMMOUCH	Professeur (ENSET, Rabat, Maroc)	(Rapporteur)
A. SAADANE	Maître de Conférences-HDR (Ecole Polytechnique de Nante	es)
M-C. LARABI	Maître de Conférences (Université de Poitiers)	
D. ABOUTAJDINE	Professeur (Faculté des Sciences Rabat, Maroc)	
A. TAMTAOUI	Professeur (INPT, Rabat, Maroc)	

REMERCIEMENTS

Le travail présenté dans cette thèse représente le fruit d'une collaboration franco-marocaine entre le Laboratoire de Recherche en Informatique et Télécommunications (LRIT) de la Faculté des Sciences de Rabat, Université Mohamed-V, Maroc, dirigé par le Professeur Driss ABOUTA-JDINE et le Laboratoire Signal, Image et Communications (SIC) de la Faculté des Sciences de Poitiers, France, dirigé par le Professeur Christine FERNANDEZ-MALOIGNE.

En rédigeant cette dernière page de manuscrit, je suis obligée de reconnaitre que la réalisation de cette thèse a exigé beaucoup de patience et de dévotion personnelle, tout au long ces nombreuses années passées au sein des deux laboratoires. Ce fruit récolté issu de beaucoup de recherche et de beaucoup d'aide reçue de nombreuses personnes que je tiens à remercier ici.

Je ne pourrais pas commencer les remerciements sans parler de la personne qui m'a promis qu'ensemble, nous allons œuvrer à l'achèvement d'une thèse de qualité, mon directeur de thèse, Hakim SAADANE, Professeur à l'École Polytechnique de l'Université de Nantes (France). Je le remercie pour sa présence, ses réunions prolongées au sein de l'IRCCYN et SIC, ses appels téléphoniques à travers la France, ses critiques constructives et surtout ses conseils avisés qui m'ont aidé à ouvrir pas mal de voies dans le domaine de la vidéo.

Une pensée particulière est accordée à mon co-directeur de thèse, Mohammed-Chaker LARABI, Maître de Conférences à l'Université de Poitiers (France), pour sa grande disponibilité, son ouverture d'esprit, sa bonne humeur, ses conseils sincères et fructueux durant nos nombreuses discussions et surtout pour ses empreintes sur mes dossiers d'inscription et ceux des postes ATER. Mes remerciements sont adressés également à mon directeur de thèse Driss ABOUTAJDINE, Professeur à la Faculté des Sciences de Rabat (Maroc), qui a cru en moi et m'a proposée cette cotutelle de thèse internationale et qui m'a soutenu lorsque la signature du contrat de cotutelle a été retardée. J'exprime ma profonde gratitude envers mon co-directeur de thèse Ahmed TAM-TAOUI, Professeur à l'Institut National des Postes et Télécommunications de Rabat (Maroc), pour sa gentillesse et ses encouragements qui n'a cessé de me prodiguer tout au long de l'élaboration de ma thèse.

Que Monsieur El Mustapha MOUADDIB, Professeur à l'Université de Picardie Jules Verne,

Amiens (France), trouve ici, l'expression de mes sincères remerciements d'avoir accepté de présider cette thèse.

Que Monsieur Mohamed DAOUDI, Professeur à Télécom Lille 1 (France) et Monsieur Ahmed HAMMOUCH, Professeur à l'Ecole Normale Supérieure de l'Enseignement Technique de Rabat (Maroc) trouvent mes plus vifs remerciements pour l'intérêt qu'ils ont porté à ce travail, je leur suis très reconnaissante d'avoir bien voulu en être les rapporteurs et de faire partie de mon jury de thèse.

Mes remerciements sont également adressés aux personnels du :

- CNRST (Centre Nationale de Recherche Scientifique et Technique), pour la gestion de la bourse d'excellence marocaine.
- SCAC (Service de Coopération et d'Action Culturelle) de l'Ambassade de France au Maroc, pour le deuxième financement de cette thèse.
- EGIDE, centre français pour l'accueil et les échanges internationaux, pour l'organisation et la gestion de mes séjours en France.

Un grand merci à mes collègues des deux laboratoires dont la gentillesse et la bonne humeur, ont su rendre ces années agréables. Merci à ceux ou celles qui, de prés ou de loin, ont contribué à l'aboutissement des ces travaux de recherche.

Une pensée profonde comblée d'amour et d'affection à ma mère et feu mon père, lumière de ma vie, mes sœurs et mes frères qui tout au long de ces longues années de thèse m'ont aidé à garder le cap et à croire que tout cela aurait une fin heureuse. Je tiens à vous remercier pour le soutien du début jusqu'à la fin de ce travail. Même si mes voyages de 2000Km de mon pays, semblaient mystérieux, vous savez maintenant que je travaillais. Malgré ma présence en pointillés dans l'environnement familial, vous m'avez toujours soutenu et une très grande partie du travail que j'achève maintenant vous est due! Mille fois merci.

Je n'oublierai certainement pas mes copines qui m'ont soutenue durant des moments ou j'ai eu l'impression de courir derrière un mirage. Merci d'avoir partagé mes moments de bonheur et de tristesse.

A mon mari, mon rayon du bonheur qui s'est infiltré splendidement dans ma vie, qui était toujours à mes cotés, qui a su enrichir ce qui était en moi et qui a fait naître et nourrir de nouvelles valeurs de la vie. Merci pour tout.

Résumé

L'explosion des applications multimédias est en grande partie liée à l'efficacité des techniques de compression utilisées. H.264/AVC, également appelée MPEG-4 Part 10, est la dernière née des normes de codage vidéo. Elle est plus efficace que les standards précédents (MPEG1, MPEG2part 4, H26x...) et permet d'atteindre des gains de compression significatifs. Comme pour les autres standards, le contrôle de débit est un élément clé de H.264/AVC car il permet de réguler la qualité visuelle de la séquence reconstruite tout en respectant les contraintes imposées par le canal de transmission. Conformément au débit cible spécifié, le contrôle de débit détermine d'une manière appropriée les paramètres de quantification. Fondamentalement, une première fonction Débit-Quantification modélise le débit en fonction du paramètre de quantification (QP) et une deuxième fonction Distorsion - Quantification estime la distorsion (ou la qualité) de la vidéo reconstruite en fonction du QP utilisé. L'ensemble de ces deux fonctions conduit à une relation (généralement quadratique) liant QP au nombre de bits cible et aux statistiques de l'unité de codage de base (Trame, Slice, Macrobloc ou ensemble de Macroblocs). Ces fonctions ont naturellement fait l'objet de plusieurs études. Les modèles qui ont été adoptés et donc préconisés par le groupe de normalisation ne présentent pas en général les meilleures performances et sont loin d'être optimaux.

Cette thèse s'inscrit dans ce contexte. Son objectif principal est de développer et de concevoir de nouvelles techniques visant à améliorer les performances du contrôle de débit et par là même celles de la norme H.264/AVC. Ces techniques s'appuient à la fois sur une analyse fine des principales limitations actuelles et sur une large étude bibliographique pour proposer une détermination plus appropriée du paramètre de quantification, une allocation de bits plus fine et une intégration des propriétés du système visuel humain pour une meilleure qualité de la vidéo reconstruite.

Pour une meilleure estimation du paramètre de quantification, deux modèles Débit-Quantification (R-Q) sont proposés. Le premier modèle conçu pour les trames Intra, est un modèle non-linéaire. Il est utilisé pour déterminer de manière optimale le paramètre de quantification initial, tout en exploitant la relation entre le débit cible et la complexité des trames Intra. Le second modèle, logarithmique, est conçu pour les unités de codage Inter. Outre ses performances, ce dernier mo-

dèle se substitue aux deux modèles utilisés par le contrôleur de débit H.264/AVC, et offre ainsi une réduction de la complexité calculatoire.

L'allocation de bits au niveau trame du profil de base du standard H.264/AVC reste basique. Elle suppose que les caractéristiques des GOPs (Group Of Pictures) sont similaires et alloue équitablement le nombre de bits cible aux unités de codage indépendamment de leur complexité. Pour une allocation plus fine, le nouveau modèle proposé exploite deux mesures de complexité. La première est un rapport de mouvement déterminé à partir de bits réels utilisés pour coder les trames précédentes. La seconde mesure exploite la différence entre trames adjacentes et l'histogramme de cette différence.

Enfin, et pour un meilleur contrôle de la qualité visuelle de la vidéo reconstruite, une carte de saillance est intégrée au processus d'allocation de bits. La carte de saillance qui est générée par une approche Bottom-up, simule l'attention visuelle humaine. Au niveau trame elle est utilisée pour réajuster le paramètre de quantification. Ce réajustement permet d'attribuer plus de bits aux trames contenant plus de régions saillantes (censées donc être plus importantes que les autres). Au niveau macrobloc, la carte de saillance est exploitée pour répartir efficacement le nombre de bits entre les macroblocs d'une même trame. Cette répartition par "région" d'intérêt permet d'améliorer la qualité visuelle de la trame.

Les nombreuses simulations effectuées montrent que chacun des modèles proposés, quand il est comparé avec deux algorithmes récents du contrôle de débit (JM15.0 et JVT-O016), apporte des améliorations significatives aussi bien au niveau du débit moyen que du PSNR moyen. Une qualité plus consistante et donc un lissage de qualité à travers les trames est également observé.

Mots-clés : Codeur Vidéo H.264/AVC, Allocation de bits sélective, Unité de base (Trame, Macrobloc), Système Visuel Humain, Fonction de Sensibilité au Contraste, Modèle Débit-Quantification, Carte de saillance, Complexité spatiale/temporelle.

ABSTRACT

The explosion of multimedia applications is largely due to the efficiency of the compression techniques used. H.264/AVC, also known as MPEG-4 Part 10 is the newest video coding standard. It is more effective than previous standards (MPEG1, MPEG2, part 4, H26x ...) and achieves significant compression gains. As for other standards, the rate control is a key element of H.264/AVC because it helps to regulate the visual quality of the reconstructed sequence while respecting the bandwidth constraints imposed by the channel transmission. In accordance with the specified target bit-rate, the rate control algorithm determines appropriately the quantization parameters. Basically, a first Rate-Quantization function elaborates a relationship between the rate and the quantization parameter (QP). A second function called Distortion-Quantization parameter. These two functions lead together to a relationship (usually quadratic) between the quantization parameter, the target number of bits and the basic unit (Frame, Slice, macroblock or set of macroblocks) statistics. These functions have been the subject of several studies. The models that have been adopted and therefore recommended by the group standardization, do not generally offer better performances and they are far from optimal.

This thesis is in this context. Its main objective is to develop and design new techniques to improve the performance of the rate control algorithm and those of the H.264/AVC standard. These techniques are based on both a detailed analysis of the major current limitations and a wide literature review. Our purpose is to provide a more appropriate determination of the quantization parameter, a selective bit allocation that integrates Human Visual System properties and enhances the reconstructed video quality.

To determine accurately the quantization parameter, two Rate-Quantization models (R-Q) have been proposed. The first model designed for Intra-Frames, is a non-linear one. It is used to determine the optimal initial quantization parameter, while exploiting the relationship between the target bit-rate and the complexity of Intra-Frames. The second model is a logarithmic one and it is designed for Inter coding units. It replaces the two models used by the H.264/AVC rate controller and reduces the computational complexity. The frame layer bit allocation of the H.264/AVC baseline profile remains basic. It assumes that GOPs (Groups Of Pictures) have similar characteristics and the target number of bits are fairly allocated to coding units regardless of their complexity. For a more accurate bit allocation, a new model has been proposed including two complexity measures. The first is a motion ratio determined from the actual bits used to encode the previous frames. The second measure uses the difference between adjacent frames and the histogram of this difference.

Finally, to better control the visual quality of the reconstructed video, a saliency map is included into the bit allocation process. The saliency map generated by a bottom-up approach, simulates the human visual attention. It has been used to adjust the quantization parameter at frame layer. This adjustment allows the assignment of more bits to frames containing more salient regions (supposed to be more important than others). At macroblock layer, the saliency map is exploited to efficiently allocate the number of bits among the macroblocks of the same frame. This bit repartition by "region" of interest improves the visual quality of the frame.

Experimental simulations show that the proposed models, when compared with two recent algorithms of rate control (JVT-O016 and JM15.0), improve significantly the coding performances in terms of average bit-rates and PSNR. More consistent quality and therefore a quality smoothness through frames is also observed.

Keywords: H.264/AVC Video Encoder, Selective bit allocation, Basic Unit (frame, macroblock), Human Visual System, Contrast Sensitivity Function, Rate-Quantization Model, Salience Map, Spatial/Temporal complexity.

TABLE DES MATIÈRES

Re	emer	cieme	nts		i
Re	ésum	né			iii
A	bstra	nct			v
In	trod	uction	Général	$ \mathbf{e} $	1
1	Cor	ntrôle (de débit	dans les standards vidéo	7
	1.1	Fond	ement d	u contrôle de débit	7
		1.1.1	Introdu	uction	7
		1.1.2	Succin	cte description de la théorie de débit-distorsion	9
		1.1.3	Fonctie	$ons \ op{\'e}rationnelles \ du \ d{\'e}bit{-}distorsion \ (ORD) \ \ldots \ \ldots \ .$	10
		1.1.4	Problè	mes d'optimisation débit-distorsion (RDO)	11
	1.2	Algor	ithmes of	du contrôle de débit des codeurs précédents	13
		1.2.1	Algorit	hme du contrôle de débit $TM5$	13
		1.2.2	Algorit	hme du contrôle de débit TMN8	15
		1.2.3	L'algor	rithme du contrôle de débit VM8	18
	1.3	Schér	na détai	llé du contrôleur de débit de H.264/AVC	19
		1.3.1	Introdu	uction	19
		1.3.2	Contrô	leur de débit de H.264/AVC	20
			1.3.2.1	Allocation de bits niveau GOP	21
			1.3.2.2	Calcul du QP initial de chaque GOP (QP Initializer)	22
			1.3.2.3	Estimation de la complexité	23
			1.3.2.4	Allocation de bits au niveau unité de base	23
			1.3.2.5	Modèle Débit-Quantification $(R-Q)$	26

			1.3.2.6 Limiteur du paramètre de quantification (QP Clipping)	27
			1.3.2.7 Modèle du buffer virtuel	27
	1.4	Discu	ssion et conclusion	27
2	Inst	iffisan	ces et avancées récentes du schéma du contrôle de débit de H.264 .	31
	2.1	Insuf	fisances du schéma du contrôle de débit	31
		2.1.1	Détermination des QPs des trames I et des unités de base P	31
		2.1.2	Estimation de la mesure de complexité	34
		2.1.3	Allocation de bits	35
	2.2	Avano	cées récentes du contrôle de débit de H.264/AVC	36
		2.2.1	$QPs \ des \ trames \ I \ et \ des \ unit \acute{es} \ de \ base \ P$ $\ldots \ldots \ldots \ldots \ldots$	36
		2.2.2	Allocation de bits	48
		2.2.3	$Optimisation \ bas{\'e} \ sur \ la \ structure \ hybride \ du \ codeur \ \ . \ . \ .$	57
			2.2.3.1 Sélection du mode de codage et de la taille des unités de codage .	58
			2.2.3.2 Sélection du débit trame	59
			2.2.3.3 Contrôle de débit à plusieurs niveaux	59
	2.3	Discu	ssion et conclusion	60
3	Mod	délisat	ion Débit-Quantification	63
	3.1	Modè	èles débit-quantification (R-Q)	63
		3.1.1	Modèle R - Q de la trame I et la première trame P	63
		3.1.2	Modèle R - Q des unités de codage P	66
	3.2	Analy	yse de la performance des modèles R-Q proposés	68
		3.2.1	Critères d'évaluation	68
			3.2.1.1 Mesure de la qualité	68
			3.2.1.2 Courbe Débit-Distorsion (R-D)	69
			3.2.1.3 Mesure de la complexité de calcul, du débit et du buffer	69
		3.2.2	Environnement expérimental	70
			3.2.2.1 Plateforme	70
			3.2.2.2 Séquences testées	70
		3.2.3	Logiciel utilisé	71
		3.2.4	\tilde{R} ésultats de simulations	72
			3.2.4.1 Performance du modèle Intra R-Q	73
			3.2.4.2 Performance du modèle Inter $R-Q$	75
			3.2.4.3 Performance des modèles combinés Intra et Inter	78
	3.3	Conc	lusion	81
4	Alle	ocation	n de bits au niveau des trames P	87
	4.1	Alloc	ation de bits au niveau trame P	87
		4.1.1	Allocation de bits basée sur des mesures de complexité	87
		4.1.2	Allocation de bits basée sur la saillance visuelle	90

			4.1.2.1 Présentation de la carte de saillance	90
			4.1.2.2 QP ajusté à base de carte de saillance	92
	4.2	Analy	se de la performance des approches proposées	93
		4.2.1	Performance de l'allocation de bits à base des mesures de complexité	94
		4.2.2	Performance de l'allocation de bits basée sur la saillance visuelle 1	.08
	4.3	Concl	usion \ldots \ldots \ldots \ldots \ldots \ldots 1	14
5	Ver	s un co	ontrôle perceptuel du débit H.264/AVC	15
	5.1	Appro	oches basées perception au niveau macrobloc	15
		5.1.1	A just ement perceptuel des QPs des macroblocs P	15
		5.1.2	Allocation de bits au niveau macrobloc P basée saillance visuelle 1	17
		5.1.3	Ajustement des QPs des macroblocs I et P	18
		5.1.4	MAD basé sur les propriétés du Système Visuel Humain 1	18
	5.2	Analy	rse de la performance des approches proposées	20
		5.2.1	$Performance \ d'ajustement \ perceptuel \ des \ QPs \ des \ macroblocs \ P \ 1$	20
		5.2.2	Performance de l'allocation de bits basée saillance visuelle 1	23
		5.2.3	Performance de l'approche d'ajustement des QPs des macro-	
			blocs I et P	.28
		5.2.4	Performance du MAD utilisant les propriétés du SVH 1	.30
	5.3	Concl	usion \ldots \ldots \ldots \ldots \ldots \ldots 1	.35
Co	onclu	sion g	${ m \acute{e}n\acute{e}rale~et~perspectives~\ldots~\ldots~\ldots~\ldots~.}$	37
Li	ste d	es pub	lications	41
		p as		
	Anr	nexe .		43
\mathbf{A}	nnex	е		43
		A1	Standard H.264/AVC	43
			A1.1 Introduction	43
			A1.2 Couche codage vidéo	.44
			A1.3 Profils et niveaux	45
			A1.4 Description technique	47
			A1.5 Prédiction Intra-Image	47
			A1.6 Prédiction Inter-Image	52
			A1.7 Transformation et Quantification	55
			A1.8 Codage entropique 1	.60
			A1.9 Filtre de suppression des effets de blocs 1	64
Bi	bliog	graphie	$\cdot \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots $	65

LISTE DES TABLEAUX

3.1	Conditions de test
3.2	Comparaison entre le modèle Intra R-Q et JM15.0 en termes du QP initial, des gains en PSNR et en débit binaire
3.3	Comparaison entre le modèle Inter R-Q et JM15.0 en termes des gains en PSNR, débit binaire et temps de calcul
3.4	Comparaison entre les modèles combinés Intra&Inter et JM15.0 en termes des gains en PSNR, débit binaire et temps
4.1	Comparaison des moyennes de PSNR et des différences de débit binaire entre le logiciel de référence JM15.0 et l'approche proposée "Notre+JM10.2" et entre cette dernière implémentée au sein du JVT-O016 ("Notre+JVT-O016") et l'algorithme du JVT-O016
4.2	Comparaison de la moyenne du PSNR et celle du SSIM et les ratios de débit binaire entre le logiciel de référence JM15.0 et l'approche proposée
5.1	Comparaison de la moyenne du PSNR et celle du SSIM et du débit binaire entre le JM15.0 et l'approche d'ajustement du QP appliquée au niveau macrobloc 120
5.2	Différences des moyennes des débits binaires, PSNR et PSNR-HVS entre le logiciel de référence JM15.0 et l'algorithme proposé "Algo_global"
5.3	Comparaison des gains en termes de débit binaire, PSNR, PSNR-HVS et SSIM entre le logiciel de référence JM15.0 et l'algorithme proposé " $RQ+QP_{Ajuste}$ " 128
5.4	Comparaison entre l'algorithme de référence JM15.0 et l'approche utilisant le MAD_{CSF} en termes de différence des débits binaires et des scores subjectifs
A1	Signification des modes de prédiction des blocs $4x4$ de luminance. \ldots \ldots 149
A2	Signification des modes de prédiction des blocs 16x16 de luminance
A3	Tailles du pas de quantification dans H.264/AVC

A4	Valeurs du PF selon la position de pixels	158
A5	Valeurs du facteur de multiplication MF selon le QP et la position (i,j)	159
A6	Mots-codes du codage de Golomb exponentiel	161
A7	Types de codage du paramètre v	162
A8	Partie de la table du paramètre "coded-block-pattern"	162

TABLE DES FIGURES

1.1	Structure d'un système multimédia	8
1.2	Variations du débit binaire de séquences vidéo à contenus différents	9
1.3	Fonction du débit-distorsion d'une source gaussienne.	10
1.4	Courbe opérationnelle R-D et celle d'un modèle R-D	11
1.5	Diagramme du contrôle de débit TMN8 au niveau macrobloc.	17
1.6	Processus de codage d'un macrobloc relié au contrôle de débit	19
1.7	A- Codage à débit binaire variable (VBR), B- Codage à débit binaire constant	
	(CBR)	20
1.8	Eléments constitutifs du contrôleur de débit de H.264/AVC	21
2.1	Impact du QP initial sur les performances R-D des Séquences vidéo (100 trames,	
	QCIF, 30 Hz, 10 KBps) \ldots	33
2.2	Comparaison du MAD réel et celui prédit par le modèle linéaire de la séquence Coastguard-Akiyo codée à 80Kbps.	33
2.3	Relation entre le MAD et le nombre de bits de texture (3 ^{ème} trame de la séquence "Foreman" codée à QP=25).	34
2.4	Comparaison des cartes de distorsion en utilisant (A) SSIM et (B) MAD pour la 36 ^{ème} trame de la séquence de Foreman codée à 20Kbps	35
2.5	Similarité entre le débit binaire et le gradient de la séquence "Foreman" : A- Gra-	00
	dient des différentes trames, B- Débit binaire des différentes trames codées en Intra.	38
2.6	Relation entre le débit binaire et le gradient de la séquence "Foreman" codée à	
	QP=28.	39
2.7	Relation entre le débit binaire et la mesure de complexité : Grad*SOH des sé- quences : A- "Stefan" codée à QP_I=20, B- "Foreman" codée à QP_I=32, et C-	
	"Football" codée à QP_I=40	40
2.8	Relation entre PSNR et QP de la séquence de "Foreman"	41
2.9	Relation entre le débit binaire R et le facteur - ρ de la Séquence "Foreman"	43

$\begin{array}{c} 2.10\\ 2.11\end{array}$	position d'une unité de base (Trame ou macrobloc) m	44
	R-D, B- : Débit binaire des composantes Y, UV	46
 2.12 2.13 2.14 	Comparaison entre l'algorithme du contrôle de débit de H.264 et celui utilisant la technique proposée (Séquence "Foreman" codée à 64Kbps)	47 53 56
3.1	Relation entre le rapport (RI/δ) et le débit binaire total RT des séquences vidéo "Carphone" "Coastguard" "Foreman" et "Suzie"	64
3.2	Relation entre le rapport (δ/RI) et QP_I des séquences vidéo "Carphone", "Coast-	01
3.3	guard" et "Foreman"	65
3.4	Allure typique d'une courbe débit-distorsion (B-D)	69
3.5	31 ^{ème} Trames des séquences vidéo : A-"Akiyo", B-"Bridge-Close", C- "Carphone", D-"Claire", E-"Coastguard", F-"Container", G-"Foreman", H-"Grandma", I-"Highway'	",
	"Suzie"	70
3.6	Comparaison des valeurs du PSNR/Trames P entre JM15.0 et le modèle Inter. Séquences manipulées : A- "Silent" codée à 35Kbps et B- "News" codée à 20Kbps.	77
3.7	Comparaison du nombre de bits/Trames P entre JM15.0 et le modèle <i>Inter</i> . Sé- quences manipulées : A- "Silent" codée à 35Kbps et B- "News" codée à 20Kbps.	77
3.8	Comparaison de la qualité en termes de PSNR entre JM15.0 et les modèles In- tra&Inter. Séquences manipulées : A- "Akiyo", B- "Bridge-Close", C- "Container",	
2.0	D- "Grandma", E- "News" et F- "Salesman" codées à différents débits binaires	80
3.9	Comparaison du temps de codage entre JM15.0 et les modeles Intra&Inter. Se- quences manipulées : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grand- ma", E- "News" et F- "Salesman" codées à différents débits binaires.	82
3.10	Comparaison des valeurs du PSNR/Trames P entre JM15.0 et les modèles In- tra&Inter combinés. Séquences manipulées : A- "Bridge-Close" codée à 15Kbps, B- "Container" codée à 45Kbps, C- "Grandma" codée à 20Kbps et D- "Salesman"	0
3.11	codée à 30Kbps	83
	30Kbps	84
4.1	Complexité de mouvement $C_1(1, i - 1, n)$, A- et répartition de bits/Trame, B- De la séquence "Foreman" codée à 20Kbps	89
4.2	Algorithme de génération de la carte d'importance perceptuelle	91

Cartes de saillance de la : 75 ^{ème} trame de la séquence "Container" : A1- Originale, B1- Luminance et C1- Carte de saillance, 31 ^{ème} trame de la séquence "Mother- Daughter" : A2- Originale, B2- Luminance et C2- Carte de saillance, et 56 ^{ème} trame de la séquence "Salesman" : A3- Originale, B3- Luminance et C3- Carte de saillance	92
Courbes Débit-Distorsion de différentes séquences codées à des débits binaires va- riables lorsque l'approche "Notre+JM10.2" est appliquée et comparée avec l'algo- rithme du JM15.0 : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grandma", E- "News" et F- "Salesman"	95
Courbes Débit-Distorsion de différentes séquences codées à des débits binaires variables lorsque l'approche "Notre+JVT-O016" est appliquée et comparée avec l'algorithme du JVT-O016 : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grandma", E- "News" et F- "Salesman"	96
Courbes Débit-Distorsion de différentes séquences codées à des débits binaires va- riables lorsque l'approche "Notre+JVT-O016" est appliquée et comparée à la fois avec les algorithmes du JM15.0 et du JVT-O016 : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grandma", E- "News" et F- "Salesman"	97
Variations du PSNR (A, C) et Répartition des bits (B, D) à travers les trames P des séquences "Bridge-Close" et "Silent" codées à 35 et 64Kbps respectivement, lorsque l'approche d'allocation de bits combinée avec le modèle Intra R-Q est comparée avec ce dernier.	103
Variations du PSNR à travers les trames P de quatre séquences vidéo codées à différents débits binaires dans deux cas : L'approche "Notre+JM10.2" est appliquée et comparée avec JM15.0. "Akiyo" codée à 25Kbps : A et "News" codée à 35Kbps : B. Et l'approche "Notre+JVT-O016" est appliquée et comparée avec le JVT-O016. "Silent" codée à 45Kbps : C et "Salesman" codée à "64kbps" : D	104
Répartition des bits à travers les trames P de quatre séquences vidéo codées à différents débits binaires dans deux cas : L'approche "Notre+JM10.2" est appliquée et comparée avec JM15.0. "Akiyo" codée à 25Kbps : A et "News" codée à 35Kbps : B. Et l'approche "Notre+JVT-O016" est appliquée et comparée avec le JVT-O016. "Silent" codée à 45Kbps : C et "Salesman" codée à "64kbps" : D	105
Variations du niveau du buffer pour quatre séquences codées à différents débits bi- naires et ayant 3GOPs (Période Intra=30) dans deux cas : 1- Le JM15.0 est appli- qué sur "Akiyo" codée à 35Kbps, "Grandma" codée à 45Kbps, "Bridge-Close" codée à 64Kbps et "Container" codée à "45Kbps" (A, B). 2- L'approche "Notre+JM10.2" est appliquée sur les mêmes séquences (C, D)	106
Variations du PSNR de quatre séquences codées à des différents débits binaires et ayant 3GOPs, lorsque l'approche "Notre+JM10.2" est appliquée et comparée avec JM15.0. A : "Akiyo" à 35Kbps, B : "Grandma" à 45Kbps, C : "Bridge-Close" à 64Kbps et D : "Container" à 45Kbps	107
	Cartes de saillance de la : 75 ^{me} trame de la séquence "Container" : A1- Originale, B1- Luminance et C1- Carte de saillance, 31 ^{eme} trame de la séquence "Molther- Daughter" : A2- Originale, B2- Luminance et C2- Carte de saillance, et 56 ^{eme} trame de la séquence "Salesman" : A3- Originale, B3- Luminance et C3- Carte de saillance

4.124.13	Variation du PSNR moyen de quatre séquences : "Akiyo" : A, "Bridge-Close" : B, "Claire" : C et "Container" : D codées à des différents débits binaires variant de 15 à 64Kbps lorsque les approches "Notre+JM10.2" et "Notre+JVT-O016" sont appliquées et comparées respectivement au JM15.0 et au JVT-O016 Comparaison des trames reconstruites de la séquence "Container" codée à 50Kbps. La 62 ^{ème} trame : A- JM15.0 est appliqué (PSNR=38.03dB, 3240Bits/Pic), B- l'ap- proche proposée est appliquée (PSNR=36.81dB, 520Bits/Pic). La 71 ^{ème} trame : C- JM15.0 est appliqué (PSNR=38.05dB, 3272Bits/Pic). D- l'approche proposée est	109
	appliquée (PSNR=38.13dB, 896Bits/Pic)	112
4.14	Comparaison des trames reconstruites de la séquence "Miss-America" codée à 64Kbps. La 25 ^{ème} trame : A- JM15.0 est appliqué (PSNR=43.63dB, 3760Bits/Pic), B- l'approche proposée est appliquée (PSNR=41.15dB, 464Bits/Pic). La 77 ^{ème} trame : C- JM15.0 est appliqué (PSNR=43.08dB, 3680Bits/Pic), D- l'approche proposée est appliquée (PSNR=41.72dB, 1000Bits/Pic)	113
4.15	Comparaison des trames reconstruites de la séquence "Mother-Daughter" codée à 80Kbps. La 11 ^{ème} trame : A- JM15.0 est appliqué (PSNR=44.22dB, 13088Bits/Pic), B- l'approche proposée est appliquée (PSNR=40.41dB, 264Bits/Pic). La 74 ^{ème} trame : C- JM15.0 est appliqué (PSNR=40.31dB, 6952Bits/Pic), D- l'approche proposée est appliquée (PSNR=37.93dB, 784Bits/Pic)	113
5.1	Fonction de Sensibilité au Contraste (CSF 2D) Isotrope et achromatique	119
5.2	Comparaison des trames reconstruites de la séquence "Grandma"codée à 60Kbps. La 25 ^{ème} trame : A- JM15.0 est appliqué (PSNR=38.57dB, 2016Bits/Pic) ; B- l'approche proposée est appliquée (PSNR=36.64dB, 464Bits/Pic). La 84 ^{ème} trame : C-JM15.0 est appliqué (PSNR=37.55dB, 1904Bits/Pic) ; D- l'approche proposée est appliquée (PSNR=38.65dB, 912Bits/Pic)	122
5.3	Comparaison des trames reconstruites de la séquence "Silent" codée à 75Kbps. La 5 ^{ème} trame : A- JM15.0 est appliqué (PSNR=32.93dB, 2800Bits/Pic) ; B- l'approche proposée est appliquée (PSNR=32.93dB, 280Bits/Pic). La 62 ^{ème} trame : C- JM15.0 est appliqué (PSNR=34.93dB, 3344Bits/Pic) ; D- l'approche proposée est appliquée	
۳.4	$(PSNR=35.66dB, 1448Bits/Pic) \dots \dots$	123
5.4	Comparaison des trames reconstruites de la sequence "Claire" codee a 80Kbps. La 18 ^{ème} trame : A- JM15.0 est appliqué (PSNR=43.36dB, 3728Bits/Pic); B- l'approche proposée est appliquée (PSNR=39.76dB, 872Bits/Pic). La 68 ^{ème} trame : C-JM15.0 est appliqué (PSNR=44.93dB, 3912Bits/Pic); D- l'approche proposée est	
55	appliquée (PSNR=43.50dB, 1176Bits/Pic)	124
<u> </u>	(Milieu) Notre approche proposée et (Droit) JM15.0.	127
5.6	La 10 ^{eme} trame de la séquence "News" codée à 45Kbps. (Gauche) 'Irame Originale, (Milieu) Notre approche proposée et (Droit) JM15.0.	127
5.7	La 90 ^{ème} trame de la séquence "Mother-Daughter" codée à 60Kbps. (Gauche) Trama Originale (Milicu) Notre approche proposée et (Dreit) 1M15.0	100
5.8	Salle psycho-visuelle 4/E209 du laboratoire SIC.	$\frac{128}{131}$

5.9	Méthode DSIS.	133
A1	Principe du codage H.264/AVC d'un macrobloc	145
A2	Les parties spécifiques des profils dans $H.264/AVC.$	146
A3	A- Labellisation des échantillons de prédiction (4x4), B- Huit des neuf modes de	
	prédiction Intra-4x4.	148
A4	Modes de prédiction des blocs 4x4 de luminance	148
A5	Modes de prédiction des blocs 16x16 de luminance.	149
A6	Blocs adjacents codés en Intra-4x4.	151
A7	A- Partitions de macrobloc : 16x16, 8x16, 16x8, 8x8. B- Sous-partitions de macro-	
	bloc: $8x8$, $4x8$, $8x4$, $4x4$.	152
A8	Exemple de la prédiction entière et la prédiction sous-pixel	153
A9	Localisation des vecteurs spatiaux utilisés pour la prédiction.	154
A10	Matrices H1, H2 et H3 des trois transformées différentes Appliquées dans H.264/AVC.	155
A11	Ordre de transmission des blocs résiduels dans un macrobloc	156
A12	Parcours en zigzag des blocs 4x4 de luminance.	163
A13	La performance du filtre anti-blocs des images très compressées, sans et avec filtre.	164

INTRODUCTION GÉNÉRALE

Les technologies de l'information et de la communication, généralement définies comme l'ensemble des dispositifs et des systèmes informatiques de stockage, de communication, de traitement et de gestion de données audio et/ou vidéo, connaissent un essor considérable. Cette croissance se manifeste à travers les nouveaux services multimédias qui sont régulièrement proposés. Le développement et la diversité de ces services de communications engendrent un fort accroissement de la quantité des données échangées. Même si les réseaux de communications actuels utilisent des bandes passantes de plus en plus larges et exploitent des espaces de stockage de plus en plus importants, il n'en demeure pas moins que leurs capacités restent limitées et une saturation progressive peut freiner l'apparition de nouvelles technologies.

C'est pour éviter une telle saturation que les méthodes de compression se développent et que des architectures et des algorithmes performants sont régulièrement proposés dans la littérature. Sans compression, une vidéo couleur Haute Définition par exemple, avec une taille de 1920*1080 pixels par image, 25 images par seconde, 8 bits par pixel et au format 4 :2 :2, génère un débit de 102.8 Moctets/sec et n'autorise le stockage que de 43 secondes sur un DVD de 4.5 Goctets.

La compression est un processus qui vise à réduire la quantité d'information nécessaire à la représentation d'un signal sans altérer sa qualité. Pour ce faire, la démarche consiste à exploiter la redondance statistique qui caractérise tout signal. Pour une vidéo, la redondance est à la fois spatiale, due à la corrélation entre pixels voisins, et temporelle (corrélation entre trames voisines). La compression peut être sans pertes ou avec pertes. La compression sans pertes, comme sans nom l'indique, ne tolère aucune perte de données originales. De ce fait elle est réversible. La compression d'information reste limitée dans ce cas et est essentiellement fournie par le codage statistique des données redondantes. Pour des applications nécessitant une réduction plus importante de l'information, la compression avec pertes est plus adaptée. L'opération de quantification qui dans ce cas permet la suppression, et donc la perte irréversible, de l'information redondante et/ou non pertinente introduit des dégradations dans le signal reconstruit. Dans le cas de la compression d'image, l'enjeu est alors de rechercher la représentation la plus compacte de l'image reconstruite qui reste visuellement indiscernable de celle de l'image originale. H.264/AVC est actuellement la dernière norme de codage vidéo. Comparée aux normes précédentes, elle offre une qualité visuelle équivalente à un débit deux fois plus faible. Elle possède les mêmes éléments fonctionnels de base que les autres standards : une transformée pour la réduction de la corrélation spatiale, une quantification pour la réduction et le contrôle du débit binaire, une prédiction compensée en mouvement pour la minimisation de la corrélation temporelle et un codage entropique pour la réduction de la corrélation statistique. Elle se différencie des autres normes dans les détails de chaque élément fonctionnel : une prédiction Intra-image, une nouvelle transformation DCT entière de blocs de taille 4x4, de multiples images de référence, des tailles variables de blocs et une compensation de mouvement jusqu'au quart de pixel, un filtre de déblocage et un codage entropique amélioré. Cette norme permet en outre d'envisager une grande diversité d'applications grâce à la définition de profils intégrant chacun différents éléments. Le profil de base est ainsi plus orienté vers les applications comme la vidéo-téléphonie, la vidéoconférence, la vidéo mobile et les communications sans fil. Le profil principal est destiné à la télévision numérique et le stockage, alors que le profil étendu est plus adapté aux applications de diffusion de média et de vidéo streaming.

• Contexte de la thèse

Les travaux menés dans le cadre de cette thèse s'inscrivent dans le contexte des technologies émergentes en général et celui des applications multimédias en particulier. Ils se focalisent sur la compression de la composante vidéo et sa transmission à très bas débits. Ils s'intéressent au processus du contrôle de débit de la norme H.264/AVC. Ce processus qui ne fait pas partie de la norme, permet d'ajuster les paramètres de quantification pour satisfaire les contraintes imposées par la bande passante du canal de transmission. Pour des applications à débits constants, le contrôle de débit cherchera à faire correspondre le débit réel au débit cible en tenant compte des contraintes imposées par le buffer. Pour des applications à débits variables, le contrôle de débit autorisera la fluctuation du débit de sortie pour maintenir une qualité constante. Fondamentalement, le contrôle de débit inclut deux parties : une allocation de bits et une réalisation de cette dernière via le calcul du paramètre de quantification (QP). Il exploite également deux modèles : un premier pour établir une relation entre le débit de sortie et le paramètre de quantification utilisé et un second pour estimer la distorsion de la vidéo reconstruite en fonction de QP. Ces modèles ont, naturellement fait l'objet de plusieurs études. Les modèles qui ont été adoptés et donc préconisés par le groupe de normalisation ne présentent pas en général les meilleures performances et sont loin d'être optimaux pour les applications orientées très bas débits.

• Objectifs de la thèse

L'objectif principal de cette thèse est de développer et de concevoir de nouvelles techniques permettant d'améliorer les performances du contrôle de débit dans la norme H.264/AVC. Ces techniques s'appuient sur une étude préalable des défauts affectant le contrôle de débit actuel pour proposer une allocation de bits plus fine, une détermination plus appropriée du paramètre de quantification et une intégration des propriétés du système visuel humain pour une meilleure qualité de la vidéo reconstruite. La démarche qui en résulte s'articule autour de quatre axes.

- ◇ Le premier axe s'intéresse aux unités de codage Intra (I) et cherche à améliorer leur codage. Comme il existe une dépendance entre les trames I et les trames Inter (P) qui les suivent, le codage de la séquence vidéo globale est également amélioré. Dans H.264/AVC, le paramètre de quantification initial est estimé sans aucune prise en considération des caractéristiques de la trame I. La démarche consiste donc à remédier à cette lacune en cherchant à s'appuyer sur la complexité de la trame I et sur son débit binaire cible pour déterminer d'une manière optimale son paramètre de quantification initial.
- ◇ Le second axe de recherche s'intéresse aux unités de codage Inter (Trames P ou macroblocs P). Dans H.264/AVC, les deux modèles utilisés, exploitent une mesure de complexité qui est basée sur le principe de stationnarité et de linéarité. Cette supposition devient approximative dans le cas où les séquences vidéo sont mouvementées ou à changement de scènes. Ceci contribue à la défaillance de l'algorithme adopté par H.264/AVC et à la dégradation de la qualité de la séquence reconstruite. Pour remédier à cette altération de la qualité, la démarche, similaire à celle du premier axe, consiste à d'abord mieux caractériser les trames P. En s'appuyant sur cette nouvelle caractérisation, il s'agira ensuite d'établir pour ces trames, une relation plus fine et dans la mesure du possible moins complexe que celle de H.264/AVC, entre le pas de quantification et le débit binaire cible.
- Le troisième axe se focalise sur une nouvelle approche d'allocation de bits au niveau des trames P. Dans l'algorithme du contrôle de débit adopté par H.264/AVC, l'allocation de bits au niveau trame s'effectue en divisant équitablement le budget de bits disponible entre les images P sans tenir compte de leurs complexités. Pour pallier cette déficience et mieux contrôler la qualité perçue de la séquence reconstruite, la démarche cherche à développer des approches permettant d'allouer plus de bits aux trames à complexité élevée. Pour une allocation fine et précise, deux approches peuvent être envisagées. Dans la première, ce sont les caractéristiques statistiques des trames qui seront exploitées pour définir des mesures de complexité. La deuxième approche s'appuiera sur les propriétés du Système Visuel Humain en général et sur les modèles d'attention visuelle en particulier. Ces modèles génèrent des cartes de saillance qui peuvent être utilisées pour définir de nouvelles mesures de complexité visuelle. L'ensemble de ces mesures, statistiques ou visuelles, permettent une hiérarchisation des trames en fonction de leur importance et donc une répartition efficace du budget de bits disponible.
- ◇ Dans le dernier axe enfin, l'allocation de bits est naturellement considérée au niveau macrobloc. L'intégration à différents niveaux (Trame et macrobloc) du contenu de la séquence vidéo, permet une allocation plus fine des bits puisqu'à l'intérieur d'un GOP (Group Of Pictures), les trames sont différentes et qu'au sein d'une trame, la complexité des macroblocs varie. Dans le processus d'allocation de bits de H.264/AVC, le nombre de bits cible de chaque macrobloc est estimé en incorporant une mesure de rapport des MADs (Mean Average Difference). Cette mesure ne permet pas toujours une bonne identification des macroblocs importants dans une trame donnée. Une telle identification est hautement sou-

haitable afin de permettre au processus d'allocation de bits de promouvoir les macroblocs importants. L'importance visée ici étant l'importance visuelle, trois voies exploitant toutes la notion de carte de saillance seront exploitées. La première voie cherchera à capitaliser le travail mené au niveau trame. L'ajustement adaptatif du paramètre de quantification est dans ce cas étendu au niveau macrobloc Inter. L'objectif est évidemment de réduire le débit binaire tout en maintenant une qualité visuelle similaire à celle obtenue par H.264/AVC. La deuxième voie cherchera à améliorer le processus d'allocation de bits. La classification en macroblocs saillants et non saillants fournie par la carte d'attention visuelle est exploitée pour ajuster rigoureusement le budget de bits calculé par l'algorithme du contrôle de H.264/AVC. Le "surplus" de bits affecté aux macroblocs saillants est déduit de celui attribué aux macroblocs non saillants. En procédant ainsi, l'objectif est d'améliorer la qualité visuelle tout en respectant la contrainte de débit imposée par le canal de transmission. La troisième voie enfin, exploitera la carte de saillance pour ajuster perceptuellement le paramètre de quantification de toute sorte de macroblocs (Intra et Inter). L'objectif est d'améliorer la qualité visuelle de la séquence vidéo traitée et de réduire considérablement le débit binaire.

• Organisation du manuscrit

L'organisation du manuscrit est directement liée aux objectifs fixés. Elle peut être exhibée comme suit :

- ◇ Le premier chapitre présente des généralités sur la théorie débit-distorsion et les problèmes liés à son optimisation. Les algorithmes du contrôle de débits adoptés par les codeurs vidéo tels que : H.263, MPEG-2, et MPEG-4 sont ensuite succinctement exposés. Une description détaillée du schéma du contrôleur de débit adopté par la norme H.264/AVC est ensuite présentée. La mise en place de cette description facilite la lecture et la compréhension des chapitres suivants. Ce premier chapitre se conclut par une brève présentation de l'évolution des versions du processus du contrôle de débit adoptées par H.264/AVC.
- ◇ Le deuxième chapitre constitue une étude bibliographique des différentes techniques mises en œuvre pour améliorer la performance du processus du contrôle de débit de H.264/AVC. Les principales insuffisances sont d'abord exposées. Les principales solutions bibliographiques pour pallier à ces insuffisances sont ensuite décrites. Les pistes explorées s'orientent en général vers la création de modèles de débit-distorsion et/ou débit-quantification, l'optimisation objective du contrôle de débit permettant d'exploiter les caractéristiques du Système Visuel Humain et l'optimisation basée sur l'exploitation de la structure hybride des codeurs vidéos. D'autres pistes intéressantes, mais n'appartenant à aucune des catégories citées ci-dessus, sont également présentées en fin de chapitre.
- ◊ Le chapitre 3 constitue la première contribution de cette thèse. Il décrit les deux modèles qui sont proposés pour les trames Intra et les unités de base Inter. Ces modèles permettent de résoudre les insuffisances présentées dans le second chapitre et évitent celles des travaux proposés dans la littérature. L'apport de ces modèles qui sont basés sur une étude expé-

4

rimentale assez exhaustive, est évalué sur la base d'une comparaison avec l'algorithme du contrôle de débit préconisé par H.264/AVC (JM15.0).

- ◊ Le chapitre 4 s'intéresse à la seconde contribution de cette thèse et décrit l'amélioration proposée pour l'allocation de bits au niveau trames Inter. Les deux nouvelles approches proposées dans le cadre de cette amélioration sont alors présentées et justifiées. Une évaluation des performances est ensuite proposée pour chacune de ces deux approches.
- ◊ Le dernier chapitre est dédié à l'optimisation du contrôle de débit au niveau macrobloc. Les contributions qu'il propose et qu'il décrit, agissent, au niveau macrobloc donc, sur le processus d'allocation de bits et sur l'ajustement du paramètre de quantification. L'apport de ces contributions est également présenté et discuté.
- ◊ Enfin, ce mémoire s'achève par une partie conclusive sur les principaux apports et résultats obtenus et les perspectives de recherche liées au processus du contrôle de débit.

CHAPITRE CONTRÔLE DE DÉBIT DANS LES STANDARDS VIDÉO

Ce chapitre est consacré à la présentation et à l'étude des algorithmes existants du contrôle de débit dans les standards de codage vidéo. La partie introductive de ce chapitre mettra le point sur l'intérêt d'appliquer un algorithme du contrôle de débit au sein d'un codeur vidéo et présentera brièvement la théorie du débit-distorsion. Trois algorithmes du contrôle de débit adoptés par les codeurs vidéo précédents tels que H.263, MPEG-2 et MPEG-4 seront succinctement présentés pour mieux détailler celui préconisé par la dernière norme de codage vidéo H.264/AVC.

1.1 Fondement du contrôle de débit

Dans les systèmes de codage d'images et de séquences vidéo, la fluctuation du taux de compression est tributaire de la quantité des données que l'on accepte de perdre. La variabilité spatiale et temporelle des contenus vidéo causent des variations dans l'efficacité du codage, qui se traduisent par des fluctuations du débit et de la qualité à la sortie de l'encodeur. Ainsi, le contrôle de débit est la partie centrale des standards de codage vidéo permettant de contrôler les paramètres d'encodage afin de maintenir une qualité consistante sous certaines contraintes imposées par la bande passante du canal de transmission.

1.1.1 Introduction

Les applications multimédias à travers les réseaux de communications sont devenues très répandues grâce à l'augmentation de la bande passante des canaux de transmission et l'amélioration des techniques de compression. Généralement, un système de communication multimédia est composé de quatre niveaux : Application, compression, transport et transmission (*Figure 1.1*).

Situé au niveau compression, le contrôleur de débit représente la partie centrale des standards de codage vidéo. Il permet de maîtriser le volume de données binaires généré. Pour satisfaire



FIGURE 1.1 – Structure d'un système multimédia.

les contraintes imposées par le canal de transmission ou le support de stockage. L'adoption de ce mécanisme permet d'éviter tout débordement de la bande passante allouée par le canal de transmission. Ces débordements causés par la variabilité spatiale et temporelle des données vidéo, se traduisant par des fluctuations du débit et de la qualité à la sortie de l'encodeur. La *Figure 1.2* illustre les variations du débit de certaines séquences vidéo codées par le standard H.264/AVC [7], sans utilisation de l'algorithme du contrôle de débit. Les contenus des séquences choisies varient de ceux comportant beaucoup de mouvement ("Foreman" et "Carphone") à ceux contenant une activité temporelle moyenne et une activité spatiale intéressante ("Salesman" et "Container"). Les changements de débit observés peuvent poser des problèmes de débordement dans la majorité des cas, induisant des dégradations qui peuvent être rédhibitoires au niveau application.

L'algorithme du contrôle de débit constitue une partie informative qui ne figure pas dans la spécification décrivant les standards de codage vidéo. Ceci offre une flexibilité pour développer des schémas adaptés à des applications spécifiques.

Outre le respect du débit binaire imposé, le contrôle de débit doit également prendre en compte les challenges de :

- Contraintes : Imposées par la compression, ces contraintes peuvent être liées à la capacité de transfert des canaux lors d'une transmission sur réseaux. Elles peuvent également être liées à la capacité de stockage disponible ou imposées par la qualité du service fourni à l'utilisateur final.
- Complexité : La mise en place d'un algorithme du contrôle de débit cherchera toujours à minimiser le rapport entre la complexité calculatoire et la performance recherchée.
- Distorsion : Il existe une relation étroite entre le débit et la distorsion. Cette relation



FIGURE 1.2 – Variations du débit binaire de séquences vidéo à contenus différents.

est centrale au codage de séquences d'images, où nous voulons typiquement minimiser la distorsion avec une contrainte sur le débit. La distorsion D est une fonction décroissante du débit binaire R. Ainsi, le problème fondamental dans le contrôle de débit peut être énoncé comme suit :

$$minD$$
 sous contrainte $R \le R_{max}$ (1.1)

avec R_{max} représente le débit maximal imposé par la bande passante du canal de transmission.

Ces challenges sont généralement pris en compte par la théorie du débit-distorsion (R-D : Rate-Distortion) et les techniques d'optimisation qui constituent une brique de base dans la conception d'algorithme du contrôle de débit.

1.1.2 Succincte description de la théorie de débit-distorsion

Initiée depuis plus de cinquante ans par l'œuvre de Shannon [79] [80], cette théorie [17] constitue le fondement théorique du contrôle de débit et un élément central de la théorie de l'information [28]. Elle décrit le compromis entre le taux de compression et la distorsion qui résulte dans le cas d'un codage source avec pertes. Le scénario de base de ce type de codage est le suivant : Une diminution de la distorsion entraîne une augmentation du débit binaire et inversement. Les fonctions R-D sont définies pour décrire la limite inférieure du débit binaire à une distorsion donnée. Par exemple, si la distorsion est mesurée par l'erreur quadratique moyenne (MSE : Mean Square Error), la limite inférieure du débit pour une source gaussienne de variance

 δ^2 , est :

$$R(D) = \begin{cases} \frac{1}{2} \times \log\left(\frac{\delta^2}{D}\right), & Si & 0 \le D \le \delta^2 \\ 0, & Si & D > \delta^2 \end{cases}$$
(1.2)

Alternativement, nous pouvons exprimer la distorsion en fonction du débit :

$$D(R) = \delta^2 \times 2^{-2 \times R} \tag{1.3}$$

La Figure 1.3 illustre une fonction R-D d'une source à distribution gaussienne exprimée par l'Eq. 1.2. Cependant, il n'existe aucune garantie que la limite inférieure du débit puisse être obtenue pratiquement dans les schémas de codage vidéo [93].



FIGURE 1.3 – Fonction du débit-distorsion d'une source gaussienne.

Par rapport à la théorie du débit-distorsion, la théorie opérationnelle du débit-distorsion (ORD : Operationel Rate-Distortion) est plus applicable dans la compression vidéo.

1.1.3 Fonctions opérationnelles du débit-distorsion (ORD)

La théorie opérationnelle R-D s'applique à des sources compressées avec pertes dont le nombre de couples débit-distorsion est fini. Ces couples constituent la fonction ORD [77] et fournissent des informations suffisantes pour que le contrôle de débit soit optimal. Un quantificateur appartenant à l'ensemble des quantificateurs de la fonction ORD doit vérifier :

$$Q_{ORDF} = \{q : q \in Q, R(q) \ge R(p) \qquad \Rightarrow D(q) < D(p), \forall p \in Q\}$$

$$(1.4)$$

 $Q=\{q_0,q_1,..,q_{M-1}\}$ est l'ensemble de tous les quantificateurs admis par le standard de codage, R(qi) et D(qi) représentent respectivement, le débit et la distorsion pour un quantificateur donné. Toutefois, les schémas du contrôle de débit à base de fonctions ORD ne fonctionnent pas efficacement pour un ensemble d'applications de codage vidéo, en particulier pour celles en temps réel, puisque la génération d'une courbe ORD à partir de pairs R-D réels nécessite une complexité calculatoire accrue et engendre un retard intolérable. De plus, tous les couples R-D réels doivent être disponibles, ce qui peut ne pas être possible. Ainsi, de nombreux modèles R-D ont été adoptés par les standards de codage vidéo. Certains supposent que les statistiques de la source sont stationnaires et possèdent une distribution gaussienne [34] ou Laplacienne [26]. Cette supposition a conduit à des modèles basés sur la théorie R-D [26] [72]. D'autres modèles ont été développés empiriquement en exploitant les différentes techniques de régression [24] [25] [29] [49] [61] [92]. Cependant, ces modèles peuvent souffrir des erreurs d'estimation relativement importantes comme le montre la *Figure 1.4*, où les cercles représentent les pairs R-D des quantificateurs admissibles par le standard de codage et la courbe en ligne continue, montre l'approximation élaborée pour l'obtention d'un modèle R-D.



FIGURE 1.4 – Courbe opérationnelle R-D et celle d'un modèle R-D.

Les détails de certains de ces modèles qui sont basés sur les fonctions R-D vont être discutés dans le *Chapitre 2*.

1.1.4 Problèmes d'optimisation débit-distorsion (RDO)

L'objectif général d'une optimisation débit-distorsion est de minimiser la distorsion sous contrainte de débit (Eq. 1.1). Cet objectif peut être spécifié en minimisant la distorsion moyenne (MINAVE) ou la distorsion maximale (MINMAX) [78] des séquences vidéo. Un algorithme du contrôle de débit peut aussi viser le contrôle de la qualité (En minimisant la variation de la distorsion durant la totalité de la séquence vidéo : MINVAR). Comme le débit binaire R_n et la distorsion D_n de la $n^{\text{ème}}$ trame dépendent non seulement de son paramètre de quantification QP_n mais aussi des paramètres de quantification des trames voisines (les *i* trames précédentes et les *j* trames suivantes). Le problème non contraint MINAVE peut être formulé comme suit :

$$QP^* = \{QP_1^*, \dots, QP_N^*\} = \underbrace{argmin}_{(QP_1, \dots, QP_N)} \sum_{n=1}^N D_n(QP_{n-i}, \dots, QP_{n+j})$$
(1.5)

 QP^* représente le paramètre de quantification optimal, N est le nombre de trames considérées dans le problème d'optimisation. Les variables i et j doivent satisfaire les conditions aux limites imposées par le standard. Similairement, le problème du MINMAX peut être formulé comme suit :

$$QP^* = \{QP_1^*, \dots, QP_N^*\} = \underbrace{argmin}_{(QP_1, \dots, QP_N)} \left(\underbrace{max}_{n \in \{1, \dots, N\}} \{D_n(QP_{n-i}, \dots, QP_{n+j})\} \right)$$
(1.6)

Le problème du MINVAR peut aussi être formulé comme suit :

$$QP^* = \{QP_1^*, \dots, QP_N^*\} = \underbrace{argmin}_{(QP_1, \dots, QP_N)} \sum_{n=1}^N |D_n(QP_{n-i}, \dots, QP_{n+j}) - D_{n-1}(QP_{n-i-1}, \dots, QP_{n+j-1})| \quad (1.7)$$

Dans le contexte du codage vidéo, les problèmes non contraints, Eq. 1.5, Eq. 1.6 et Eq. 1.7, intègrent des contraintes telles que celles imposées par la bande passante du canal, le buffer du décodeur et le délai de transmission. Par exemple, lorsque le stockage vidéo est limité par la capacité de stockage R_{max} , le problème du *MINAVE* (Eq. 1.5) devient :

$$QP^* = \{QP_1^*, \dots, QP_N^*\} = \underbrace{argmin}_{(QP_1, \dots, QP_N)} \sum_{n=1}^N D_n(QP_{n-i}, \dots, QP_{n+j})$$

$$sous \quad contrainte \qquad \sum_{n=1}^N R_n(QP_{n-i}, \dots, QP_{n+j}) \le R_{max}$$

$$(1.8)$$

Sous la contrainte du buffer du décodeur, le problème du MINAVE (Eq. 1.5) s'écrit :

$$QP^* = \{QP_1^*, \dots, QP_N^*\} = \underbrace{argmin}_{(QP_1, \dots, QP_N)} \sum_{n=1}^N D_n(QP_{n-i}, \dots, QP_{n+j})$$

$$sous \quad contrainte \qquad 0 \le B_n \le B_{max}$$
(1.9)

 B_n et B_{max} symbolisent respectivement le niveau du buffer et sa taille maximale. Il est à noter également, que pour diverses applications, les contraintes peuvent être différentes et multiples. Comment allouer de manière optimale le nombre de bits disponible, constitue un problème essentiel de l'algorithme du contrôle de débit. Deux approches bien connues ont été appliquées pour résoudre ce problème d'allocation optimale; à savoir l'optimisation de Lagrange [31] et la programmation dynamique (DP : Dynamic Programming) [16]. L'allocation de bits optimale a été abordée dans [38] en utilisant l'approche du multiplicateur de Lagrange dans l'analyse débitdistorsion. Cependant, la méthode du multiplicateur de Lagrange peut souffrir d'un problème fondamental lié à l'obtention de valeurs réelles ou négatives du nombre de bits. Des extensions améliorées de cette technique ont été implémentées en intégrant l'allocation de bits non négative [33]. La méthode lagrangienne est utilisée également dans la conversion d'un problème à contraintes (*MINAVE* ou *MINMAX*) tel que :

min
$$D(QP)$$
 sous contrainte $R(QP) \le R_{max}$ (1.10)

en un problème sans contraintes en utilisant un multiplicateur de Lagrange positif λ :

$$min \qquad D\left(QP\right) + \lambda \times R\left(QP\right) \tag{1.11}$$

La méthode du multiplicateur de Lagrange indique que si $QP^*(\lambda)$ est une solution optimale de l'Eq. 1.11, elle l'est également pour l'Eq. 1.10 [78]. Dans [31], une démonstration par récursivité de ce théorème prouve que $QP^*(\lambda)$ représente effectivement la solution optimale :

$$D\left(QP^{*}\left(\lambda\right)\right) + \lambda \times R\left(QP^{*}\left(\lambda\right)\right) \le D\left(QP\left(\lambda\right)\right) + \lambda \times R\left(QP\left(\lambda\right)\right) \tag{1.12}$$

Un théorème n'ayant pas une solution pour un problème à contraintes (Eq. 1.10) peut trouver une, en résolvant le problème sans contraintes (Eq. 1.11). Cependant, si $R(QP^*(\lambda))$ est égal à la bande supérieure R_{max} , $QP^*(\lambda)$ est la solution optimale du problème à contraintes de l'Eq. 1.10 [78].

La programmation dynamique peut être également employée dans l'allocation de bits afin de réaliser une distorsion globale minimale à travers un arbre (treillis) à quantificateurs bien précis et caractéristiques R-D bien déterminées. L'algorithme de Viterbi (VA) [32] [93] et celui du plus court chemin [27] sont les plus largement employés. Le principe de la programmation dynamique consiste à construire d'abord un treillis de nœuds et cherche le plus court chemin à travers les treillis. Chaque nœud représente une étape. Cette dernière est définie par un vecteur (r, d, s)tel que s indique l'indice de l'étape, r est le nombre total de bits utilisé et d est la somme des distorsions causées par les r bits. Le treillis est construit en choisissant un quantificateur qui crée une nouvelle étape (r, d, s). Finalement, le chemin ayant le coût minimal d est choisi.

La théorie du débit-distorsion a été amplement étudiée par la communauté de la théorie de l'information afin d'assortir les limites et les points réalisables de cette approche. [69] [87] donnent un aperçu général des méthodes débit-distorsion relatives au codage vidéo.

1.2 Algorithmes du contrôle de débit des codeurs précédents

Un algorithme du contrôle de débit est dit efficace lorsqu'il permet d'améliorer les performances débit-distorsion tout en ayant une faible complexité calculatoire et une flexibilité lui permettant de s'adapter à différentes applications. Ainsi, il est utile d'examiner l'évolution de ces algorithmes dans les normes de codage vidéo les plus répandues, car cette évolution peut refléter le progrès des techniques utilisées dans le contrôle de débit.

H.263 [6], MPEG-2 [5], MPEG-4 [8] et H264/AVC ont mis en œuvre une série de modèles de tests destinés à vérifier l'efficacité des algorithmes du contrôle de débit proposés. Parmi les derniers modèles de simulation, nous citons : MPEG-2 TM5 (Test Model 5) [10], H.263 TMN8 (Test Model Near-term 8) [13], MPEG-4 VM8 (Verification Model 8) [12] et H.264/AVC JM (Joint Model) anciennement nommé TML (Test Model Long-term) [85]. Nous allons donner un bref aperçu de ces algorithmes avant de détailler celui préconisé par la norme H.264/AVC.

1.2.1 Algorithme du contrôle de débit TM5

L'algorithme du contrôle de débit TM5 est désigné pour le standard MPEG-2. Cette norme est une extension du standard MPEG-1 [3], permettant un codage très flexible à des débits binaires élevés et une meilleure résiliation d'erreur. Les applications visées par ce standard sont le stockage sur DVD (Digital Versatile Disc) et la diffusion de télévision numérique DTV (Digital Television). En comparaison avec les standards précédents tels que MPEG-1 et H.261 [4], MPEG-2 supporte le mode entrelacé aussi bien que le mode progressif. Une taille de 16x8 pixels par bloc est prise en considération dans la compensation de mouvement. La composante continue (DC) des coefficients transformés par DCT (Discrete Cosinus Transform) est quantifice sur 10 bits au lieu de 8 bits afin de fournir une meilleure qualité de la vidéo. Une quantification non uniforme ainsi que des tables améliorées VLC (Variable Length Code) ont été conçues. La caractéristique majeure de cette norme de codage vidéo consiste en la définition de plusieurs modes de scalabilité (Spatiale, Temporelle, SNR-Signal to Noise Ratio-). Cette norme vise une gamme de débits binaires élevés et définit plusieurs profils et niveaux de souplesse permettant de choisir différents outils pour diverses applications.

L'algorithme TM5 est composé de deux étapes principales. L'allocation de bits constitue la première étape. Elle consiste à déterminer le nombre de bits cible pour coder chaque trame dans un GOP (Group Of Pictures). Dans la seconde étape, le paramètre de quantification QP (Quantization Parameter) est déterminé pour chaque macrobloc en utilisant son activité spatiale et l'état du buffer virtuel.

 Allocation de bits cible : Le but d'une allocation de bits est d'atteindre le débit binaire moyen au niveau GOP. Cette allocation dépend du type d'images traitées : I, P ou B. Pour chaque type, un modèle de complexité est utilisé pour estimer le nombre de bits nécessaire pour coder une trame donnée en utilisant un paramètre de quantification spécifique. Ce modèle est exprimé par :

$$X_I = Sb_I \times QP_I, X_P = Sb_P \times QP_P, X_B = Sb_B \times QP_B \tag{1.13}$$

 X_I , X_P et X_B représentent respectivement les complexités des trames I, P et B. De même Sb_I , Sb_P et Sb_B sont le nombre de bits utilisé pour coder chaque type de trames et QP_I , QP_P et QP_B sont les paramètres de quantification correspondants. Le modèle de complexité est mis à jour après le codage de chaque trame en fonction du paramètre de quantification moyen et du nombre de bits utilisé pour coder une trame donnée.

Le nombre de bits cible est ensuite déterminé pour la trame suivante selon les équations ci-dessous :

$$T_{I} = max \left\{ \frac{R}{1 + \frac{N_{P} \times X_{P}}{X_{I} \times K_{P}} + \frac{N_{B} \times X_{B}}{X_{I} \times K_{B}}}, \frac{BitRate}{8 \times Fr} \right\}$$
(1.14)

$$T_P = max \left\{ \frac{R}{N_P + \frac{N_P \times K_P \times X_B}{K_B \times X_P}}, \frac{BitRate}{8 \times Fr} \right\}$$
(1.15)

$$T_B = max \left\{ \frac{R}{N_B + \frac{N_P \times K_B \times X_P}{K_P \times X_B}}, \frac{BitRate}{8 \times Fr} \right\}$$
(1.16)

Les constantes K_P et K_B sont égales respectivement à 1 et 1.4. R est le nombre de bits restant pour coder les images du GOP courant. Fr et BitRate représentent respectivement, le débit trame et celui utilisé pour coder toute la séquence. T_I , T_P et T_B représentent le nombre de bits cible des trames I, P et B. N_P et N_B sont le nombre des trames P et B.

- Détermination du paramètre de quantification : La valeur du pas du paramètre de quantification est calculée au niveau macrobloc en utilisant le remplissage du buffer. Ce remplissage est prédit suivant l'Eq. 1.17:

$$dd_j^m = dd_j^0 + B_{j-1} - \frac{T_{I,P,B} \times (j-1)}{MB_{number}}$$
(1.17)

Avec m=I,P,B et dd_j^m est la valeur initiale du remplissage du buffer qui correspond au type d'images traitées. B_{j-1} est le nombre total de bits des macroblocs qui précédent le macobloc (j-1). MB_{number} est le nombre total des macroblocs contenus dans la trame. Le pas du paramètre de quantification $QP_{step,j}$ du $j^{\text{ème}}$ macrobloc est ensuite ajusté en fonction du remplissage du buffer dd_j suivant l'Eq. 1.18 :

$$QP_{step,j} = \frac{dd_j \times 31 \times Fr}{2 \times BitRate}$$
(1.18)

Finalement, le paramètre de quantification du $j^{\text{ème}}$ macrobloc est déterminé par :

$$QP_{MB,j} = QP_{step,j} \times N_act_j \tag{1.19}$$

Où N_act_j est l'activité spatiale normalisée du $j^{\text{ème}}$ macrobloc. $QP_{MB,j}$ est borné par 1 et 31.

La totalité de l'algorithme du contrôle de débit TM5 est basée sur une hypothèse supposant que la distorsion augmente linéairement avec le paramètre de quantification et que le débit binaire est inversement proportionnel à cette distorsion. Cette seconde supposition est basée sur un modèle débit-distorsion extrêmement simplifié.

1.2.2 Algorithme du contrôle de débit TMN8

L'algorithme du contrôle de débit TMN8 est désigné pour le standard de codage vidéo H.263. Ce dernier est destiné aux applications vidéo à faible débit. Il a été d'abord développé en 1995 afin d'améliorer la performance de compression de son antécédent H.261aux débits inférieurs à 64Kbps. Les différences entre H.263 et H.261 résident dans : l'adoption d'une précision d'un demi pixel utilisée en compensation de mouvement, une prise en charge de cinq résolutions (QCIF, CIF, SQCIF, 4CIF et 16CIF), une prédiction avancée des trames P/B, similaire à celle utilisée dans MEPG-2 et un codage entropique arithmétique.

15
L'algorithme du contrôle de débit TMN8 inclut un schéma de saut de trames, un contrôle de débit aux niveaux trame et macrobloc.

- Saut de trame : Cette phase inclut les étapes suivantes :

Etape1 : Estumer le nombre de bits du buffer en utilisant le nombre de bits réels employés pour coder la $(k-1)^{\text{ème}}$ trame (R_{k-1}) :

$$Buf = max \left(Buf_{prev} + R_{k-1} - \frac{BitRate}{Fr}, 0 \right)$$
(1.20)

Avec Buf_{prev} représente le nombre de bits du buffer lors du codage de la trame précédente. BitRate et Fr sont le débit binaire total pour coder une séquence donnée et le débit trame. Buf représente le remplissage du buffer.

Etape2 : Déterminer le nombre de trames qui vont être sautées : Si Buf > M, il faut sauter les trames, où M=Bitrate/Fr est le seuil de remplissage du buffer. Pour chaque trame sautée, l'entité (Bitrate/Fr) bits sera déduite du remplissage du buffer. Si L trames sont sautées, la mise à jour du buffer est effectuée comme suit :

$$Buf = max \left(Buf_{prev} + R_{k-1} - (L-1) \times \frac{BitRate}{Fr}, 0 \right)$$
(1.21)

 Contrôle de débit au niveau trame : A ce niveau, le mécanisme du contrôle de débit détermine le nombre de bits cible utilisé pour coder la trame suivante en se basant sur le remplissage du buffer :

$$T = \frac{BitRate}{Fr} - \Delta \tag{1.22}$$

Avec Δ défini comme suit :

$$\Delta = \begin{cases} \frac{Buf}{Fr}, & Si \quad Buf > Z \times M\\ Buf - Z \times M, & Autrement \end{cases}$$
(1.23)

La valeur par défaut de Z est de 0.1. A cause des applications visées par le standard, les faibles variations du nombre de bits cible T sont autorisées.

- Contrôle de débit au niveau macrobloc : A ce niveau, le pas de quantification (QP_{step}) est conçu pour atteindre le débit binaire cible. Sa détermination est basée sur le modèle logarithmique suivant :

$$R\left(QP_{step}\right) = \begin{cases} \frac{e}{ln2} \times \frac{\sigma^2}{QP_{step}^2}, & Si & \frac{\sigma^2}{QP_{step}^2} \le \frac{1}{2e} \\ \frac{1}{2}log_2 \times \left(2 \times e^2 \times \frac{\sigma^2}{QP_{step}^2}\right), & Si & \frac{\sigma^2}{QP_{step}^2} > \frac{1}{2e} \end{cases}$$
(1.24)

R est le nombre de bits cible et σ est l'écart type du macrobloc. Le contrôle de débit est effectué selon les 5 étapes suivantes (*Figure 1.5*) :



FIGURE 1.5 – Diagramme du contrôle de débit TMN8 au niveau macrobloc.

• **Initialisation :** L'estimation du mouvement de la totalité de la trame doit être effectuée avant le contrôle de débit au niveau macrobloc. La somme des variances pondérées des erreurs de prédiction S', est calculée comme suit :

$$S' = \sum_{k=1}^{N} \alpha_k \times \sigma_k \tag{1.25}$$

Avec

$$\alpha_k = \begin{cases} 2 \times \frac{T}{16^2 \times N_{mb}} \times (1 - \sigma_k) + \sigma_k, & Si \\ 1, & Sinon \end{cases} \frac{T}{16^2 \times N_{mb}} < 0.5 \tag{1.26}$$

 σ_k^2 est la variance des composantes de luminance et chrominance du $k^{\text{ème}}$ macrobloc. Si le macroboc est de type Intra, $\sigma_k^2 = \sigma_k^2/3$. T représente le nombre de bits cible de la trame et N_{mb} est le nombre de macroblocs contenus dans la trame.

• Calcul du paramètre de quantification optimisé selon :

$$QP_k^* = \begin{cases} \sqrt{\frac{16^2 \times K \times \sigma_i \times S_i'}{L \times \alpha_i}}, & Si \quad L > 0\\ 2 \times (QP_{Prev} + 2), & Sinon \end{cases}$$
(1.27)

La valeur de la constante L est égale à $\beta_i - 16^2 \times N_i \times C$, où β_i est le nombre de bits pour coder les trames passées, N_i est le nombre des macroblocs qui ont été codées dans la trame actuelle. K et C représentent les paramètres du modèle. QP_{Prev} est la valeur du paramètre de quantification du macrobloc précédent.

Ajustement du paramètre de quantification et codage du macrobloc : QP est fixé à (QP*/2) arrondi à l'entier le plus proche sélectionné dans l'ensemble {1,2,..., 31}. Ensuite le macrobloc est codé avec un QP égal à QP_{prev} +DQUANT, avec DQUANT=QP-QP_{prev}. Si DQUANT>2, alors DQUANT=2. Si DQUANT<-2, alors DQUANT=-2.

- Mise à jour des paramètres du modèle.
- Condition de bouclage : Si i=N_{mb}, le processus est arrêté. Sinon, le macrobloc suivant d'indice (i+1) sera traité.

L'algorithme TMN8 peut atteindre des débits cibles précis, à cause de l'ajustement du paramètre de quantification au niveau du macrobloc. Cependant, le modèle R-D reste trop simple en exploitant uniquement la variance, pour caractériser les statistiques de la trame.

1.2.3 L'algorithme du contrôle de débit VM8

L'algorithme du contrôle de débit VMN8 est exploité par le standard MPEG-4. Ce dernier supporte une large variété d'applications multimédias à diverses fonctionnalités. Dans ce contexte, une scène est considérée comme un ensemble d'objets vidéo (VOs : Video Objects) ayant des propriétés intrinsèques telles que la forme, le mouvement et la texture. Les instances d'un objet donné à un temps donné sont appelées les plans de l'objet (VOP's : Video Object Planes). MPEG-4 adopte un schéma du contrôle de débit évolutif (SRC : Scalable Rate Control) intégré dans le modèle de vérification VM. L'approche de cet algorithme est composée des étapes suivantes :

 Niveau trame : Le contrôle de débit au niveau trame est divisé en quatre étapes : Initialisation, détermination du débit cible, calcul du paramètre de quantification et mise à jour de certains coefficients après codage. Pour calculer le paramètre de quantification, un modèle quadratique débit-quantification (R-Q : Rate-Quantization) est utilisé :

$$R = \frac{X_1 \times S}{QP} + \frac{X_2 \times S}{QP^2} \tag{1.28}$$

 X_1 et X_2 sont les paramètres du modèle qui sont mis à jour par une méthode de régression linéaire, S est la mesure de complexité représentée par le MAD (Mean Absolute Difference), QP est le paramètre de quantification et R est le débit binaire cible.

 Niveau multiple VO (MVO) : La modélisation R-D pour les objets vidéo multiples (MVO : Multiple Video Object) est une simple extension de celle adoptée par un seul VO (SVO : Single Video Object) :

$$R_{i} = \frac{X_{1,i} \times S_{i}}{QP_{i}} + \frac{X_{2,i} \times S_{i}}{QP_{i}^{2}}$$
(1.29)

où i est l'indice de l'objet vidéo.

- **Niveau macrobloc :** Trois étapes composent le contrôle de débit au niveau macrobloc. Une estimation du débit binaire cible débute ces étapes, ensuite, le paramètre de quantification est calculé et finalement une mise à jour du modèle R-D est effectuée. Le nombre de bits R_k^{MB} du $k^{\text{ème}}$ macrobloc est estimé comme suit :

$$R_{k}^{MB} = \begin{cases} MAD_{k}^{2} \times \frac{Y_{1}}{(QP_{k}^{MB})^{2}}, & Si \quad R_{bpp}^{VO} > 0.085\\ MAD_{k} \times \frac{Y_{2}}{(QP_{k}^{MB})^{2}} + MAD_{k} \times \frac{Y_{3}}{(QP_{k}^{MB})^{2}}, & Autrement \end{cases}$$
(1.30)

Avec Y_1 , Y_2 et Y_3 sont les paramètres du modèle. QP_k^{MB} représente le paramètre de quantification et R_{bpp}^{VO} désigne le nombre de bits par pixels de la texture d'un VOP.

Le contrôle de débit SRC adopté par MPEG-4, offre une scalabilité aux différents objets vidéo. Cependant, cet algorithme souffre souvent des grandes erreurs du contrôle relatives à la précision limitée et à la robustesse de son modèle débit-quantification.

1.3 Schéma détaillé du contrôleur de débit de H.264/AVC

1.3.1 Introduction

Le contrôle de débit de H.264/AVC est plus complexe que celui des standards décrits dans la partie précédente, du fait que le paramètre de quantification est utilisé aussi bien dans le contrôle de débit que dans la décision du mode de codage d'un macrobloc, réalisée par la technique d'op-timisation débit-distorsion.

Pour choisir le mode de codage d'un macrobloc par exemple, son paramètre de quantification doit être déterminé en premier en utilisant des statistiques telles que le nombre de bits d'entête et sa mesure de complexité MAD. Or, ces statistiques ne sont disponibles qu'après avoir décidé du mode de codage. Le dilemme qui en résulte est celui typique du "Chicken and egg" (*Figure. 1.6*). La résolution de ce dilemme peut être effectuée en respectant la procédure suivante : Statistiques du MB courant -> Contrôle de débit -> Paramètre de Quantification -> Décision de Mode -> Statistiques du MB courant -> Codage.

Selon cette procédure, le processus du contrôle de débit est exécuté en premier afin de déterminer le paramètre de quantification. Ce dernier sera ensuite exploité dans la décision du mode de codage du macrobloc courant. L'exécution du contrôle de débit a besoin de certaines statistiques. Ces dernières vont être déterminées en utilisant une technique de prédiction linéaire qui consiste à prédire les statistiques du macrobloc courant à partir de celles du macrobloc précédent [9] [56] [57] [85].



FIGURE 1.6 – Processus de codage d'un macrobloc relié au contrôle de débit.

Dans le codeur H.264/AVC, le contrôle de débit peut s'exécuter à trois niveaux : Niveau

GOP, niveau trame et niveau unité de base (qui peut être une tranche (Slice), un macrobloc ou un ensemble de macroblocs). Dans ce qui suit, nous décrivons les éléments constitutifs du contrôleur de débit de H.264/AVC.

1.3.2 Contrôleur de débit de H.264/AVC

Le contrôle de débit de H.264/AVC supporte deux types du contrôle : Le contrôle à débit constant (CBR : Constant Bit-Rate) et le contrôle à débit variable (VBR : Variable Bit-Rate) qui est apparu avec l'avènement des réseaux ATM.



FIGURE 1.7 – A- Codage à débit binaire variable (VBR), B- Codage à débit binaire constant (CBR).

Le codage VBR (*Figure. 1.7A*) est employé dans les applications où les trames vidéo ont besoin d'être présentées avec une qualité plus stable. Ainsi, l'utilisateur doit fournir deux entrées : la vidéo source et la valeur du paramètre de quantification (QP). Le problème majeur se pose au niveau de la sélection du paramètre de quantification adéquat vu que la complexité des images change continuellement dans les séquences vidéo réelles.

En réalité, les contraintes imposées par la taille du buffer du décodeur et la bande passante du canal obligent à coder la vidéo à un débit binaire presque constant. Pour atteindre cet objectif, la *Figure. 1.7B* suggère que la variation du paramètre de quantification doit se faire dynamiquement en se basant sur une estimation de la complexité de la source. Cette estimation permettra de réaliser une allocation de bits appropriée pour chaque macrobloc, image et groupe d'images. Dans le cas d'un tel codage, l'utilisateur doit définir le débit binaire cible au lieu de spécifier le paramètre de quantification.

La *Figure. 1.8* donne un aperçu général sur les différents blocs constituant le contrôleur de débit de H.264/AVC. L'exécution de l'algorithme du contrôle de débit peut être subdivisée en deux étapes importantes :

• Allocation de bits : Exécutée par les deux blocs "GOP bit allocation" et "Basic Unit

bit allocation", cette étape consiste à estimer le nombre de bits cible utilisé pour coder la trame courante (Respectivement le macrobloc courant).

• Détermination du paramètre de quantification : Dans cette étape, une estimation de la complexité de la trame (ou macrobloc) est réalisée via le bloc "Complexity Estimation", ensuite, le paramètre de quantification est calculé en appliquant un modèle quadratique schématisé par le bloc "Quadratic R-D Model". Finalement, le paramètre de quantification déterminé sera borné à l'intérieur du bloc "QP clipping".

Le rôle de l'ensemble de ces blocs sera amplement détaillé dans les sous-paragraphes qui suivent.



FIGURE 1.8 – Eléments constitutifs du contrôleur de débit de H.264/AVC.

1.3.2.1 Allocation de bits niveau GOP

En se basant sur le débit binaire cible "Target bit-rate" fixé par l'utilisateur et sur le remplissage du buffer "Buffer Fullness", le contrôle de débit au niveau GOP calcule le nombre total de bits utilisé pour coder toutes les trames restantes dans chaque GOP.

Lorsque la $i^{\text{ème}}$ image du n^{ème} GOP est codée, le nombre total de bits des trames restantes est

ajusté selon :

$$B_{i,n} = \begin{cases} \frac{BitRate_{i,n}}{Fr} \times N - CBL_{1,n}, & i = 1\\ B_{i-1,n} + \frac{BitRate_{i,n} - BitRate_{i-1,n}}{Fr} \times (N - i + 1) - R_{i-1,n}, & i = 1, 2, \dots, N \end{cases}$$
(1.31)

Fr est le débit trame prédéfini, N est le nombre total des images dans le $n^{\text{ème}}$ GOP, $BitRate_{i,n}$ représente le débit binaire disponible à la $i^{\text{ème}}$ image du $n^{\text{ème}}$ GOP et $CBL_{1,n}$ symbolise l'occupation du buffer virtuel (CBL: Current Buffer Level) dépendant de la trame courante. $R_{i-1,n}$ est le nombre de bits actuels générés par le codage de la $(i-1)^{\text{ème}}$ image. Pour le codage à débit binaire constant, $BitRate_{i,n}$ est toujours égal à $BitRate_{i-1,n}$ et la deuxième formule de l'Eq. 1.31 peut être simplifiée comme suit :

$$B_{i,n} = B_{i-1,n} - R_{i-1,n} \tag{1.32}$$

Le remplissage du buffer virtuel $CBL_{i,n}$ est mise à jour après le codage de chaque image comme suit :

$$CBL_{i,n} = \begin{cases} 0, & n = 1\\ CBL_{i-1,n} + R_{i-1,n} - \frac{CBL_{n-1}(N)}{Fr}, & Autres & GOPs \\ \frac{BitRate_{i-1,n}}{Fr}, & i = 2, 3, \dots, N \end{cases}$$
(1.33)

1.3.2.2 Calcul du QP initial de chaque GOP (QP Initializer)

Pour chaque GOP appartenant à une séquence vidéo, un paramètre de quantification doit être initialisé. Ce dernier est utilisé pour coder la trame I et la première trame P. Il est prédéfini en utilisant le nombre de bits par pixel (bpp : bits per pixel). Ce dernier se base sur la largeur de la bande passante du canal et la longueur du GOP :

$$bpp = \frac{BitRate_{1,1}}{Fr \times N_{pixel}} \tag{1.34}$$

 N_{pixel} est le nombre de pixels par trame.

Pour le premier GOP, le paramètre de quantification initial QP_I est défini comme suit :

$$QP_I = \begin{cases} 35, & bpp \le l_1 \\ 25, & l_1 < bpp \le l_2 \\ 20, & l_2 < bpp \le l_3 \\ 10, & bpp > l_3 \end{cases}$$
(1.35)

Avec $l_1 = 0.15$, $l_2 = 0.45$ et $l_3 = 0.9$ lorsque le format de l'image est QCIF/CIF. Pour les formats plus larges que le CIF, l_1 , l_2 et l_3 prennent respectivement les valeurs suivantes : 0.6, 1.4 et 2.4.

Pour les autres GOPs, le paramètre de quantification initial QP_I_n de la trame I du $n^{\text{ème}}$ GOP est donné par :

$$QP_I_n = \frac{TotalQP_P_{n-1}}{NP_{n-1}} - min\left\{2, \frac{N}{15}\right\}$$
(1.36)

 $TotalQP_P_{n-1}$ est la somme des valeurs de QPs des trames P du $(n-1)^{\text{ème}}$ GOP et NP_{n-1} est le nombre de trames P dans le GOP précédent. La valeur du QP_I_n obtenue est ensuite ajustée :

$$QP_I_n = \begin{cases} QP_I_n - 1, & Si & QP_I_n > QP_I_{n-1} - 2\\ QP_I_n, & Si & QP_I_n \le QP_I_{n-1} + 2 \end{cases}$$
(1.37)

Finalement la valeur du QP_I_n est bornée :

$$QP_I_n = \min\{51, \max\{QP_I_n, 1\}\}$$
(1.38)

1.3.2.3 Estimation de la complexité

Pour estimer la complexité des unités de base (Trame ou macrobloc), le contrôleur de débit utilise le MAD. Pour résoudre le dilemme du "Chicken and egg", précédemment décrit, un modèle linéaire est introduit pour prédire le MAD de la trame courante (ou du macrobloc courant) (MAD_i) à partir de celui de la trame précédente (MAD_{i-1}) (ou de celui du macrobloc précédent).

$$MAD_i = a_1 \times MAD_{i-1} + a_2 \tag{1.39}$$

 a_1 et a_2 sont les coefficients du modèle de prédiction linéaire. Leurs valeurs initiales sont respectivement 1 et 0. Ensuite, elles sont mises à jour après le codage de chaque trame (ou macrobloc).

1.3.2.4 Allocation de bits au niveau unité de base

L'introduction de l'unité de base dans le contrôle de débit, a rendu ce dernier scalable à différents niveaux de granularité tels que : niveau trame, niveau tranche, niveau macrobloc ou ensemble de macroblocs contigus. Le nombre des unités de base dans une trame peut être calculé par :

$$N_{unit} = \frac{N_{mbpic}}{N_{mbunit}} \tag{1.40}$$

 N_{mbpic} et N_{mbunit} représentent respectivement le nombre de macroblocs par trame et par unité de base.

Si l'unité de base est plus petite qu'une image, le bloc "Basic Unit Bit Allocation" de la *Figure. 1.8* sera subdivisé en deux niveaux : l'un pour la trame elle-même et l'autre pour le macrobloc (ou ensemble de macroblocs). Dans ce qui suit, nous allons décrire le contrôle de débit au

niveau trame et au niveau macrobloc.

a) Allocation de bits au niveau trame

Le contrôle de débit au niveau trame permet de déterminer le nombre de bits cible pour coder la trame courante P, tout en se basant sur le nombre de bits cible déterminé au niveau GOP. Le nombre de bits cible est déterminé en fonction de l'occupation du buffer et de la complexité de l'image. Deux étapes sont à citer :

• Etape 1-Détermination de l'occupation du buffer cible : Un niveau de buffer cible (TBL : Target Buffer Level) est prédéfini pour chaque trame P. L'objectif du calcul de ce dernier consiste à déterminer le nombre de bits cible utilisé dans l'estimation du paramètre de quantification de chaque trame P. Puisque le paramètre de quantification de la première trame P est généré au niveau du GOP, TBL est prédéfini à partir de la seconde trame P du $n^{\text{ème}}$ GOP. Il est initialisé comme suit :

$$TBL_{2,n} = CBL_{2,n} \tag{1.41}$$

 $CBL_{2,n}$ est l'occupation du buffer actuel après avoir codé la première trame P du $n^{\text{ème}}$ GOP. Le niveau du buffer cible des autres trames est déterminé par :

$$TBL_{i+1,n} = TBL_{i,n} - \frac{TBL_{2,n}}{NP_n - 1} + \frac{\widetilde{WP_{i,n}} \times (L+1) \times BitRate_{i,n}}{Fr \times (\widetilde{WP_{i,n}} + \widetilde{WB_{i,n}} \times L)} - \frac{BitRate_{i,n}}{Fr}$$
(1.42)

L est le nombre de trames B successives entre deux trames P, NP_n représente le nombre de trames P dans le $n^{\text{ème}}$ GOP, $\widetilde{WP_{i,n}}$ est la moyenne de complexité pondérée de l'image P et $\widetilde{WB_{i,n}}$ est celle de l'image B. Ces complexités pondérées sont calculées comme suit :

$$\widetilde{WP_{i,n}} = \frac{WP_{i,n}}{8} + \frac{7 \times \widetilde{WP_{i,n}}}{8}$$

$$\widetilde{WB_{i,n}} = \frac{WB_{i,n}}{8} + \frac{7 \times \widetilde{WB_{i,n}}}{8}$$

$$WP_{i,n} = R_{i,n} \times QP_P_{i,n}$$

$$WB_{i,n} = \frac{R_{i,n} \times QP_B_{i,n}}{1.3636}$$
(1.43)

 $QP_P_{i,n}$ et $QP_B_{i,n}$ représentent respectivement les paramètres de quantification des trames P et B. Dans le cas où il n'y a pas de trame B entre les deux trames P, l'Eq. 1.42 se ramène à :

$$TBL_{i+1,n} = TBL_{i,n} - \frac{TBL_{2,n}}{NP_n - 1}$$
(1.44)

Il peut être facilement montré que $TBL_{N,n}$ est égal à 0. Ainsi, si le remplissage du buffer actuel est exactement égal au niveau du buffer cible prédéfini, ceci prouve que chaque GOP consomme son propre budget de bits. Cependant, puisque le modèle de débit-distorsion et le modèle de prédiction du *MAD* ne sont pas exacts, une différence subsiste. Pour pouvoir réduire cette différence, une seconde étape est nécessaire.

• Etape 2-Calcul microscopique (Calcul du débit binaire cible) :

Le nombre de bits cible alloué à la $i^{\text{ème}}$ trame du $n^{\text{ème}}$ GOP est déterminé en se basant sur le niveau du buffer cible $TBL_{i,n}$, la largeur de la bande disponible du canal et l'occupation du buffer actuel :

$$\widetilde{T_{i,n}} = \frac{BitRate_{i,n}}{Fr} + \gamma \times (TBL_{i,n} - CBL_{i,n})$$
(1.45)

où γ est une constante de valeur typique égale à 0.5 lorsqu'il n'existe pas de trame B et de 0.25 autrement. Le nombre de bits restant $(B_{i,n})$ est également pris en considération lorsque le nombre de bits est calculé :

$$\widehat{T_{i,n}} = \frac{WP_{i-1,n} \times B_{i,n}}{WP_{i-1,n} \times NP_n + WB_{i-1,n} \times NB_n}$$
(1.46)

 NB_n représente le nombre de trames B non sauvegardées. Le nombre de bits cible est une combinaison pondérée de $\widetilde{T_{i,n}}$ et $\widehat{T_{i,n}}$:

$$T_{i,n} = \beta \times \widehat{T_{i,n}} + (1 - \beta) \times \widetilde{T_{i,n}}$$
(1.47)

 β est une constante de valeur typique égale à 0.5 lorsqu'il n'existe pas de trame B et de 0.9 autrement. Pour se conformer à la condition du HRD (Hypothetical Reference Decoder), le nombre de bits cibles est ajusté comme suit :

$$T_{i,n} = max \{Z_{i,n}, T_{i,n}\}$$

$$T_{i,n} = min \{U_{i,n}, T_{i,n}\}$$
(1.48)

 $Z_{i,n}$ et $U_{i,n}$ représentent respectivement la bande inférieure et supérieure de l'intervalle dont lequel doit varier le nombre de bits cible. Ces deux limites sont calculées par :

$$Z_{i,n} = \begin{cases} \frac{BitRate_{i,n}}{Fr}, & Si = 1\\ Z_{i-1,n} + \frac{BitRate_{i,n}}{Fr} - R_{i,n}, & Autrement \end{cases}$$
(1.49)

$$U_{i,n} = \begin{cases} (BitRate_{i,n} \times t_1) \times 0.9, & Si = 1\\ Z_{i-1,n} + \left(\frac{BitRate_{i,n}}{Fr} - R_{i,n}\right) \times 0.9, & Autrement \end{cases}$$
(1.50)

 t_1 est le temps passé par la première trame dans le buffer des images codées.

b) Allocation de bits au niveau macrobloc

Lorsque l'unité de base n'est pas sélectionnée en tant que trame, un niveau du contrôle de débit est ajouté (niveau macrobloc). Similairement au contrôle de débit au niveau trame, un

nombre de bits cible est à estimer pour chaque macrobloc. Les étapes de cette estimation peuvent être résumées comme suit :

- Etape 1 : Prédire le *MAD* du macrobloc courant dans la trame courante par le modèle linéaire décrit par l'*Eq. 1.39* en utilisant le *MAD* actuel de l'unité de base située à la même position dans la trame précédente.
- Etape 2 : Calculer le nombre de bits de texture $(Texture_{j,i,n})$ du $j^{\text{ème}}$ macrobloc dans la $i^{\text{ème}}$ trame du $n^{\text{ème}}$ GOP. Cette étape est subdivisée en deux sous étapes :
 - * Etape 2.1 : Calculer le nombre de bits cible d'un macrobloc donné :

$$\widetilde{b_{j,i,n}} = Tbr_{i,n} \times \frac{MAD_{j,i,n}^2}{\sum_{k=N_{unit}-j}^{N_{unit}-1} MAD_{k,i-1,n}^2}$$
(1.51)

 $Tbr_{i,n}$ représente, le nombre de bits restants qui seront alloués aux macroblocs non codés dans la trame courante. La valeur initiale de $Tbr_{i,n}$ est égale à $T_{i,n}$.

* Etape 2.2 : Calculer la moyenne des bits d'entête générés par tous les macroblocs :

$$\widetilde{m_{hdr,j}} = \widetilde{m_{hdr,j-1}} \times \left(1 - \frac{1}{j}\right) + \frac{\widetilde{m_{hdr,j}}}{j}$$

$$m_{hdr} = \widetilde{m_{hdr,j}} \times \frac{1}{N_{unit}} + m_{hdr,1} \times \left(1 - \frac{j}{N_{unit}}\right); \qquad 1 \le j < N_{unit}$$

$$(1.52)$$

Où $\widehat{m_{hdr,j}}$ est le nombre de bits d'entête générés suite au codage du $j^{\text{ème}}$ macrobloc de la trame courante. $\widetilde{m_{hdr,j}}$ représente le nombre de bits d'entête du $j^{\text{ème}}$ macrobloc estimé à partir de ceux du macrobloc précédent. $m_{hdr,1}$ symbolise l'estimation de bits d'entête générés par tous les macroblocs de la trame précédente.

* *Etape 2.3* : Calculer le nombre de bits de texture du $j^{\text{ème}}$ marcobloc appartenant à la $i^{\text{ème}}$ trame du $n^{\text{ème}}$ GOP :

$$BitsTex_{j,i,n} = \widetilde{b_{j,i,n}} - m_{hdr} \tag{1.53}$$

1.3.2.5 Modèle Débit-Quantification (R-Q)

Le cœur de l'algorithme du contrôle de débit est un modèle quadratique. Ce dernier est appliqué pour calculer le paramètre de quantification de la trame courante (respectivement le macrobloc courant). Il décrit la relation entre le pas de quantification de la $i^{\text{ème}}$ trame $(QP_{step,i,n})$, sa complexité $(MAD_{i,n})$ et son nombre de bits de texture $(BitsTex_{i,n})$:

$$BitsTex_{i,n} = c_1 \times \frac{MAD_{i,n}}{Q_{step,i,n}} + c_2 \times \frac{MAD_{i,n}}{QP_{step,i,n}^2}$$
(1.54)

 $BitsTex_{i,n}$ est une différence entre le nombre de bits cible $(T_{i,n})$ et le nombre de bits d'entête généré après codage de la trame précédente. Les paramètres du modèle c_1 et c_2 sont mis à jour par une méthode de régression linéaire. Une fois le pas de quantification déterminé, le paramètre de quantification $QP_{i,n}$ de la trame courante (respectivement macrobloc) est calculé par la relation suivante :

$$QP_{step,i,n} = 2^{\frac{QP_{i,n}}{6}} \tag{1.55}$$

La valeur finale du paramètre de quantification est comprise entre 1 et 51 suivant les exigences du standard H.264/AVC.

1.3.2.6 Limiteur du paramètre de quantification (QP Clipping)

Pour garantir la stabilité et minimiser les variations perceptibles de la qualité, un limiteur des valeurs des paramètres de quantification est employé. Ce limiteur est surtout bénéfique dans le cas de séquences vidéo comportant des changements de scènes, où des variations rapides des paramètres de quantification se produisent. Ainsi, ce limiteur est appliqué afin de borner les changements de ces paramètres pour qu'ils ne dépassent pas ± 2 unités entre deux images successives.

1.3.2.7 Modèle du buffer virtuel

N'importe quel décodeur est équipé d'un buffer utilisé pour lisser les variations du débit binaire. Le codeur correspondant doit produire un train binaire simulant le remplissage réel du buffer du décodeur pour que la contrainte liée au buffer soit satisfaite. Le niveau du remplissage du buffer doit être compris entre θ et la *capacité du buffer*. Il est calculé par une différence entre le nombre de bits généré après codage d'une trame donnée et le débit binaire utilisé pour coder la totalité de la séquence. L'utilisateur spécifie des valeurs appropriées au remplissage initial du buffer et à sa capacité.

A la fin du codage d'une unité de base donnée, une étape de post-codage est exécutée. Durant cette dernière, deux objectifs principaux sont à réaliser : Une mise à jour des paramètres du modèle linéaire (a_1 et a_2 de l'Eq. 1.39), et ceux du modèle quadratique (c_1 et c_2 de l'Eq. 1.54), et un contrôle du processus du saut de trames (Frame Skipping) déterminant le nombre de trames susceptibles d'être sautées.

Après avoir codé la trame courante, le nombre de bits générés $R_{i,n}$ est additionné à l'occupation actuelle du buffer. Le processus du saut de trames consiste à contrôler l'occupation du buffer en ajustant le paramètre N_{post} initialisé à 0 et incrémenté jusqu'à ce que la condition suivante soit satisfaite :

$$CBL_{i+N_{nost},n} < B_s \times 0.8 \tag{1.56}$$

 B_s est la taille du buffer. Le remplissage du buffer est alors mis à jour comme suit :

$$CBL_{i+k+1,n} = CBL_{i+k,n} - \frac{BitRate_{i+k+n}}{Fr}; \qquad 1 \le k < N_{post}$$
(1.57)

1.4 Discussion et conclusion

Les algorithmes du contrôle de débit dans les différents standards de codage vidéo ont connu des développements pouvant être résumés comme suit :

- Les modèles débit-distorsion ont été introduits depuis MPEG-4 et H.263. Ces modèles sont utiles dans le processus du contrôle de débit vu qu'ils fournissent suffisamment d'informations pour choisir le paramètre de quantification le plus adéquat. L'introduction des techniques du contrôle de débit fondées sur les modèles débit-distorsion (R-D) a représenté un véritable succès pour les standards de codage vidéo tel que H.264/AVC. Ces modèles sont complexes vu que l'optimisation débit-distorsion est complémentaire au contrôle de débit.
- Pour déterminer le nombre de bits cible pour coder une trame donnée, des techniques du contrôle de buffer ont été introduites. Dans les algorithmes du contrôle de débit des anciens standards de codage vidéo, le paramètre de quantification est mis à jour directement en se basant sur le niveau du buffer. En comparaison avec les algorithmes du nouveaux standard (H.264/AVC), les paramètres de quantification ne dépendent pas directement du niveau de buffer puisque des caractéristiques R-D de la vidéo source sont également disponibles à partir des modèles R-D. Dans les deux normes MPEG-4 et H.264/AVC, l'étape de post-encodage est utilisée pour éviter le débordement (overflow) et le manque de bits (underflow) du buffer.
- Les algorithmes du contrôle de débit dans les normes de codage vidéo sont adaptatifs vu l'exploitation des activités au niveau trame ou macrobloc (VOP-level). La prise en considération des caractéristiques des trames est utile pour construire des modèles R-D précis capables d'ajuster les paramètres de quantification.
- Trois niveaux du contrôle de débit sont supportés dans les normes récentes de codage vidéo telles que MPEG-4 et H.264/AVC alors que ceux du MPEG-2 et H.263 supportent deux niveaux. L'ajout de niveaux dans l'algorithme du contrôle de débit rend ce dernier plus précis.

Plusieurs versions du codeur H.264/AVC ont vu le jour. La structure de ce dernier ainsi que l'algorithme du contrôle de débit qui lui est associé, constituent des champs libres pour les concepteurs. Dans un premier temps, les modifications apportées aux différentes versions ont ciblé l'amélioration des fonctionnalités des composantes constituant la boucle de codage (Transformation, quantification, codage entropique...) tout en résolvant les bugs générés lors de l'exécution du codeur. Le contrôle de débit tel qu'il a été introduit ci-dessus (Section I.3), est dit original et il a été implémenté dans toutes les versions JM du codeur H.264/AVC sans amélioration (du JM1.0 jusqu'au JM14.0 [1]). A partir de la version JM14.1, le schéma du contrôle de débit a subit certains changements. Des modes du contrôle de débit ont été définis. Nous citons à titre d'exemple :

- Le mode 2 : Application du contrôle de débit original (Section I.3) auquel une nouvelle sélection du paramètre de quantification pour les trames I ou B est ajoutée. Cette sélection inclut la notion de codage hiérarchique bidirectionnel.
- Le mode 3 : Application du contrôle de débit original (Section I.3) muni d'un modèle quadratique hybride. De plus, le débit binaire employé pour coder les trames P est calculé en incluant des ratios de bits déterminés suivant le type de la trame et son niveau de prédiction temporelle.

Parallèlement à ces améliorations, le groupe JVT (Joint Video Team) chargé de développer le standard H.264/AVC; a proposé certaines versions améliorées de l'algorithme du contrôle de

débit telles que JVT-G012, JVT-O016 [2] [106] et autres. Ces propositions consistent à améliorer les fonctionnalités d'une ou plusieurs composantes du contrôleur de débit illustré par la *Figure. 1.8.*

Un nouveau modèle quadratique est mis en œuvre dans l'algorithme du JVT-O016 :

$$T_{i,n} = A_{i,n} \times QP_{step,i,n}^2 + B_{i,n} \times QP_{step,i,n} + C_{i,n} + H_{i,n}^P$$

$$(1.58)$$

 $H_{i,n}^P$ est le nombre de bits d'entête de la trame codée précédemment. $A_{i,n}$, $B_{i,n}$ et $C_{i,n}$ représentent les coefficients du nouveau modèle. Ils sont déterminés par une méthode de régression linéaire robuste exploitant les corrélations spatio-temporelles des trames. Une prédiction précise des bits d'entête ainsi qu'un nouveau modèle linéaire estimant la mesure de complexité ont été implémentés dans l'algorithme JVT-O016. Finalement, une optimisation lagrangienne permet de déterminer des paramètres de quantification optimaux tout en minimisant la distorsion totale à travers la séquence vidéo.

L'algorithme du contrôle de débit tel qu'il est introduit dans le standard H.264/AVC, souffre d'une panoplie de failles qui fera l'objet du chapitre suivant au niveau duquel nous allons les détailler et présenter les techniques récentes mises en œuvre pour les corriger afin d'améliorer la performance de l'algorithme du contrôle de débit adopté par le standard H.264/AVC.

Chapitre

2 Insuffisances et avancées récentes du schéma du contrôle de débit de H.264

Dans ce chapitre, nous allons présenter les principales insuffisances liées au contrôleur de débit du standard H.264/AVC. Ensuite, certaines solutions proposées dans la littérature sont exposées. Finalement, nous présentons un ensemble d'avancées récentes pouvant être explorées pour améliorer davantage les fonctionnalités du contrôleur de débit.

2.1 Insuffisances du schéma du contrôle de débit

Le schéma du contrôle de débit TM5 (Test Model 5) [10] utilisé au niveau du MPEG-2 [5], est une base pour la majorité des algorithmes du contrôle de débit décrits dans le chapitre précédent. Le processus d'allocation de bits au sein du TM5, suppose implicitement que les séquences vidéo sont stationnaires, que tous les GOPs possèdent des caractéristiques similaires et que les trames de même type ont une complexité de codage semblable, ce qui est habituellement incertain. A ces approximations s'ajoutent pour le contrôle de débit de H.264/AVC [9] [56] [57] [85], d'autres insuffisances liées principalement :

- A la détermination des paramètres de quantification (QPs) de la trame Intra (I), de la première trame Inter (P) et des autres unités de base P (Trames ou macroblocs).
- A l'utilisation du MAD comme mesure de complexité,
- Et au processus d'allocation de bits aux niveaux trame et macrobloc.

Dans ce qui suit, nous détaillons ces insuffisances tout en en exhibant les solutions proposées dans la littérature pour y remédier.

2.1.1 Détermination des QPs des trames I et des unités de base P

a) Calcul du QP initial de la trame I et la $1^{\acute{e}re}$ trame P

Le choix d'un paramètre de quantification (QP) initial approprié pour les trames I et la première trame P, a un énorme impact sur l'amélioration des performances Débit-Distorsion (R-D) du schéma du contrôle de débit. Les trames P étant prédites directement ou indirectement par compensation de mouvement à partir des trames I, un choix erroné du QP initial générera des erreurs pouvant se propager dans l'ensemble de la séquence vidéo.

La détermination du QP initial des deux premières trames (I et P) du 1^{er} GOP, est basée sur le nombre de bits par pixels (*bpp*) (*Chapitre 1, Eq. 1.34*). Cette grandeur exploite le débit binaire cible, la fréquence trame et le nombre de pixels par trame. Pour les trames I des GOPs suivants, le calcul du QP initial est basé sur la moyenne des paramètres de quantification des trames P du GOP précédent (*Chapitre 1, Eq. 1.36*). Cette détermination manque de précision. D'abord parce que le calcul du *bpp* ne prend nullement en compte les caractéristiques de de la trame I. L'utilisation du *bpp* ensuite génère des variations importantes des valeurs QP.

La Figure 2.1 illustre l'impact du choix du QP initial sur la qualité mesurée par la valeur moyenne du PSNR (Peak Signal to Noise Ratio) de certaines séquences vidéo ("Akiyo", "Bridge-Close", "Carphone", et "Claire"). Ces dernières sont codées à 10Kbps avec quatre valeurs différentes du QP initial (10, 20, 30 et 40). Ce QP est utilisé par la trame I et la première trame P. Pour les autres trames P, les paramètres de quantification sont calculés par les deux modèles décrits dans le *Chapitre 1* et rappelés dans le paragraphe ci dessous. Il ressort de la Figure 2.1 que l'utilisation du *bpp* ne permet pas de déterminer la valeur optimale du QP. Pour la séquence "Bridge-Close" par exemple, la valeur du QP initial optimal (correspondant au PSNR moyen le plus élevé), est égale à 10 alors qu'elle est fixée à 35 par le contrôleur de débit de H.264/AVC. Une détermination adéquate du QP initial peut améliorer la qualité de la totalité de la séquence d'environ +4dB.

b) Calcul des QPs des autres unités de base P

Deux modèles sont utilisés pour estimer les paramètres de quantification du reste des unités de base P à savoir le modèle linéaire (*Chapitre 1, Eq. 1.39*) et le modèle quadratique (*Chapitre 1, Eq. 1.54*). Le modèle linéaire prédit le *MAD* de l'unité de base courante à partir du *MAD* réel de l'unité de base précédente. Le modèle quadratique exploite le *MAD* prédit et le nombre de bits cible de l'unité de base courante pour calculer le pas de quantification.

Le modèle linéaire peut constituer une source d'altération de la performance du contrôleur de débit. Ce modèle suppose que la complexité varie graduellement d'une unité de base à une autre. Pour les vidéos non stationnaires où il y a des changements de scènes, l'information collectée à travers les unités de base précédentes est erronée et engendre donc une mauvaise prédiction du MAD. La Figure 2.2 trace le MAD réel et celui prédit par le modèle linéaire pour une séquence vidéo non stationnaire comportant un seul changement de scène se produisant à la $40^{\rm ème}$ trame. Les valeurs du MAD prédit par le modèle linéaire et celles du MAD réel, obtenu donc après avoir codé les trames, sont proches sauf dans le cas où le changement de scène se produit.

A ce manque de précision du modèle linéaire, se rajoute une approximation grossière induite par le modèle quadratique R-Q. Pour un paramètre de quantification fixe, ce modèle (*Chapitre*



FIGURE 2.1 – Impact du QP initial sur les performances R-D des Séquences vidéo (100 trames, QCIF, 30Hz, 10KBps)



FIGURE 2.2 – Comparaison du MAD réel et celui prédit par le modèle linéaire de la séquence Coastguard-Akiyo codée à 80Kbps.

1, Eq. 1.54) stipule que le nombre de bits de texture est directement proportionnel au MAD. La Figure 2.3, donnant la relation expérimentale entre le nombre de bits de texture et le MAD de chaque macrobloc de la 3^{ème} trame de la séquence "Foreman", montre que la règle de proportionnalité n'est pas vérifiée. Des ajustements sont donc nécessaires pour rendre ce modèle plus précis [94].



FIGURE 2.3 – Relation entre le MAD et le nombre de bits de texture (3^{ème} trame de la séquence "Foreman" codée à QP=25).

2.1.2 Estimation de la mesure de complexité

Dans le contrôleur de débit classique de H.264/AVC, le MAD est utilisé comme mesure de complexité :

$$MAD = \frac{\sum_{i,j} |source_{i,j} - prediction_{i,j}|}{256}$$
(2.1)

On note que le MAD est calculé pour un macrobloc donné (respectivement pour une trame donnée). Il représente la somme des différences absolues des valeurs des pixels situés aux positions (i,j) du macrobloc (ou trame) source (source_{i,j}) et ceux du macrobloc (ou trame) prédit (prediction_{i,j}).

La mesure de distorsion a un profond impact sur les performances d'un codeur. Pour montrer l'inefficacité du *MAD* comme mesure de distorsion, nous l'avons comparé par exemple, à la métrique *SSIM* (Structural Similarity Index Metric) [97]. *SSIM* permet d'évaluer la qualité des images en intégrant certaines propriétés du Système Visuel Humain (SVH) adaptées à l'extraction des informations structurales de l'image. Cette mesure évalue la similarité de structure, de luminance et de contraste entre deux blocs d'images.

Sur la Figure 2.4, sont tracées les cartes de distorsion après voir appliqué sur la $36^{\text{ème}}$ trame de la séquence "Foreman" codée à 20Kbps, la SSIM et le MAD. Les dissemblances entre les images



FIGURE 2.4 – Comparaison des cartes de distorsion en utilisant (A) SSIM et (B) MAD pour la 36^{ème} trame de la séquence de Foreman codée à 20Kbps.

sont détectées localement en utilisant la SSIM et globalement en appliquant le MAD ce qui rend ce dernier moins précis.

2.1.3 Allocation de bits

L'idée de base d'une allocation de bits optimale est de développer une technique effective et quantifiable, capable d'attribuer plus de bits aux unités de base à forte complexité et moins de bits à celles à faible complexité. Le processus d'allocation de bits tel qu'introduit dans l'algorithme du contrôle de débit contribue également à l'échec de ce dernier car la répartition du budget de bits est effectuée équitablement entre les unités de base sans prise en considération de leurs complexités spatiales ou temporelles. Pour rappel, l'allocation de bits peut s'effectuer à deux niveaux : trame et macrobloc.

Au niveau trame, et sans la considération des trames B, l'Eq. 1.46 (Chapitre 1) devient :

$$\widehat{T_{i,n}} = \frac{B_{i,n}}{NP_n} \tag{2.2}$$

Cette équation suppose qu'il existe une similarité de la complexité de codage entre deux trames voisines. Cette hypothèse de stationnarité est néanmoins incertaine, vu qu'il existe des séquences vidéo comportant des changements de scènes ou des mouvements locaux. Cette division équitable de bits entre trames adjacentes ne permet pas de mettre en évidence l'importance du contenu spatial ou temporel d'une trame donnée par rapport à ses voisines. Ceci contribue à la détérioration des performances R-D de la séquence vidéo.

Au niveau macrobloc, l'Eq. 1.51 (Chapitre 1) utilise un rapport (MAD_{Ratio}) entre le MAD prédit par le modèle linéaire du $j^{\text{ème}}$ macrobloc de la $i^{\text{ème}}$ trame du $n^{\text{ème}}$ GOP ($MADP_{j,i,n}$), et la moyenne des MADs de tous les macroblocs précédents :

$$MAD_{Ratio}(j, i, n) = \frac{MADP_{j,i,n}^{2}}{\sum_{k=N_{unit}-j}^{N_{unit}-1} MADP_{k,i-1,n}^{2}}$$
(2.3)

$$\widetilde{b_{j,i,n}} = Tbr_{i,n} \times MAD_{Ratio}\left(j,i,n\right)$$
(2.4)

Pour rappel, N_{unit} représente le nombre de macroblocs par trame. $Tbr_{i,n}$ est le nombre de bits restants qui seront alloués aux macroblocs non codés dans la trame courante.

Cette mesure est supposée être plus corrélée avec le nombre de bits cible de chaque macrobloc

que le *MAD* prédit par le modèle linéaire. Elle peut être intéressante dans le cas de séquences vidéo non-typiques composées de deux ou plusieurs vidéos à changements de scène, mais reste inefficace dans le cas de séquences non ou peu mouvementées.

Dans le paragraphe suivant nous allons exposer les solutions bibliographiques proposées dans la littérature pour améliorer les performances de l'algorithme du contrôle de débit et remédier à ses insuffisances.

2.2 Avancées récentes du contrôle de débit de H.264/AVC

L'amélioration des performances R-D peut être effectuée en incluant deux types d'optimisation :

- Optimisation basée sur l'amélioration des fonctionnalités des composantes constituant le contrôleur de débit. Ces améliorations peuvent être effectuées au niveau du modèle quadratique (R-Q), au niveau du processus d'allocation de bits, au niveau de l'ajustement du paramètre de quantification ou au niveau de l'estimation de la mesure de complexité.
- Optimisation exploitant la structure hybride du codeur vidéo. Cette optimisation peut être réalisée au niveau de la détermination du mode de codage d'un macrobloc ou du débit trame.

Nous commençons par présenter l'ensemble des optimisations réalisées au sein du contrôleur de débit.

2.2.1 QPs des trames I et des unités de base P

Dans les applications vidéo, le contrôle de débit se traduit toujours par un problème d'optimisation avec contraintes. Puisque la quantité d'informations des séquences vidéo compressées est intrinsèquement variable, un buffer est placé entre le codeur vidéo et le canal de transmission afin de lisser la variation du débit et satisfaire les contraintes du débit et du délai de transmission. Le débit binaire cible pour coder une trame donnée est ajusté en se basant sur l'état du buffer et les données collectées à travers le codage des trames précédentes. Le paramètre de quantification est ensuite déterminé via une relation entre le débit et le quantificateur. Les solutions bibliographiques proposées dans la littérature cherchent à pallier les défauts du contrôle de débit de H.264/AVC cités précédemment. Elles proposent de nouveaux modèles débit-quantification (R-Q) pour déterminer le paramètre de quantification initial et de nouvelles mesures de complexité pour remplacer le MAD.

a) Calcul du QP initial de la trame I et la 1^{i} trame P

Dans la littérature, plusieurs solutions sont proposées afin d'estimer plus finement le paramètre de quantification initial [45] [46] [50] [58] [65] [70] [107] [108]. Kim et al. [50] modifient la formule de calcul du bpp (*Chapitre 1, Eq. 1.34*) en introduisant un facteur de pondération :

$$bpp_{mod} = \frac{a}{b} \times bpp \tag{2.5}$$

où a et b sont des constantes égales à 0.4 et 2 respectivement. En se basant sur bpp_{mod} , les auteurs introduisent une table de référence dérivée de tests expérimentaux intensifs menés avec plusieurs séquences vidéo. Cette table permet de choisir de manière adaptative le paramètre de quantification initial, suivant la valeur du bpp_{mod} . Malgré l'amélioration apportée par cette approche (+0.25dB en PSNR moyen), elle représente néanmoins un inconvénient qui réside dans la détermination empirique de la table de référence liant le paramètre de quantification initial et le bpp_{mod} .

Zhou et al. [107] proposent un schéma adaptatif de calcul du paramètre de quantification en utilisant des statistiques de la trame I et sa corrélation avec les trames P. Le modèle proposé exprime la relation entre le débit (R), le pas de quantification (QP_{step}) et sa variance (σ) via un modèle quadratique R-Q donné par :

$$R = a \times ln^2 \left(\frac{\sigma}{QP_{step}}\right) + b \times ln \left(\frac{\sigma}{QP_{step}}\right) + c \tag{2.6}$$

Les constantes a, b et c sont égales respectivement à 0.2346, 0.5657 et 0.6206. R est déterminé comme suit :

$$R = M \times \frac{BitRate}{Fr} \times \frac{L}{(L+M+1)}$$
(2.7)

où M est le nombre des trames P et L est un rapport entre la variance de la trame I et la moyenne des écarts-types des trames P. Le nouveau paramètre de quantification déterminé ainsi, permet d'améliorer la qualité mesurée en PSNR moyen (+1.53dB). De plus, les tests d'évaluation subjective montrent un lissage de qualité à travers les séquences traitées. Ces améliorations observées dans le cas d'un codage à bas débits, restent toutefois négligeables lorsque les séquences vidéo sont codées à hauts débits. De plus, cette approche ne peut être envisagée dans le cas des applications temps réel puisqu'il faut préalablement disposer des trames P pour calculer le rapport L.

Dans [70], le paramètre de quantification initial est déterminé en fonction de la longueur du GOP et la moyenne de la valeur absolue (MAV : Mean Average Value) des coefficients DCT (Discrete Cosinus Transform) de la trame I (F(u, v)).

$$QP_I = 0.022 \times MAV_{DCT}^{1.933}$$
$$MAV_{DCT} = \frac{1}{M \times N} \sum_{u=0,v=0}^{M-1,N-1} ABS(F(u,v))$$
(2.8)

Ces coefficients ne sont disponibles qu'après avoir codé la trame I et exécuté l'optimisation débit-distorsion ce qui n'est pas conforme au standard H.264/AVC. Cette approche proposée a permis de réduire le nombre de trames sautées (Frame Skipping) en comparaison avec le

contrôleur de débit de H.264/AVC tout en obtenant les mêmes performances de débit et du PSNR moyen.

Il est à noter que dans [70] et [107], les mesures de complexité incluses dans l'algorithme du contrôle de débit, ne sont disponibles qu'après avoir codées les trames I et P; or le processus de codage doit succédé le contrôle de débit.

Jing et al. [45] proposent un modèle R-Q basé sur une mesure du gradient de la trame I.



FIGURE 2.5 – Similarité entre le débit binaire et le gradient de la séquence "Foreman" : A-Gradient des différentes trames, B- Débit binaire des différentes trames codées en Intra.

La Figure 2.5A représente cette mesure du gradient pour les différentes trames de la séquence "Foreman". La Figure 2.5B représente les débits binaires des trames de la même séquence codées en mode Intra avec six paramètres de quantification variant entre 20 et 40. Les six courbes ont la même allure. Cette allure est similaire à celle du gradient représentée dans la Figure 2.5A. Cette similarité traduit la corrélation qui existe entre le nombre de bits et cette mesure de complexité. La Figure 2.6 établit une relation linéaire entre ces deux facteurs.

Finalement, les auteurs utilisent une relation polynomiale liant le paramètre de quantification (QP), le débit (R_k) et leur mesure du gradient de la $k^{\text{ème}}$ trame :

$$R_k (QP_I) = Grad \times F_k (QP_I)$$

$$F_k (QP_I) = a \times QP_I^2 + b \times QP_I + c$$
(2.9)

où a, b et c sont les coefficients du modèle. La mise à jour de ce modèle est effectuée comme suit :

$$R_{k+1}(QP_I) = Grad \times F_k(QP_I) \times \alpha_k, \qquad \alpha_k = \frac{R_{k,code}}{R_{k,cible}}$$
(2.10)

 $R_{k,code}$ et $R_{k,cible}$ représentent respectivement les nombres de bits réel et cible de la $k^{\text{ème}}$ trame I. Après avoir codé une trame I donnée, le modèle est automatiquement mis à jour et est utilisé



FIGURE 2.6 – Relation entre le débit binaire et le gradient de la séquence "Foreman" codée à QP=28.

pour les trames I suivantes.

La même étude est menée dans [46] et au lieu d'utiliser le modèle polynomial (*Eq. 2.9*), les auteurs proposent un modèle exponentiel liant le pas de quantification (QP_{step}) et le débit binaire :

$$R(QP_{step}) = Grad \times F(QP_{step}), \qquad F(QP_{step}) = a \times QP_{step}^{b}$$
(2.11)

où a>0 et b<0 sont les paramètres du modèle et qui dépendent du contenu de l'image. Les auteurs justifient le choix de ce modèle en se basant sur l'étude menée par kamaci et al. [47]. Ces derniers montrent, empiriquement, que la densité de Cauchy est plus adéquate que la densité Laplacienne traditionnelle dans l'estimation de la distribution des coefficients DCT. Pour les trames codées en Intra, l'estimation du débit est plus exacte lorsqu'une fonction d'entropie à base de densité de Cauchy est utilisée. Une approximation de la fonction d'entropie similaire à la fonction $F(QP_{step})$ de l'Eq. 2.11 est utilisée. Jing et al. [46] exploitent ainsi cette étude et adoptent $F(QP_{step})$ avec une combinaison du gradient afin de réaliser une bonne estimation du débit binaire des trames Intra. Les résultats de simulation montrent que le modèle R-Q s'adapte au changement du contenu des trames I tout en estimant avec précision leurs débits binaires (3.5% d'inadéquation du modèle proposé et 75% d'amélioration en comparaison avec la densité de Cauchy traditionnelle). Cette méthode ne s'applique pas à la première trame I du premier GOP qui continue toujours à utiliser le *bpp* pour déterminer son paramètre de quantification. Pour les trames I des GOPs suivants, une mise à jour des paramètres a et b est effectuée. La valeur de b est fixée empiriquement à -0.8, tandis que a est mis à jour selon :

$$a_{k+1} = \alpha \times a_k + (1 - \alpha) \times \frac{R_k}{Grad_k \times QP^b_{step,k}}$$
(2.12)

 R_k , $Grad_k$ et $QP_{step,k}$ représentent respectivement, le débit binaire réel après codage de la $k^{\text{ème}}$ trame I, son gradient et son pas de quantification. α est une constante égale à 0.5.

Zhou et al. [108] proposent une nouvelle mesure de complexité pour les vidéos couleurs. Cette

mesure est une combinaison du gradient et de la somme des histogrammes (SOH) de la composante de luminance Y et des deux composantes de chrominance U et V de la trame I. La mesure de complexité MC est déterminée comme suit :

$$MC = Grad \times SOH \tag{2.13}$$

où Grad est la mesure du gradient définie par :

$$Grad = \sum_{i=0}^{M_Y-2} \sum_{j=0}^{N_Y-2} \left(|Y_{i,j} - Y_{i,j+1}| + |Y_{i,j} - Y_{i+1,j}| \right) / (M_Y \times N_Y) + \sum_{i=0}^{M_U-2} \sum_{j=0}^{N_U-2} \left(|U_{i,j} - U_{i,j+1}| + |U_{i,j} - U_{i+1,j}| \right) / (M_U \times N_U) + \sum_{i=0}^{M_V-2} \sum_{j=0}^{N_V-2} \left(|V_{i,j} - V_{i,j+1}| + |V_{i,j} - V_{i+1,j}| \right) / (M_V \times N_V)$$
(2.14)

 $M_{Y,U,V}$ et $N_{Y,U,V}$ représent ent les dimensions des composantes Y, U et V. SOH est la somme des histogrammes définie par :

$$SOH = \sum_{l=0}^{255} (log_2(Hist_Y[l])) + (log_2(Hist_U[l])) + (log_2(Hist_V[l]))$$
(2.15)

 $Hist_Y, Hist_U$ et $Hist_V$ représentent les histogrammes de luminance (Y) et des chrominances (U, V).



FIGURE 2.7 – Relation entre le débit binaire et la mesure de complexité : Grad*SOH des séquences : A- "Stefan" codée à QP_I=20, B- "Foreman" codée à QP_I=32, et C- "Football" codée à QP_I=40

Trois séquences vidéo ("Stefan", "Foreman" et "Football") sont codées en mode Intra avec différents paramètres de quantification. Les débits binaires générés après codage de ces séquences ainsi que les mesures de complexité de chaque trame sont représentés sur la Figure 2.7. Une relation linéaire liant le débit réel et MC est observée. L'exactitude de cette modélisation est aussi spécifiée par le coefficient de corrélation R^2 , qui indique que cette approximation (Ligne droite) est compatible avec les valeurs réelles générées (Points bleus) (Puisque la valeur de R^2 comprise à priori entre 0 et 1, est proche de 1). Pour étudier la relation entre le débit binaire, la mesure de complexité et le paramètre de quantification, les mêmes expérimentations ci-dessus sont effectuées. Le modèle R-Q proposé est exponentiel :

$$R = \alpha_I \times Grad \times SOH \times e^{-\beta_I \times QP_I}$$
(2.16)

où α_I et β_I sont les coefficients du modèle. β_I est fixé empiriquement à 0.1 et α_I est mis à jour comme suit :

$$\alpha_I = \frac{R_{act}}{Grad_{act} \times SOH_{act} \times e^{-\beta_I \times QP_I_{act}}}$$
(2.17)

où R_{act} est le nombre de bits actuel après avoir codé la trame I courante, QP_I_{act} est son paramètre de quantification et $Grad_{act} \times SOH_{act}$ est sa mesure de complexité. La valeur initiale de α_I est égale à 30. Pour la première trame I, les valeurs de α_I et β_I sont fixées empiriquement indépendamment du contenu de ces trames dans les séquences vidéo. Les résultats expérimentaux montrent que l'approche proposée réduit effectivement le saut de trames tout en améliorant la qualité (+1.8dB de gain en PSNR moyen).

Lim et al. [58] exploitent la mesure du PSNR pour déterminer le paramètre de quantification initial. Le but est d'avoir une qualité consistante à travers les trames qui précèdent ou suivent la trame I. Les auteurs montrent, sur la *Figure 2.8*, qu'il existe, sauf pour les faibles valeurs de QP, une relation linéaire entre le PSNR de la trame P et celui de la trame I.



FIGURE 2.8 – Relation entre PSNR et QP de la séquence de "Foreman".

Le PSNR cible de la trame I $(PSNR_I_n)$ dans un GOP peut être déterminé en utilisant une moyenne pondérée des valeurs des PSNRs des trames P appartenant au GOP précédent :

$$PSNR_I_n = \left(\frac{1}{k}\sum_{i=1}^k PSNR_P_{i,n-1}\right) \times 0.8 + PSNR_P_{k,n-1} \times 0.2$$
(2.18)

 $PSNR_P_{i,n-1}$ est la valeur du PSNR de la $i^{\text{ème}}$ trame P dans le GOP précédent (n-1) et k représente le nombre des trames P. En utilisant la méthode des moindres carrées de l'analyse de régression linéaire, la relation linéaire entre le paramètre de quantification initial et le PSNR cible de la trame I est donnée par :

$$PSNR_I_n = \alpha_0 + \alpha_1 \times QP_I_n \tag{2.19}$$

où α_0 et α_1 sont les coefficients du modèle linéaire. Cette approche ne peut pas être appliquée sur le premier GOP de la séquence vidéo traitée. Le paramètre de quantification de la 1^{ère} trame I du 1^{er} GOP est déterminé classiquement à partir du nombre de bits par pixel (*bpp*). Cette détermination peut générer des erreurs pouvant se propager à travers les trames P compensées en mouvement à partir de la trame I. Ceci contribue à la dégradation des performances R-D vu que les valeurs des PSNRs de ces trames P sont exploitées par la suite pour déterminer le paramètre initial de la trame I du GOP suivant.

b) Calcul des QPs des autres unités de base P

L'utilisation des deux modèles linéaire et quadratique contribue également au mauvais choix du paramètre de quantification adéquat pour coder une unité de base P. Afin d'améliorer l'efficacité de l'algorithme du contrôle de débit, des changements sont proposés pour les deux modèles exploités.

Dans [45], le modèle R-Q de l'Eq. 2.10 est employé pour déterminer le paramètre de quantification d'une trame P à changements de scènes. Ce type de trame est généralement difficile à prédire à partir des trames précédentes en utilisant les processus d'estimation et de compensation de mouvement. Une séquence vidéo comportant des changements de scènes codée avec des paramètres de quantification erronés génère un fort débordement du buffer. En conséquence, le saut de trame qui devient nécessaire, affecte sévèrement la qualité subjective de la vidéo reconstruite. L'idée consiste alors, à coder les trames P à changements de scènes importantes comme des trames I en adoptant le modèle R-Q proposé. Le changement de scène est détecté dans une séquence vidéo lorsque la différence des histogrammes (DOH) de deux trames consécutives est supérieure à un seuil TH :

$$DOH_{n} = \sum_{i=0}^{q-1} |Hist_Y_{n}(i) - Hist_Y_{n-1}(i)| + \sum_{i=0}^{q-1} |Hist_U_{n}(i) - Hist_U_{n-1}(i)| + \sum_{i=0}^{q-1} |Hist_V_{n}(i) - Hist_V_{n-1}(i)|$$
(2.20)

où q est le nombre des niveaux (q=256). $Hist_{Y,U,V}_n$ et $Hist_{Y,U,V}_{n-1}$ sont les histogrammes des composantes de luminance Y et des chrominances (U, V) des trames n et n-1. Le seuil TH est déterminé empiriquement. La combinaison du modèle R-Q et la mesure DOHpermet d'éviter le débordement du buffer et de réduire le nombre de trames sautées causées par les changements de scènes tout en améliorant le PSNR moyen de +0.51dB.

Dans [58], une modélisation en ρ -domaine est adoptée pour calculer le paramètre de quantification des trames P. ρ est le pourcentage des zéros obtenu après transformation et quantification des coefficients. Les auteurs exploitent les études effectuées dans [35] [36]et [81] qui montrent qu'il existe une relation linéaire entre le débit binaire R et le facteur ρ . Ces mêmes études montrent également qu'une relation non linéaire est observée entre le facteur ρ et QP, le paramètre de quantification.



FIGURE 2.9 – Relation entre le débit binaire R et le facteur - ρ de la Séquence "Foreman"

La Figure 2.9 illustre ces résultats en montrant le lien entre le débit binaire R, le coefficient ρ et le paramètre de quantification de dix trames de la séquence "Foreman". En se basant sur ces résultats, un nouveau modèle liant le paramètre de quantification (QP_P) et ρ est proposé pour les trames P :

$$E_{QP_P_n} = \beta_0 + \beta_1 \times E_{\rho_n}, \qquad E_{QP_P_n} = e^{-\left(-\frac{QP_P_n}{12.5}\right)}, \qquad E_{\rho_n} = e^{-\rho}$$
(2.21)

où β_0 et β_1 sont les coefficients du modèle et sont déterminés par une méthode de régression linéaire. Le modèle proposé est appliqué lorsque le débit réel utilisé pour coder la trame précédente est compris entre l'intervalle des valeurs [0.9 1.1] multipliées par le débit cible de la trame courante. L'application du modèle R-Q décrit ci-dessus engendre une estimation précise du débit cible puisque le taux d'erreur relatif à la différence entre ce dernier et le débit réel reste légèrement petit. Cette estimation précise de bits, permet de maintenir la stabilité du buffer tout en réduisant sa capacité d'occupation qui peut brusquement augmenter à cause d'un changement de scènes. La qualité est également maintenue consistante à travers les trames (faible variance des valeurs du PSNR) sauf dans le cas d'un changement de scènes, ou une variation des valeurs du PSNR est observée accompagnée d'une légère amélioration du PSNR moyen.

Min et al. [65] proposent deux mesures de similarité à base de MAD spatial et temporel. Le MAD spatial est calculé comme suit :

$$MAD_{spa} = b_1 \times MAD_d + b_2 \tag{2.22}$$

 MAD_d est la différence entre l'unité de base originale et celle reconstruite. Il est déterminé uniquement lorsque le processus de compensation de mouvement est effectué en utilisant deux modes de codage uniquement (Inter16x16 et IntraA16x16) (Voir Annexe 1). b_1 et b_2 sont les coefficients du modèle. Le MAD temporel (MAD_{tem}) est déterminé par le modèle linéaire de l'Eq. 1.39 (Chapitre 1). Pour choisir le meilleur modèle de prédiction du MAD, deux mesures de similarité sont calculées à savoir :

$$\mu_{spa} = \sum_{K} \left(MAD_{spa} - MAD_{act} \right)^2, \qquad \mu_{tem} = \sum_{K} \left(MAD_{tem} - MAD_{act} \right)^2 \tag{2.23}$$

 MAD_{act} est le MAD réel et K est le nombre des unités de base. Le MAD sera prédit spatialement si la variance spatiale est supérieure à la variance temporelle et inversement. Finalement, le MAD estimé par l'un de ces deux modèles (spatial ou temporel) est utilisé ensuite par le modèle quadratique (*Chapitre 1, Eq. 1.54*). L'intégration de ces deux catégories de MAD permet d'obtenir une légère amélioration en PSNR (+0.272dB). Il est à noter que le standard H.264/AVC supporte deux types de codage Intra et Inter. Le codage Intra inclut 13 modes qui diffèrent selon la taille du bloc à coder (4x4 ou 16x16 pixels). Le codage Inter inclut 9 modes. Le processus d'optimisation débit-distorsion permet de choisir le mode de codage optimal qui peut ne pas être INTRA16x16 ou INTER16x16 choisi par l'approche proposée. Le MAD spatial basé uniquement sur l'un de ces deux modes choisis peut être erroné.

Yin et al. [103] proposent une nouvelle idée permettant d'exploiter les corrélations spatiotemporelles.



FIGURE 2.10 – position d'une unité de base (Trame ou macrobloc) m.

Sur la Figure 2.10, le MAD prédit du $m^{\text{ème}}$ macrobloc est déterminé en fonction des valeurs des MADs des macroblocs adjacents $(m_1, m_2 \text{ et } m_3)$ et de la valeur du MAD du macrobloc

situé à la même position dans la trame précédente :

$$MAD_{n}^{Pred}(m) = \frac{|MAD_{n}(m_{1}) + MAD_{n}(m_{2}) + MAD_{n}(m_{3}) + MAD_{n-1}(m)|}{4}$$
(2.24)

Le modèle quadratique est aussi substitué par celui utilisé dans l'algorithme du contrôle de débit TMN8 [13] :

$$R(m) = MAD_n^{Pred}(m) \times \left(\frac{e}{ln2} \times \frac{\alpha}{QP_{step}^2(m)}\right)$$
(2.25)

où R(m) est le nombre de bits cible pour coder le $m^{\text{ème}}$ macrobloc, e est la base de la fonction logarithmique utilisée et α est le coefficient du modèle R-Q. L'intégration de cette approche dans l'algorithme du contrôle dé débit conduit à une légère amélioration de la qualité (+0.43dB en PSNR moyen), et une très légère réduction de la variance des valeurs du PSNR. L'utilisation d'un modèle R-Q quadratique simplifié (Eq. 1.54), peut constituer une limitation de l'approche proposée car la relation de proportionnalité entre le MAD prédit et le nombre de bits cible n'est pas toujours garantie (Voir Section 2.1.1 (b)).

Dans [102], le MAD du macrobloc courant est remplacé par la variance des coefficients AC obtenus après calcul de la DCT. Cette approche reste toutefois non conforme au contexte H.264/AVC vu que les coefficients transformés ne sont disponibles qu'après avoir codé l'unité de base courante.

Dans [94] et [95], des améliorations sont apportées au modèle quadratique utilisé par le contrôleur de débit de H.264/AVC. Les auteurs montrent que la proportionnalité existante entre le nombre de bits de texture et le MAD prédit linéairement, n'est pas toujours valide. Le modèle quadratique amélioré est défini comme suit :

$$\frac{R}{MAD} = \frac{c_0}{MAD} + c_1 \times QP + c_2 \times QP^2 \tag{2.26}$$

où c_0 , c_1 et c_2 sont les coefficients du modèle. Ils sont calculés empiriquement lorsque le modèle est appliqué à la première trame I. Ils sont mis à jour ensuite, par une méthode de régression linéaire. L'adoption de ce nouveau modèle quadratique génère des débits binaires réels égaux aux débits cibles. Le gain en PSNR est amélioré de +1.09dB.

Dans [22] et [23], une modélisation R-D séparable et considérant l'ensemble des trois composantes (Y, U et V) est adoptée. Afin d'analyser l'impact des différentes composantes sur l'efficacité du codeur vidéo, la séquence vidéo couleur "Mobile" a été compressée en utilisant le codeur H.264/AVC. Les courbes R-D obtenues sont données sur la *Figure 2.11A*. Ces courbes sont différentes puisque les trois composantes possèdent différentes textures. La *Figure 2.11B* représente le débit binaire généré en fonction du numéro de la trame lorsque la même séquence est codée avec un paramètre de quantification fixe (QP=32). Bien que le débit binaire des composantes U et V soit inférieur à celui de la composante Y, les caractéristiques R-D des composantes couleurs sont différentes de celles de la composante de luminance et doivent être modélisées séparément pour accroître l'efficacité du codage.



FIGURE 2.11 – Résultats de test de la séquence couleur "Mobile" (QCIF,QP=32), A- : courbes R-D, B- : Débit binaire des composantes Y, UV

Deux modèles quadratiques sont proposés. Le premier permet de calculer le paramètre de quantification pour chacune des trois composantes :

$$R_{YUV} = \frac{a_{1,1} \times MAD + a_{1,2}}{QP} + \frac{a_{2,1} \times MAD + a_{2,2}}{QP^2}$$
(2.27)

 $a_{1,1}$, $a_{1,2}$, $a_{2,1}$ et $a_{2,2}$ sont les coefficients du modèle R-Q. Ils sont mis à jour par la méthode de régression linéaire. Le second modèle sert à représenter une nouvelle mesure de distorsion :

$$D_{YUV} = \sqrt{\frac{1}{\delta} \times \left((b_{1,1} + 0.5 \times b_{1,2} \times b_{1,2}) \times QP + (b_{2,1} + 0.5 \times b_{2,2} \times b_{1,2}) \times QP^2 \right)}$$
(2.28)

 $b_{1,1}$, $b_{1,2}$, $b_{2,1}$ et $b_{2,2}$ sont les coefficients du modèle mis à jour par une méthode de régression linéaire. δ est une constante égale à 1.5. Cette technique de modélisation séparable permet de réduire l'erreur d'estimation relative à la différence entre le débit cible et celui réel. Les auteurs proposent en plus, une approche basée sur la minimisation de la variation de la distorsion (*MINVAR*, *Chapitre 1*, *Eq. 1.7*) afin de réduire la fluctuation de distorsion entre trames adjacentes. Pour rappel, la variation V_n de la distorsion D_n de la $n^{\text{ème}}$ trame est contrôlée par la contrainte des fluctuations du débit binaire. Le critère de la variation de la distorsion dans les systèmes de codage vidéo est défini par :

$$V_n(QP_{n-i-1},\ldots,QP_{n+j}) = |D_n(QP_{n-i},\ldots,QP_{n+j}) - D_{n-1}(QP_{n-i-1},\ldots,QP_{n+j-1})| \quad (2.29)$$

Les variables i et j sont des entiers positifs représentant l'ensemble des trames précédentes et suivantes. Les paramètres de quantification optimaux estimés selon le critère MINVAR sont donnés par :

$$QP^* = (QP_1^*, \dots, QP_N^*) = \underbrace{argmin}_{(QP_1, \dots, QP_N)} \sum_{n=2}^N V_n \left(QP_{n-i-1}, \dots, QP_{n+j} \right)$$
(2.30)

N représente le nombre de trames dans la séquence. Les variations du buffer peuvent aussi être prises en considération dans ce problème de minimisation de la variation de la qualité. Ainsi, la variation minimale (*MINVAR*) peut être reformulée comme suit :

$$QP_n^* = \underbrace{argmin}_{QP_n} V_n(QP_n) \qquad sous \qquad contrainte \qquad R_n^l \le R_n(QP_n) \le R_n^u \qquad (2.31)$$

Avec R_n^l et R_n^u représentent les limites inférieure et supérieure du débit binaire cible de la $n^{\text{ème}}$ trame. $V_n(QP_n)$ est la fonction normalisée de la variation de la distorsion définie par :

$$V_n(QP_n) = \frac{|D_{YUV,n}(QP_n) - D_{YUV,n-1}(QP_{n-1})|}{D_{YUV,n-1}(QP_{n-1})}$$
(2.32)

La Figure 2.12 montre la comparaison entre l'algorithme traditionnel du contrôle de débit de H.264/AVC et l'algorithme du contrôle de débit à base de modélisation séparable et de lissage de qualité (MINVAR). Les fluctuations du PSNR (distorsion) sont larges lorsque le contrôleur de débit de H.264/AVC est appliqué alors que la qualité est plus lisse lorsque la la technique proposée est adoptée. De plus, la variation globale de la qualité (mesurée en PSNR) est réduite de 0.13dB.



FIGURE 2.12 – Comparaison entre l'algorithme du contrôle de débit de H.264 et celui utilisant la technique proposée (Séquence "Foreman" codée à 64Kbps).

2.2.2 Allocation de bits

Le but d'une allocation de bits est de maximiser la qualité globale de la vidéo, tout en minimisant/réduisant la variation de la distorsion entre trames adjacentes et en respectant les contraintes du débit binaire cible, du débit trame et de la taille du buffer du codeur/décodeur. Le problème d'allocation de bits peut être formulé comme suit. Soit $QP = \{1, 2, 3, \ldots, 51\}$, l'ensemble des valeurs du paramètre de quantification admis par le standard H.264/AVC, et soit B_s la taille maximale du buffer. Trouver un paramètre de quantification optimal $QP^* = QP_1^*, QP_2^*, \ldots, QP_N^*$ tel que :

$$QP^* = \frac{1}{N} \times \sum_{i=1}^{N} \left| d_i \left(QP_i \right) - \overline{d \left(QP \right)} \right| \qquad sous \qquad contrainte \qquad 0 \le B_i \left(QP \right) \le B_s \qquad (2.33)$$

N représente le nombre total de trames dans la séquence, $d_i(QP_i)$ est la distorsion mesurée par l'erreur quadratique moyenne de la $i^{\text{ème}}$ trame dont le paramètre de quantification est QP_i , $\overline{d(QP)}$ est la distorsion moyenne de toutes les trames dans la séquence. Une solution globale et optimale est impossible à cause de l'indisponibilité des trames futures. Néanmoins à un temps spécifique du codage, il est possible de développer des solutions sous-optimales basées sur la disponibilité des informations concernant les trames codées. Ainsi, au lieu de minimiser explicitement la variation de la distorsion à travers toutes les trames, le problème peut être adressé en explorant la relation entre le budget de bits alloué et une mesure permettant d'estimer la complexité du contenu de l'unité de base. L'allocation de bits peut être effectuée en exploitant soit :

- Les mesures classiques (Variance, gradient, contour...) qui estiment l'activité spatiale d'une unité de base . Il est à noter qu'une unité de base ayant une activité élevée et codée avec le même nombre de bits que d'autres unités, engendre plus de distorsion.
- Les mesures à base de modèles d'attention visuelle exploitant certaines propriétés du Système Visuel Humain. Ces modèles permettent de détecter les zones saillantes dans une trame donnée. L'idée de base est de favoriser ce type de régions tout en les codant avec plus de bits.

Dans ce qui suit, nous détaillons ces deux approches.

a) Allocation de bits basée sur des mesures classiques

Les études qui visent à améliorer le processus d'allocation de bits aux niveaux trame ou macrobloc, peuvent être subdivisées en deux catégories : La première regroupe les études exploitant l'état du buffer [12] [13] [26]. La deuxième catégorie est celle des études exploitant le contenu des scènes vidéo [42] [43] [44] [47] [60] [63] [70] [72] [75] [85] [97] [99] [101] [107] [108]. Dans [12] [26] et [13], le nombre de bits cible initial de chaque trame P est ajusté en fonction de l'occupation courante du buffer. Le but est de maintenir l'état du buffer constant (à 50% environ) après le codage de chaque image. Evidemment, de tels schémas sont incapables de réaliser une allocation de bits optimale à cause de l'absence de correspondance avec les caractéristiques non-stationnaires des signaux vidéo.

La seconde catégorie des approches proposées consiste à allouer le budget de bits disponible

selon des mesures de complexité de la trame considérée. Habituellement, la variance des coefficients DCT, obtenus après compensation de mouvement, est utilisée pour mesurer la complexité de la trame. Ces coefficients DCT sont supposés avoir une distribution Laplacienne ou une densité de Cauchy [47] [72]. Ribas et Lei [72] améliorent le processus d'allocation de bits aux niveaux trame et macrobloc. Par exemple au niveau trame, le nombre de bits cible est estimé par l'Eq. 2.34:

$$T_i^* = \left(T - A \times N_{mb} \times \sum_{n=1}^N C_n\right) \times \frac{S_i}{\sum_{n=1}^N S_n} + A \times N_{mb} \times C_i$$
(2.34)

 $S_i = \sum_{j=1}^{N_{mb}} \sigma_{j,i}^2$ peut être interprétée comme l'énergie de la $i^{\text{ème}}$ trame, et $\sigma_{j,i}^2$ représente la variance du $j^{\text{ème}}$ macrobloc appartenant à la $i^{\text{ème}}$ trame. T_i^* et T représentent respectivement, le nombre de bits cible optimal et le budget de bits total correspondant au GOP. N, N_{mb}, A et C_i sont le nombre de trames par GOP, le nombre de macroblocs par trame, le nombre de pixels par macrobloc et le débit moyen (en bits/pixel) utilisé pour coder les vecteurs de mouvements, les bits d'entête et autres informations relatives à la $i^{\text{ème}}$ trame. Au niveau macrobloc, le nombre de bits cible est déterminé par :

$$T_{j,i}^* = A \times \left(K \times \frac{\sigma_{j,i}^2}{Q_{step,j,i}^2} + C_i \right)$$
(2.35)

Les constantes C_i et K sont mises à jour en fonction des statistiques de la trame précédente. Elles sont égales initialement à *Bitrate/Fr* et e/ln2. Cet algorithme s'achève par une détermination du paramètre de quantification suivant une optimisation lagrangienne permettant de minimiser la distorsion sous contrainte de débit binaire. Cette approche implémentée dans les algorithmes du contrôle de débit de H.263 et MPEG-4, permet d'atteindre des débits binaires réels très proches aux débits cibles, de réduire légèrement le nombre de trames sautées et d'améliorer la qualité des séquences vidéo (+0.9dB en PSNR moyen). De plus, l'occupation du buffer est maintenue presque stable et inférieure à celle obtenue par les algorithmes du contrôle de débit de H.263 et MPEG-4.

Dans [99], un modèle liant le débit et la mesure de complexité est utilisé afin de mieux gérer le budget de bits. Cette mesure considérant à la fois les vecteurs de mouvement et l'intensité des pixels est étudiée :

$$J = mdev + \lambda * \overline{R_{motion}} \tag{2.36}$$

où *mdev* est la variance moyenne définie par :

$$mdev = \frac{1}{N_{mb}} \times \sum_{j=1}^{N_{mb}} mdev_j, \qquad mdev_j = \frac{1}{256} \times \sum_{k=1}^{256} |x_{k,j} - \overline{x_j}|$$
 (2.37)

 $x_{j,k}$ et $\overline{x_j}$ représentent respectivement, l'intensité du $k^{\text{ème}}$ pixel du $j^{\text{ème}}$ macrobloc et la moyenne des intensités des pixels du macrobloc considéré. La moyenne $\overline{R_{motion}}$ définie par $\overline{R_{motion}} = \frac{1}{N_{mb}} \times R_{motion}$, tel que R_{motion} est le nombre de bits total utilisés pour coder les vecteurs de

mouvement dans la trame, λ est le multiplicateur de Lagrange. Si $\lambda = 0$, la mesure J devient très identique à σ^2 . Cependant, les mesures telles que σ^2 ou J présentent quelques problèmes intrinsèques. L'allocation de bits à base de relations linéaires R-J ou R- σ^2 est plus adaptée aux scénarios ou chaque unité de base est codée indépendamment sans considérer les dépendances pouvant exister entre les trames (ou macrobocs) d'une séquence vidéo. Ces mesures ne différencient pas effectivement les unités de base I des unités de base P/B. Ainsi, le changement de scène n'est pas traité adéquatement, ce qui dégrade la qualité de la séquence reconstruite. Pour remédier à l'insuffisance des mesures R-J ou R- σ^2 , les auteurs incluent un nouveau modèle d'allocation de bits utilisant le MAD comme mesure de complexité :

$$T_i = \frac{Bitrate}{Fr} \times \sqrt{\frac{MAD_i}{MAD_{i-1}}}$$
(2.38)

 MAD_i et $\overline{MAD_{i-1}}$ représentent respectivement, le MAD de la $i^{\text{ème}}$ trame et la moyenne des MADs de toutes les trames codées précédemment. Le modèle non linéaire R-MAD fournit un gain de +0.56dB en PSNR moyen et réduit très légèrement le saut de trames en comparaison avec le modèle linéaire R-J.

Des mesures à base de MAD améliorées sont également utilisés, même s'elles restent insuffisantes dans plusieurs cas. [44] et [101] présentent un schéma d'allocation de bits exploitant le ratio du MAD (MAD_{ratio}). Ce dernier représente le rapport entre le MAD prédit de la trame courante et la moyenne des MADs des trames P codées précédemment.

$$MAD_{ratio}\left(n\right) = \frac{MADP_{n}}{\frac{1}{n-1} \times \sum_{l=1}^{n-1} MAD_{l}}$$
(2.39)

où $MADP_n$ est le MAD prédit par l'algorithme de Li (*Chapitre 1, Eq. 1.24*) de la $n^{\text{ème}}$ trame P. MAD_l est le MAD réel généré après codage de la $l^{\text{ème}}$ trame précédente. (n-1) est le nombre des trames codées précédemment. Les auteurs utilisent ensuite ce rapport pour pondérer le nombre de bits cibles utilisé pour coder une trame donnée (T_n) .

$$T_n = \frac{R_r}{N_r} \times SF \tag{2.40}$$

où R_r représente le nombre de bits restant pour coder l'ensemble des trames restantes N_r . SF est le facteur de pondération et est donné par :

$$SF = \begin{cases} 0.2, Si & MAD_{ratio}(n) \le 0.9, \\ 0.7 \times MAD_{ratio}(n), Si & 0.9 < MAD_{ratio}(n) \le 1.05 \\ 0.85 \times MAD_{ratio}(n), Si & 1.05 < MAD_{ratio}(n) \le 1.3 \\ 0.95 \times MAD_{ratio}(n), Si & 1.3 < MAD_{ratio}(n) \le 1.8 \\ 1.5, Si & MAD_{ratio}(n) > 1.8 \end{cases}$$
(2.41)

Cette approche proposée présente certaines limitations liées à l'utilisation de la prédiction linéaire du *MAD* qui reste toujours inefficace dans le cas de changement de scènes, à l'ajustement du nombre de bits qui réduit uniquement les variations du PSNR sans amélioration de ce dernier et au manque d'une allocation de bits optimale au niveau macrobloc également. Pour résoudre ces limitations, une nouvelle mesure de complexité nommée MAD normalisé (MAD_{NORM}) est proposée. Cette dernière est un rapport entre le MAD amélioré de la $i^{\text{ème}}$ trame $(MADEP_i)$ et la moyenne des MADs des trames codées précédemment.

$$MAD_{NORM}(i) = \frac{MADEP_i}{\frac{1}{i-1} \times \sum_{l=0}^{i-1} MAD_l}$$
(2.42)

 $MADEP_i$ est déterminé par :

$$MADEP_{i} = \begin{cases} MADP_{i} \times \sqrt{D_{Ratio}(i)}, & Si & D_{Ratio}(i) < 1.5, \\ MADP_{i} \times D_{Ratio}(i) \times 0.8, & Si & 1.5 \le D_{Ratio}(i) < 3.0, \\ MADP_{i} \times [D_{Ratio}(i) - 0.6], & Si & D_{Ratio}(i) \ge 3.0 \end{cases}$$
(2.43)

avec

$$D_{Ratio}(i) = \frac{D(i, i-1)}{D(i-1, i-2)}$$
(2.44)

D(i,i-1) est la différence de PSNR entre la trame courante et les trames précédentes. Finalement, le MAD_{NORM} est inclut dans le modèle quadratique (*Chapitre 1, Eq. 1.54*) au lieu du MAD habituel et dans l'allocation de bits au niveau macrobloc. Les tests expérimentaux menés à base de cette approche permettent de réaliser des gains en PSNR de +1.2dB et une réduction de 63% de la variance entre les valeurs du PSNR des trames codées. Ceci engendre un lissage de qualité à travers la séquence. De plus, le remplissage du buffer est maintenu stable malgré le changement de scènes.

Jiang et al. [42] [43] proposent une mesure de complexité combinant à la fois, le rapport du PSNR et le rapport du *MAD*. Cette mesure de complexité s'écrit :

$$FC_n = 0.7 \times MAD_{ratio}(n) + 0.3 \times PSNR_{drop-ratio}(n)$$
(2.45)

où $MAD_{ratio}(n)$ et $PSNR_{drop-ratio}(n)$ représentent respectivement, les rapports du MAD et du $PSNR_{drop}$ de la $n^{\text{ème}}$ trame. Supposons que la $n^{\text{ème}}$ trame est sautée et interpolée à partir de la trame précédente reconstruite (*n*-1). Le $PSNR_{drop}$ représente la différence entre le PSNR de la trame sautée et celui de la trame précédente. Le nombre de bits cible pour coder la $n^{\text{ème}}$ trame est ensuite pondéré par la mesure de complexité FC_n selon :

$$T_n = \frac{R_r}{N_r} \times \begin{cases} FC_n, Si & FC_n < 1.1, \\ (1.1 + 0.7 \times (FC_n - 1.1)), Si & 1.1 \le FC_n < 2.0 \\ 1.73, Si & FC_n \ge 2.0 \end{cases}$$
(2.46)

L'adoption de cette technique permet de réduire légèrement le nombre de trames sautées tout en maintenant stable le buffer et en évitant son débordement. La qualité mesurée en PSNR est améliorée de +0.6dB avec une variation des valeurs de PSNR similaire à celle obtenue par l'algorithme du contrôle de débit de H.264/AVC.

Dans la même voie, Lin et al. [60] proposent un rapport débit-distorsion comme mesure de
complexité. Ce dernier est une relation entre le nombre de bits (B_n) et le PSNR générés dans l'étape de pré-encodage de la $n^{\text{ème}}$ trame courante :

$$Cr_n = \frac{RD_n}{\frac{1}{n} \times \sum_{j=1}^n RD_j} \qquad RD_n = \frac{B_n}{PSNR_n}$$
(2.47)

où Cr_n est la mesure de complexité de la $n^{\text{ème}}$ trame basée sur le rapport RD. Les fluctuations du PSNR sont différentes suivant le type du mode de codage adopté par le codeur H.264/AVC. Pour prendre en considération ces fluctuations, une seconde mesure à base de PSNR est définie :

$$Cp_{n} = 1 - 0.3 \times \frac{PSNR_{p,n} - PSNR_{avg-a}}{PSNR_{avg-a}},$$

$$avec \qquad PSNR_{avg-a} = PSNR_{p,n} + \frac{1}{n-1} \times \sum_{j=1}^{n-1} PSNR_{a,j},$$

$$et \qquad PSNR_{p,n} = PSNR_{p,n-1} \times \frac{PSNR_{n}}{PSNR_{n-1}}$$

$$(2.48)$$

où $PSNR_{a,n}$ est la valeur réelle du PSNR obtenue après codage de la $n^{\text{ème}}$ trame. $PSNR_{p,n}$ est le PSNR prédit linéairement à partir du PSNR réel. Pour éviter davantage les fluctuations du PSNR, le produit des deux mesures Cr_n et Cp_n est ajusté. Finalement, le nombre de bits cible est pondéré par le produit final des deux mesures de complexité (Co_n) :

$$T_n = \frac{R_r}{N_r} \times Co_n \tag{2.49}$$

Cette technique permet d'améliorer la qualité mesurée en PSNR de +1dB tout en maintenant un lissage de qualité à travers les trames (La fluctuation du PSNR est légèrement réduite surtout lorsqu'un changement de scènes est présent dans la séquence).

Dans [63] et [85], la différence de complexité entre les trames P et B voisines est introduite dans le processus d'allocation de bits tandis que la différence entre les trames de même type est ignorée. Dans ce cas, une allocation de bits optimale ne peut pas être atteinte du fait que les trames ont des caractéristiques distinctes et donc des consommations de bits différentes.

L'histogramme de différence (HOD : Histogramm Of Difference) entre trames adjacentes est exploité comme mesure de complexité dans [45] [75] et [107]. Dans [107] par exemple, les bits sont alloués aux trames P en fonction de leur mouvement local déterminé par la mesure HOD :

$$T_n = \frac{R_r}{N_r} \times \left(1 + \frac{HOD_n - \frac{1}{n} \times \sum_{i=1}^n HOD_i}{\frac{1}{n} \times \sum_{i=1}^n HOD_i} \right)$$
(2.50)

La mesure HOD est définie selon :

$$HOD(f_m, f_n) = \frac{\sum_{i \notin [-\alpha, \alpha]} hod(i)}{N_{pix}}$$
(2.51)

où hod(i) est l'histogramme du niveau *i* relatif à la différence entre les deux trames f_m et f_n . N_{pix} et α représentent respectivement, le nombre de pixels par trame et le seuil du niveau *i*. L'algorithme global incluant l'allocation de bits à base de HOD et le modèle R-Q de l'Eq. 2.6, améliore significativement les performances R-D (+1.29dB en PSNR et débit réel est identique au débit cible). De la même manière dans [75], le nombre de bits cible est déterminé comme suit :

$$T_n = \begin{cases} T_{n-1}, & Si \quad M < 8\\ T_{n-1} \times 1.25, & Si \quad M > 20\\ T_{n-1} \times 1.05, & Autrement \end{cases}$$
(2.52)

où M est la différence entre les histogrammes de deux trames adjacentes. T_{n-1} représente le nombre de bits réel obtenu après codage de la trame précédente. A travers cette allocation de bits, les fluctuations de PSNR à travers trames sont nettement réduites. La qualité est améliorée de +0.9dB.

Il est à noter que les solutions proposées dans [45] [75] [107] et [108] utilisent des seuils empiriques afin de détecter les changements de scènes. Ces seuils peuvent générer des fausses alarmes de ces changements, d'où une répartition erronée de bits.

Les méthodes précédemment décrites souffrent de la non-intégration de certaines propriétés de la perception humaine. Inclure cette dernière dans l'algorithme du contrôle de débit est fortement recommandé vu que l'utilisateur de la séquence vidéo est un humain, et l'exploitation de son comportement visuel peut être très utile afin d'atteindre un meilleur compromis entre la qualité de la vidéo et le débit. Cette optimisation perceptuelle peut être effectuée au niveau du processus d'allocation de bits ou au niveau du réajustement du paramètre de quantification.

b) Allocation de bits basée perception humaine

Dans ce paragraphe, nous donnons quelques éléments de la vision humaine ensuite nous présentons un ensemble de méthodes perceptuelles visant à améliorer le processus d'allocation de bits.

◊ Éléments de la vision humaine

Comprendre le Système Visuel Humain (SVH) est très important dans le codage vidéo puisque les résultats de codage sont présentés à l'être humain. Par exemple, nous savons que le SVH est plus sensible à l'information de luminance qu'à celle de chrominance.



FIGURE 2.13 – Structure de base de l'œil humain.

La *Figure 2.13* présente la structure de base de l'œil humain. Ce dernier est composé de plusieurs éléments qui ont chacun une influence dans le cheminement de la lumière et la compréhension du signal optique par le cerveau.

- La cornée : Une couche transparente et résistante placée sur la face avant de l'œil. Son rôle est de protéger l'avant du globe oculaire. Sa courbure dépend des individus et varie aussi avec l'âge.
- L'humeur aqueuse : Liquide salin et alcalin sous pression qui maintient ainsi la rigidité du globe oculaire.
- L'iris : Membrane colorée qui fonctionne comme un diaphragme en contrôlant la quantité de lumière qui pénètre dans l'œil.
- La pupille qui a une apparence de disque noir, est l'ouverture centrale de l'iris. Sa taille varie pour laisser passer plus ou moins de lumière (diamètre= 2mm en pleine lumière, 8mm dans l'obscurité). La dimension maximale de la pupille est affectée par le vieillissement. Les pupilles d'une personne âgée s'ouvrent moins que celles d'un jeune (en vieillissant, on a besoin d'un éclairage plus intense pour être à l'aise).
- La fovéa est la zone de la rétine où la vision des détails est la plus précise. Elle est située dans le prolongement de l'axe optique de l'œil. La fovéa est peuplée uniquement de cônes, et nous permet la meilleure résolution optique. La fovéa ne fonctionne pas chez les personnes atteintes d'achromatopsie congénitale qui est une anomalie dans laquelle les cônes sont totalement déficients. En conséquence, les achromates ont une absence de vision centrale et une acuité visuelle réduite (<2/10), ainsi qu'une absence totale de vision des couleurs et une forte photophobie.
- Le cristallin : Lentille transparente convergente placée derrière l'iris. Il permet d'avoir une vision nette de ce que nous observons grâce à sa capacité à modifier sa courbure. La cornée et le cristallin font converger la lumière qui entre dans l'œil en la focalisant sur la rétine (membrane recouverte de photorécepteurs). La perte de transparence des cellules cristallines, entraînant leur mort, s'appelle la cataracte.
- Le corps vitré constitue les 4/5 du volume de l'œil. Il est composé d'un liquide parfaitement transparent continuellement sécrété et absorbé, dont le rôle est d'assurer la structure autonome de l'œil.
- Le point aveugle (ou "tache aveugle") : Endroit où le nerf optique vient se raccorder à la rétine. Le trou dans la vision est appelé scotome.

La lumière pénètre la pupille, elle est ciblée et inversée par la cornée et le cristallin et est projetée sur la rétine. La rétine est une membrane qui tapisse l'œil. C'est une fine couche de cellules nerveuses qui convertit le signal lumineux en un signal neuronal et le transmet au cerveau par le nerf optique. L'épaisseur de la rétine est d'environ 250μ m sauf dans la zone centrale appelée fovéa (diamètre=1.5 mm) où elle s'amincit pour atteindre une centaine de microns. La rétine contient deux types de photorécepteurs : les cônes et les bâtonnets. Les cônes sont seuls présents dans la fovéa située au centre de la macula, et leur densité diminue en allant vers la périphérie. Les bâtonnets sont répartis sur la rétine périphérique, qui donne une impression générale du champ de vision avec peu d'acuité visuelle. Les deux types de photorécepteurs de la rétine ont une fonction différente : Les bâtonnets permettent la vision à faible intensité, tandis que les cônes sont sensibles aux couleurs, sous forte lumière. Ces derniers sont en mesure de percevoir les détails les plus fins et les changements les plus rapides dans les images.

Quand on regarde une vidéo, nous nous concentrons sur des zones particulières de la scène. Notre attention est fortement dépendante de certains facteurs tels que la perception, le contraste et l'orientation [98]. Nous savons aussi que les objets en mouvement attirent l'attention humaine dans une scène. Ceci est connu sous le nom de poursuite lisse des objets en mouvement (SPEM : Smooth Pursuit Eye Movements). Ce mouvement maintient l'objet intéressant sur la fovéa [74].

◊ Bilan des techniques exploitant des modèles d'attention visuelle

Récemment, de nombreuses méthodes de codage vidéo à base de modèles d'attention visuelle ont été développées. Ces dernières peuvent être grossièrement classées en trois catégories. La première classe utilise des algorithmes de vision par ordinateur permettant de détecter automatiquement les régions intéressantes dans les scènes vidéo. Par exemple, dans [48] [52] [90] les visages humains sont considérés comme des régions d'intérêt. Les méthodes de détection et de suivi de visages sont explorées afin de garder l'œil humain focalisé sur ces zones qui vont être codées avec plus de bits afin d'améliorer leur qualité visuelle. Ce type de méthodes est très efficace dans le cas ou le visage humain est prédominant dans la scène vidéo (Vidéo conférence ou téléphone), mais échoue dans la détermination d'autres types de zones d'intérêt. La seconde catégorie d'approches perceptuelles exploite les connaissances psychophysiques du SVH pour guider le processus de codage. Il est ainsi admis que le SVH peut tolérer une certaine quantité de bruit (distorsion) en fonction de sa sensibilité à la source et au type du bruit. Les techniques proposées dans [37] [59] [66] [67] [89] [105] allouent plus de bits aux zones d'images dont la distorsion peut être facilement perceptible et moins de bits aux régions qui tolèrent plus de bruit. Par exemple, dans [89], le modèle d'attention visuelle est composé de deux sous-modèles. Le premier (appelé modèle d'attention de mouvement), permet de calculer l'intensité de mouvement d'un macrobloc à la position (i,j). Cette intensité utilise les vecteurs de mouvement (mvx, mvy):

$$I_{nij} = \sqrt{\frac{mvx_{nij}^2 + mvy_{nij}^2}{maxI_n}}$$
(2.53)

 $maxI_n$ est l'intensité de mouvement maximale de la $n^{\text{ème}}$ trame. Le second sous-modèle s'intéresse à l'analyse de la texture d'une trame donnée. Cette dernière est subdivisée en trois régions : lisses, à texture structurée et à texture aléatoire . Les régions à texture aléatoire sont constituées de petits contours à orientations variables alors que les régions à texture structurées sont composées de longs bords consistants. La vision humaine est moins sensible aux distorsions existantes dans les zones à texture aléatoire. Pour déterminer ces régions, les auteurs appliquent un filtre de détection de contours à chaque trame de la séquence vidéo. Ensuite, la moyenne des contours (ES_{nij}) est déterminée comme suit :

$$ES_{nij} = \sum_{u=0}^{15} \sum_{v=0}^{15} \frac{e_{nijuv}}{16 \times 16}$$
(2.54)

 e_{nijuv} représente l'intensité du pixel situé à la position (u,v) dans le macrobloc situé à la

position (i,j) dans la $n^{\text{ème}}$ trame. En se basant sur l'intensité des contours, une densité de distribution de ces contours est déterminé par cette équation :

$$D_{nij} = \sum_{u=0}^{15} \sum_{v=0}^{15} \frac{E_{nijuv}}{16 \times 16}$$
(2.55)

 E_{nijuv} est égale à 1 lorsque l'intensité du pixel e_{nijuv} est supérieure à 50, sinon elle est nulle. Finalement l'indicateur de texture du macrobloc situé à la position (i,j) dans la $n^{\text{ème}}$ trame (TI_{nij}) est évalué en prenant en considération ses $s \times s$ macroblocs voisins :

$$TI_{nij} = \sum_{k=i-\frac{s-1}{2}}^{i+\frac{s-1}{2}} \sum_{l=j-\frac{s-1}{2}}^{j+\frac{s-1}{2}} \frac{ES_{nkl} \times D_{nkl}}{s \times s}$$
(2.56)

Le modèle final $(VDSI_{nij})$ est une combinaison du modèle d'attention de mouvement et celui d'analyse de texture :

$$VDSI_n = \begin{cases} 255, & Si & MI_{nij} > 0.4\\ TI'_{nij}, & Autrement \end{cases}$$
(2.57)

 MI_{nij} est l'indice d'attention de mouvement du macrobloc situé à la position (i,j). Il est calculé en fonction de l'intensité de mouvement I_{nij} et il est compris entre 0 et 1. TI'_{nij} est l'indice de la carte de texture déterminé en fonction du TI_{nij} .

La Figure 2.14 illustre la carte d'importance perceptuelle de la 28^{ème} et la 107^{ème} trame de la séquence "Stefan" après avoir appliqué le modèle final. Les zones blanches correspondent aux régions mouvementées à texture lisse ou structurée. Les zones noires sont statiques à texture aléatoire. Les zones de couleur grise foncée sont les zones statiques à texture lisse et les zones mouvementées à texture aléatoire. Le gris "léger" correspond aux zones mouvementées à texture structurée.



FIGURE 2.14 – Les cartes d'importance perceptuelle de la séquence "Stefan". 28^{ème} trame : A-Original trame, B- Carte d'importance. 107^{ème} trame : C- Original trame, D- Carte d'importance.

Finalement, le facteur $VDSI_{nij}$ est employé dans le réajustement du paramètre de quanti-

fication d'un macrobloc donné :

$$QP_{nij} = QP_{anc,nij} + \left(1 - \frac{VDSI_{nij}}{255}\right) \times \Delta Q$$
(2.58)

 $QP_{anc,nij}$ est le paramètre de quantification déterminé par l'algorithme original du contrôle de débit et $\Delta Q \ge 0$; ce dernier paramètre permet de limiter la variation des paramètres de quantification des macroblocs adjacents. Cette technique réduit le débit binaire de 40.61% (Codage à haut débit) sans générer de dégradation de la qualité visuelle.

La seconde catégorie d'approche perceptuelle permet également l'intégration des facteurs perceptuels au niveau du processus d'optimisation débit-distorsion. Dans [59] [66] et [105] le multiplicateur de Lagrange utilisé par ce processus est pondéré par un facteur η généré à partir du modèle d'attention proposé.

$$\lambda_{Mode} = \lambda_{Mode} \times \eta \tag{2.59}$$

Les évaluations subjectives de la qualité montrent que les artefacts visuels peuvent être réduits grâce à cette approche. Cependant, deux problèmes peuvent exister : Premièrement, les mécanismes de la sensibilité perceptuelle humaine ne sont pas encore entièrement compris et maitrisés. Deuxièmement, la sensibilité perceptuelle ne peut pas expliquer l'attention des gens. Par exemple, les régions texturées et les objets à mouvements réguliers appartiennent souvent à l'arrière plan d'une scène vidéo et peuvent ne pas attirer l'attention humaine mais ce type de régions est sensible et doit être codé efficacement.

Enfin, la dernière catégorie des approches proposées exploite des modèles récents de neuroscience afin de prédire les régions susceptibles d'attirer l'attention humaine dans un flux vidéo. Avec le développement des sciences de la vision et du cerveau humain, des progrès ont été réalisés pour bien assimiler les mécanismes biologiques de l'attention visuelle sélective humaine. Les modèles proposés dans [40] [41] et [62], exploitent certaines caractéristiques dites bas niveau telles que l'orientation, l'intensité, et le mouvement. Ces dernières sont extraites et combinées via une combinaison non linéaire inspirées des mécanismes biologiques pour générer une carte d'attention (ou carte de saillance). Dans cette carte, les zones intéressantes sont mises en évidence. Les ressources sont ensuite allouées de manière non uniforme aux régions selon leur importance détectée via cette carte. Ceci permettra d'améliorer la qualité visuelle et respecter la contrainte de bande passante imposée par le canal de transmission. Malgré les résultats prometteurs de ces techniques, le problème d'optimisation de débit et de la qualité subjective n'est pas complètement résolu.

2.2.3 Optimisation basée sur la structure hybride du codeur

Bien que les standards de codage vidéo aient défini des architectures communes de compression pour la totalité des systèmes de codage, la flexibilité d'une structure hybride de codage vidéo permet aux chercheurs d'implanter des optimisations des débits binaires reliées à plusieurs aspects tels que : modification du mode de codage d'un macrobloc, du débit trame, de la taille du bloc

2.2.3.1 Sélection du mode de codage et de la taille des unités de codage

Le standard H.264/AVC supporte des types de codage de blocs flexibles tels que *SKIP*, *INTER16x16*, *INTER16x8*, *INTER8x16*, *INTER8x8*, *INTRA16x16* et *INTRA4x4* pour les unités de codage P (Voir Annexe 1). Le contrôle de débit et de la distorsion au sein de H.264/AVC dépend du type de l'unité de base traitée ainsi que du processus d'optimisation débit-distorsion permettant de déterminer le meilleur mode de codage suivant la formulation de Lagrange :

$$J_M(M|QP,\lambda_M) = D_M(M|QP) + \lambda_M \times R_M(M|QP)$$
(2.60)

M est l'ensemble de modes de codage testé et D_M et R_M représentent leurs distorsions et leurs débits respectifs. λ_M est le multiplicateur de Lagrange déterminé expérimentalement par :

$$\lambda_M = 0.85 \times 2^{\frac{(QP-12)}{3}} \tag{2.61}$$

Ces modes génèrent des débits binaires différents et des erreurs résiduelles différentes. La méthode d'optimisation du multiplicateur de Lagrange permet de déterminer le meilleur mode de codage d'une unité de base donnée. Elle permet également de sélectionner les trames de référence qui vont être utilisées dans le processus d'estimation de mouvement.

La sélection d'un type de la trame (I, P ou B) peut être optimisée en incorporant la méthode du multiplicateur de Lagrange afin de minimiser la distorsion sous contrainte de débit binaire [54]. La stratégie consiste à déterminer à la fois le nombre, les positions et les paramètres de quantification optimaux associés aux trames de référence dans un GOP. Le multiplicateur de Lagrange optimal est déterminé à partir des paramètres de quantification des trames de référence détectées. Cette technique permet de réduire le nombre des données R-D généré après exécution du processus d'optimisation débit-distorsion. L'implantation de cette technique est très complexe mais elle peut être adoptée comme schéma du contrôle de débit non optimal.

La technique d'optimisation débit-distorsion est d'une complexité calculatoire accrue. Pour réduire cette complexité, des approches de décisions rapides des modes de codage ont été adoptées afin de réduire le nombre de modes candidats [20] [55] [91] et [104]. Pour éviter de tester l'ensemble des modes de codage supportées par le standard H.264/AVC, on commence par tester le mode le plus simple : SKIP. Il est appliqué aux macroblocs non mouvementés. Si le mode SKIP n'est pas le mode optimal, la recherche doit être lancée dans les deux groupes des modes {16x16, 16x8, 8x16} et {8x8, 8x4, 4x8, 4x4}. Dans un premier temps, on ne compare que les coûts obtenus après application des deux modes 16x16 et 8x8 (calculés par l'Eq. 2.60). Si le coût du mode 8x8 est minimal, ceci implique que le macrobloc traité a tendance à utiliser des tailles petites, et le mode optimal se trouve dans le groupe {8x8, 8x4, 4x8, 4x4}. Sinon, le mode optimal appartient au groupe {16x16, 16x8, 8x16}. Ensuite, il faut choisir entre le mode Intra (Uniquement les modes INTRA16x16 et INTRA4x4) et Inter. Il est à noter que la complexité calculatoire des modes Intra est 5 fois plus élevée que celle des modes Inter. Une décision précoce entre ces deux modes peut réduire jusqu'à la moitié du temps de calcul.

Outre une décision flexible du mode de codage, certains algorithmes du contrôle de débit adressent le problème de décision flexible de la taille d'un bloc ou d'une trame donnée. Le codage vidéo de blocs à taille variable doit réaliser un compromis entre le débit binaire et la distorsion en choisissant la taille optimale du bloc dans une trame donnée [86].

2.2.3.2 Sélection du débit trame

Dans les applications de codage à débit binaire faible, il est recommandé de sauter quelques trames tout en codant le reste à meilleure qualité, afin de parvenir à un compromis entre les représentations spatiale et temporelle du signal vidéo. Song et al. [82] proposent une méthode optimale de saut de trame basée sur les caractéristiques spatiale et temporelle des trames. Dans un premier temps, chaque GOP est subdivisé en un ensemble de sous-GOPs dont chacun est constitué de 12 trames . Pour chaque sous-GOP, un débit trame est affecté. Pour modifier ou choisir le débit trame pour une trame donnée, les auteurs définissent deux mesures à savoir : l'histogramme de différences entre trames adjacentes (HOD) et l'activité de mouvement. Cette dernière est un rapport entre le débit binaire et le PSNR générés après le processus de codage. Le débit trame d'un sous-GOP donné est ajusté en fonction de la différence entre l'activité de mouvement et la moyenne des valeurs HODs de trames appartenant au sous-GOP précédent. Cette approche permet de préserver une qualité lisse. De plus, le saut de trames brusquement causé par les scènes mouvementées peut être réduit.

2.2.3.3 Contrôle de débit à plusieurs niveaux

Plusieurs niveaux du contrôle de débit sont supportés dans les récents standards de codage vidéo tels que MPEG-4 et H.264/AVC. Ces nouvelles approches sont étendues du niveau trame à celui du macrobloc (*Chapitre 1*). Un algorithme du contrôle de débit à plusieurs niveaux fournit une robustesse et une exactitude en ce qui concerne la détermination du paramètre de quantification. Cette exactitude est accompagnée d'une complexité calculatoire accrue en comparaison avec un contrôle de débit à un seul niveau.

He et al. [35] proposent une solution d'ajustement du paramètre de quantification au niveau macrobloc outre le contrôle de débit aux niveaux GOP et trame. Ils remarquent que le paramètre de quantification au niveau trame estimé en utilisant le débit binaire cible et le modèle débitquantification (R-Q) est un nombre réel alors que le codeur supporte uniquement des valeurs entières des paramètres de quantification. L'utilisation d'une valeur arrondie diminue la précision du contrôle de débit. Ainsi, deux solutions sont proposées. La première consiste à attribuer arbitrairement à quelques macroblocs de larges valeurs entières inférieures à la valeur réelle et à d'autres macroblocs des petites valeurs entières supérieures à la valeur réelle de telle manière que la valeur moyenne du paramètre de quantification reste égale à la valeur réelle.

Chai et al. [19] proposent un schéma de codage vidéo avant plan/arrière plan en segmentant une trame donnée en des régions d'avant et d'arrière plans qui sont codées avec différents pas de quantification. Ceci a permis d'améliorer la qualité subjective des séquences vidéo décodées. L'idée de base derrière cette technique est que, les gens sont plus intéressés par les zones appartenant en avant plan et l'amélioration de la qualité objective de ces zones participe dans l'amélioration de la qualité subjective de la scène visuelle. Dans les solutions [37] [59] [66] [67] [89] et [105] (exhibées ci-dessus), le contrôle de débit est exécuté aussi au niveau macrobloc en exploitant des modèles d'attention visuelle. De telles algorithmes ont atteint deux buts : réduire le débit global sans générer une dégradation dans la qualité visuelle, ou améliorer la qualité tout en respectant la contrainte de bande passante imposée par le canal de transmission.

2.3 Discussion et conclusion

La plupart des standards de codage vidéo ne spécifient pas le mécanisme du contrôle de débit. Son implémentation est un champ qui est laissé ouvert aux concepteurs. Le but à atteindre est de développer des solutions capables de subvenir aux contraintes imposées par le type d'applications visées.

Dans ce chapitre, nous avons présenté un ensemble de techniques visant à remédier aux défauts détectés de l'algorithme du contrôle de débit adopté par la norme H.264/AVC. Ces différentes techniques récentes peuvent être prises avec précaution et peuvent être appliquées dans différents environnements. Puisque les systèmes de codage vidéo ont différentes caractéristiques, définir un problème du contrôle de débit optimal reste toujours crucial pour chaque concepteur. Généralement, la structure d'un système de codage vidéo et le type d'application visé jouent un rôle très important dans la conception des schémas du contrôle de débit. Dans [21], de nouvelles pistes à explorer ont été brièvement présentées. Ces pistes peuvent constituées un champ fertile permettant d'améliorer d'avantage le processus du contrôle de débit :

- Codage vidéo perceptuel : Certaines recherches visent à utiliser ce type de codage afin de délivrer une qualité vidéo maximale à un débit binaire réduit tout en exploitant les propriétés du Système Visuel Humain. Une approche optimale du contrôle de débit est de maximiser une métrique de qualité visuelle de telle façon qu'on essaye de déterminer les meilleurs paramètres du codage à un débit binaire cible fixe. Puisque les mesures perceptuelles sont très complexes, un contrôle de débit optimal est nécessaire pour équilibrer la différence entre la complexité de l'algorithme et la performance R-D.
- Codage conjoint source-canal : Sujet aux contraintes des systèmes de communication, le contrôle de débit en codage source-canal sert à trouver la distribution optimale de bits entre le codage source et le codage canal afin de réaliser une distorsion minimale. Dans un environnement à débit binaire variable (VBR), le contrôle de débit source-canal devrait être capable de réaliser à la fois une qualité vidéo consistante au niveau du codeur et un lissage du train binaire au niveau canal. En ce qui concerne l'environnement à débit binaire constant (CBR), le contrôle de débit optimise la qualité de la vidéo transmise en sélectionnant jointement le débit de la source et celui du canal.
- **Transcodage** : Les applications vidéo numériques bénéficient de la standardisation du format du train binaire et des méthodes de codage. Le transcodage vidéo est un processus de conversion d'une vidéo compressée à un certain débit à une autre vidéo avec un débit différent. Le contrôle de débit est l'une des fonctions clés dans un transcodeur permettant une adaptation du débit et une concaténation des trains binaires, ceci influence directement la dégradation de la qualité vidéo introduite par le processus de codage.

Dans ce chapitre nous avons étalé les insuffisances de l'algorithme du contrôle de débit adopté par H.264/AVC, ensuite, nous avons élaboré un examen de certaines solutions bibliographiques permettant de remédier à ses défauts et d'améliorer ses performances. Nous avons également exhibé brièvement, quelques domaines de recherche pouvant être de nouvelles perspectives et un champ fertile pour davantage d'études sur le contrôle de débit et de la qualité, puisque ce dernier aura toujours une importance majeure dans les applications de codage vidéo. Dans le chapitre suivant, nous allons présenter une première contribution majeure dans cette thèse. Il s'agit d'une nouvelle technique de modélisation débit-quantification (R-Q) permettant de déterminer les paramètres de quantification des trames I et des unités de base P. Cette proposition vise à corriger les défauts détectés dans le contrôle de débit du standard H.264/AVC et évite certaines insuffisances de certaines techniques proposées précédemment dans la littérature.



Dans ce chapitre, nous proposons une approche de modélisation débit-quantification (R-Q, de l'anglais Rate-Quantization). Cette dernière surpasse certaines méthodes proposées dans la littérature, et permet de remédier les insuffisances liées à l'estimation des paramètres de quantification des trames I et des unités de base P (Trames ou Macroblocs). Nous commençons par donner les détails d'implémentation des deux modèles Intra et Inter R-Q. Nous discuterons par la suite, les résultats expérimentaux en comparant notre approche avec le codeur de référence JM15.0.

3.1 Modèles débit-quantification (R-Q)

Dans cette section, nous allons décrire les détails d'élaboration des deux modèles débitquantification (R-Q) utilisés pour calculer les paramètres de quantification des unités de base Intra et Inter.

3.1.1 Modèle R-Q de la trame I et la première trame P

Dans les standards de codage vidéo, chaque groupe d'images est composé au moins d'une seule trame I suivie d'autres trames P et/ou B. Ces dernières sont prédites, directement ou indirectement, par compensation de mouvement à partir de la trame I. Améliorer la qualité de codage de la première trame permettra sans aucun doute d'améliorer celle des trames suivantes, et augmentera par conséquent les performances R-D de la totalité de la séquence. Dans cette optique, nous proposons un nouveau modèle Intra R-Q permettant d'optimiser la sélection du paramètre de quantification initial pour coder la trame I et la première trame P. Pour mettre en œuvre ce modèle, nous avons commencé par subdiviser les séquences vidéo manipulées en deux bases : Une première base nommée "*BaseTr*" utilisée dans l'élaboration du modèle Intra proposé. Cette base est composée de quatre séquences vidéo "Carphone", "Coastguard", "Foreman" et "Suzie". La seconde base appelée "*BaseTest*" est employée pour tester l'efficacité du modèle. Elle est constituée de onze séquences vidéo "Akiyo", "Bridge-Close", "Claire", "Container", "Grandma", "Highway", "Miss-America", "Mother-Daughter", "News", "Salesman" et "Silent". La description du contenu de ces séquences est détaillée dans la section ci-dessous décrivant par la même occasion l'environnement expérimental.

L'élaboration du modèle Intra R-Q est subdivisée en deux principales étapes :

- 1) Modéliser la relation entre le débit binaire fixé pour coder la totalité de la séquence (RT) et le débit binaire cible pour coder la trame I (RI).
- 2) Modéliser la relation entre le débit binaire cible pour coder la trame I (RI) et le paramètre de quantification (QP_I) .

La combinaison des deux relations permettra de déterminer le paramètre de quantification initial. Pour atteindre l'objectif de la 1^{ere} étape, nous avons mené des tests expérimentaux sur les séquences appartenant à *BaseTr*. Ces derniers consistent à coder chaque séquence vidéo à plusieurs débits binaires variant de 10Kbps à 80Kbps. Une collecte des informations suivantes est effectuée :

- Débit binaire réel généré après codage de la trame I,
- La mesure de complexité δ (Moyenne quadratique des gradients) déterminée selon :

$$\delta = \sqrt{\left(\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |Y_{i,j} - Y_{i+1,j}|\right)^2 + \left(\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} |Y_{i,j} - Y_{i,j+1}|\right)^2}$$
(3.1)

M et N sont les dimensions de la trame. $Y_{i,j}$ est la valeur du pixel de luminance à la position (i,j).

Dans un premier temps, nous avons schématisé la relation liant le rapport RI/δ et RT/Fr(Avec Fr est la fréquence trame) (comme le montre la Figure 3.1).



FIGURE 3.1 – Relation entre le rapport (RI/δ) et le débit binaire total RT des séquences vidéo "Carphone", "Coastguard", "Foreman" et "Suzie".

D'après la *Figure 3.1*, nous remarquons que la relation entre le rapport (RI/δ) et le débit global RT est non linéaire et peut être approximée par cette équation :

$$\frac{RI}{\delta} = a * \left(\frac{RT}{Fr}\right)^b \tag{3.2}$$

Les paramètres du modèle non linéaire a et b sont déterminés par la méthode des moindres carrées afin d'approximer les données obtenues précédemment. Leurs valeurs sont égales respectivement à 123.4 et 0.38.

La seconde étape de notre approche de modélisation R-Q, consiste en la génération d'un modèle liant le rapport (RI/δ) et le paramètre (QP_I) . Pour atteindre cet objectif, nous avons codé les mêmes séquences vidéo avec les mêmes débits mentionnés précédemment tout en variant le paramètre de quantification de 10 à 40 à un pas de 2.



FIGURE 3.2 – Relation entre le rapport (δ/RI) et QP_I des séquences vidéo "Carphone", "Coastguard" et "Foreman".

Dans la Figure 3.2, nous avons tracé la variation du rapport (δ/RI) en fonction du paramètre de quantification initial (QP_I) pour trois séquences vidéo : "Carphone", "Coastguard" et "Foreman". Nous remarquons que la relation est non linéaire et nous avons choisi de la modéliser :

$$\frac{\delta}{RI} = p_1 \times QP_I^2 + p_2 \times QP_I + p_3$$

$$\frac{RI}{\delta} = \frac{1}{p_1 \times QP_I^2 + p_2 \times QP_I + p_3}$$
(3.3)

où p_1 , p_2 et p_3 représentent les coefficients du modèle R-Q. Ils sont égaux respectivement à 2.443×10^{-6} , -6.934×10^{-5} et 6.68×10^{-4} .

La combinaison de l'Eq. 3.2 et l'Eq. 3.3 permet de calculer le paramètre de quantification initial optimal à partir du débit binaire global fixé au départ pour coder la totalité de la séquence vidéo. La particularité de notre approche de modélisation R-Q réside dans le fait que nous tenons compte de la complexité de la trame I représentée par δ , sans effectuer le calcul de cette dernière; contrairement aux travaux proposés dans la littérature qui effectuent un calcul de la mesure de complexité [45] [58] [70] et [107]. De plus, les tests expérimentaux menés sur les séquences vidéo appartenant à la base : BaseTest prouvent l'efficacité de cette nouvelle approche (Nous reviendrons sur cette partie au Paragraphe 3.2.4).

Pour améliorer davantage les performances R-D des séquences vidéo, nous allons proposer par la suite une seconde approche de modélisation R-Q permettant de déterminer les paramètres de quantification (QPs) des unités de codage P.

3.1.2 Modèle R-Q des unités de codage P

L'objectif d'une modélisation à base de paramètre de quantification est de développer une solution débit-quantification capable de fournir une qualité consistante à travers la totalité des séquences vidéo et quelles que soient leurs caractéristiques spatiales et temporelles.

Dans cette section, nous proposons un modèle débit-quantification capable d'estimer les paramètres de quantification des unités de codage P. Les tests expérimentaux ont montré que l'utilisation d'un paramètre de quantification constant pour coder la totalité d'une séquence vidéo, engendre de bonnes performances en termes de consistance de qualité. Ainsi, nous avons codé l'ensemble des séquences vidéo appartenant à la base BaseTr, avec un paramètre de quantification constant variant de 10 à 40 avec un pas de 2. Chaque séquence est constituée d'une seule trame I suivie de plusieurs trames P. Nous avons ensuite schématisé la relation liant les valeurs du paramètre de quantification et le nombre de bits/Trame, générés après codage de certaines trames P, pour deux séquences vidéo : "Carphone" (*Figure 3.3A*) et "Foreman" (*Figure 3.3B*).



FIGURE 3.3 – Courbes Débit-Quantification de certaines trames P des séquences vidéo -A- "Carphone" et -B- "Foreman".

Nous remarquons que cette relation peut être approximée par un modèle logarithmique R-Q :

$$QP_P = \gamma \times Log(RP) + \beta \tag{3.4}$$

 QP_P et RP représentent respectivement le paramètre de quantification et le débit binaire cible d'une unité de base P. γ et β sont les paramètres du nouveau modèle Inter R-Q. Ces derniers sont définis en utilisant une méthode de régression linéaire qui exploite les statistiques des unités de base Inter précédemment codées. La mise à jour des paramètres du modèle logarithmique peut être résumée par ces trois étapes :

 \diamond Fenêtrage : Sélection des trames précédentes contribuant à la constitution des données historiques. Ces données servent à la mise à jour des coefficients du modèle logarithmique. Cette première étape est utilisée de manière adaptative afin de lisser le changement de scènes. Soit W le nombre de trames précédentes décidé par l'algorithme suivant :

$$W = \begin{cases} \left[\frac{MAD_{t}}{MAD_{t-1}} \times MAX_SW\right], MAD_{t-1} > MAD_{t} \\ \left[\frac{MAD_{t-1}}{MAD_{t}} \times MAX_SW\right], Autrement \end{cases}$$
(3.5)

Avec t est l'instant du codage et MAX_SW représente la taille maximale de la fenêtre glissante et vaut 20. A partir de l'Eq. 3.5, W pourrait augmenter considérablement une fois qu'un changement de scène a eu lieu dans la séquence vidéo, ce qui signifie que les points de données avant le changement de scènes seront toujours pris en considération pour les trames présentes et futures.

 \diamond Élimination statistique des données aberrantes : Les paramètres du nouveau modèle sont calculés en utilisant la méthode des moindres carrées. Ce modèle est en outre calibré en rejetant les données aberrantes. Le rejet est décidé lorsque l'erreur entre le budget de bits prédit par le modèle et le nombre de bits réel utilisé par la trame, est supérieure à l'écart type à travers le nombre W des trames précédentes prises en considération.

$$std = \sqrt{\frac{1}{W} \sum_{i=0}^{W} (\gamma \times \log (RP_i) + \beta - QP_P_i)^2}$$

$$err(i) = \gamma \times \log (RP_i) + \beta - QP_P_i$$
(3.6)

Avec QP_P_i et RP_i représentent respectivement, le paramètre de quantification et le nombre de bits réel de la $i^{\text{ème}}$ unité de base. Si err(i) > std, la $i^{\text{ème}}$ unité de base va être rejetée, c.à.d. que son paramètre de quantification et son nombre de bits réel ne vont pas être inclus dans l'estimation des deux paramètres γ et β par la méthode des moindres carrées.

◊ Mise à jour des paramètres du modèle : En utilisant une technique de régression linéaire, les deux paramètres du modèle peuvent être obtenus comme suit :

$$\gamma = \frac{n \times \sum_{i=1}^{n} QP_P_i \times \log\left(RP_i\right) - \sum_{i=1}^{n} Log\left(RP_i\right) \times \sum_{i=1}^{n} QP_P_i}{n \times \sum_{i=1}^{n} \left(Log\left(RP_i\right)\right)^2 - \left(\sum_{i=1}^{n} Log\left(RP_i\right)\right)^2}$$
(3.7)

$$\beta = \frac{\sum_{i=1}^{n} QP_P_i \times \sum_{i=1}^{n} (Log(RP_i))^2 - \sum_{i=1}^{n} Log(RP_i) \times \sum_{i=1}^{n} QP_P_i \times Log(RP_i)}{n \times \sum_{i=1}^{n} (Log(RP_i))^2 - (\sum_{i=1}^{n} Log(RP_i))^2}$$
(3.8)

où n est le nombre de trames précédentes sélectionnées. Le modèle logarithmique R-Q proposé, n'utilise aucune mesure de complexité. Ceci constitue un avantage majeur en comparaison avec l'algorithme adopté par H.264/AVC et qui utilise le MAD comme mesure de complexité, et combine deux modèles : linéaire de prédiction de MAD et quadratique pour calculer les paramètres de quantification des unités de base P.

3.2 Analyse de la performance des modèles R-Q proposés

Avant de discuter les différents résultats expérimentaux obtenus, nous allons présenter les critères d'évaluation sur lesquels nous nous sommes basés pour évaluer la performance de l'approche proposée.

3.2.1 Critères d'évaluation

Parmi les critères d'évaluation adoptés par la communauté du codage vidéo et utilisés pour mettre en évidence les performances d'une approche proposée, nous citons la mesure du PSNR, la mesure de la complexité de calcul et la représentation des courbes débit-distorsion (R-D) [18].

3.2.1.1 Mesure de la qualité

Dans de nombreuses applications multimédia ou industrielles, il est nécessaire de juger de la qualité des images/vidéos traitées ou produites et en particulier de maîtriser la qualité d'une image/vidéo par rapport à celle initiale dans une chaîne de traitement. La compression vidéo affecte de plus en plus de domaines tels que la production, la post-production, la diffusion ou bien encore la transmission. Devant cette croissance effrénée, plusieurs questions se posent et notamment sur la façon d'évaluer la qualité des images. Il existe actuellement trois grandes familles de méthodes :

- Les méthodes dites subjectives utilisant un groupe d'observateurs pour qualifier la qualité des images,
- Les méthodes objectives qui utilisent les propriétés statistiques des signaux,
- Les méthodes basées sur le Système Visuel Humain qui tentent de regrouper les avantages des deux précédentes méthodes.

Il est commun, dans la communauté du codage d'images et de vidéos, de parler des pertes d'informations en termes d'Erreur Quadratique Moyenne (MSE). Cette dernière est une simple différence entre l'image originale et celle reconstruite :

$$MSE = \frac{1}{N} \times \sum_{i=1}^{N} (x_i - y_i)^2$$
(3.9)

 N, x_i et y_i représentent respectivement, le nombre d'échantillons, le signal original et celui reconstruit.

Une autre grandeur plus significative est dérivée à partir du MSE. Elle est nommée rapport signal à bruit crête (PSNR). Cette métrique représente un gain qui s'exprime en décibels, et s'obtient par l'équation suivante :

$$PSNR = 10 \times \log_{10} \left(\frac{255^2}{MSE}\right) \tag{3.10}$$

Le PSNR indique le degré de ressemblance entre l'image d'origine et l'image codée reconstruite. Étant calculé généralement sur toute l'image, il est représentatif de la qualité moyenne mais ne rend pas compte des défauts locaux. Le PSNR est souvent présenté sous forme de courbe d'évolution en fonction du débit comme décrit dans la section suivante. Autres mesures à base de certaines propriétés du Système Visuel Humain vont être définies et exploitées dans les *Chapitres* 4 et 5.

3.2.1.2 Courbe Débit-Distorsion (R-D)

Une valeur de PSNR donnée seule ne peut en aucun cas permettre d'analyser une situation de codage puisque celle-ci dépend d'un autre paramètre clé qui est le débit. Ainsi, les courbes débit-distorsion ont été adoptées afin d'apprécier le compromis entre le gain de la compression et la qualité de restitution. Un exemple de ce compromis est présenté sur la courbe de la *Figure 3.4*.



FIGURE 3.4 – Allure typique d'une courbe débit-distorsion (R-D)

Le but de l'élaboration de telles courbes étant, en général, d'atteindre la meilleure fidélité (ou la plus grande distorsion) compte tenu de la capacité du canal de transmission, qui détermine la contrainte de débit. Cette optimisation peut être faite à l'aide des techniques d'optimisation de Lagrange fondées sur la théorie du débit-distorsion.

3.2.1.3 Mesure de la complexité de calcul, du débit et du buffer

La complexité de calcul est un autre paramètre de performance important à évaluer. On se contente d'utiliser différentes méthodes pour calculer la complexité, en dépit de la situation expérimentale : le temps de codage, le débit généré et l'occupation du buffer. Ces critères sont les plus utilisés pour calculer la complexité des codeurs logiciels. Ainsi, les différences de temps de calcul, des débits et de l'occupation du buffer sont calculées entre le schéma original et celui modifié.

3.2.2 Environnement expérimental

Dans ce paragraphe, nous allons décrire la plateforme utilisée pour les différents tests expérimentaux élaborés. Ensuite nous allons présenter succinctement l'ensemble des séquences employées pour prouver l'efficacité de l'approche proposée. Finalement, nous présentons les logiciels utilisés dans notre implémentation et pour l'ensemble de comparaisons effectuées.

3.2.2.1 Plateforme

Les expériences décrites dans cette partie, ont été effectuées sur deux postes :

- PC à un processeur Intel (R) Pentium 4 de vitesse 1.61 GHz, 1Go de RAM, sous le système Microsoft Windows XP professionnel version 2.
- PC à un processeur Intel (R) Pentium (R) de vitesse 2.16 GHz, 2.00 Go de RAM, sous le système : Microsoft Windows Vista Edition Familiale Basique.

3.2.2.2 Séquences testées

Les séquences vidéo à tester sont choisies à partir des documents vidéo de l'IUT qui sont largement adoptés par les chercheurs et les scientifiques dans le but de couvrir une large gamme contenant le détail, le mouvement des objets et les différents types de fond, ainsi que le mouvement de la caméra. Le format utilisé des séquences d'images est QCIF. Les séquences manipulées sont les suivantes :



FIGURE 3.5 – 31^{ème} Trames des séquences vidéo : A-"Akiyo", B-"Bridge-Close", C- "Carphone", D-"Claire", E-"Coastguard", F-"Container", G-"Foreman", H-"Grandma", I-"Highway", J-"Miss-America", K-"Mother-Daughter", L-"News", M-"Salesman", N-"Silent", O-"Suzie".

3.2. ANALYSE DE LA PERFORMANCE DES MODÈLES R-Q PROPOSÉS

- "Akiyo", "Claire", "Miss-America" et "Grandma" (*Figure 3.5A*, *3.5D*, *3.5J*, *3.5H*) : Quatre séquences vidéo avec peu de mouvements et de détails. Chacune d'elle représente une dame assise devant un arrière plan figé à couleur plus ou moins unique, et parlant devant une caméra avec un minimum de mouvement corporel focalisé sur la tête et les épaules.
- "Brigde Close" (*Figure 3.5B*) : Cette séquence est à mouvement modéré. L'arrière-plan est figé et constitué d'un ensemble de détails. L'ensemble constitue une vue prise de loin par une caméra.
- "Carphone" (*Figure 3.5C*) : Cette séquence comporte beaucoup de mouvements pouvant être résumés au niveau de la tête, le corps et les expressions faciales exagérées. En outre, l'image subit les perturbations de prise de vues causées par le porteur de la caméra.
- "Coastguard" (*Figure 3.5E*) : Cette séquence représente beaucoup de mouvements et de détails spatiaux. Son arrière plan est variant.
- "Container" (*Figure 3.5F*) : Cette séquence est une vue prise de loin de deux bateaux ayant un mouvement modéré.
- "Foreman" (*Figure 3.5G*) : Cette séquence est classée parmi les séquences mouvementées. Elle représente un travailleur en génie civil parlant devant une caméra, avec des expressions faciales animées et des gestes de main. La caméra se déplace du travailleur vers le site des travaux et vice versa, ce qui résulte un mouvement significatif de la globalité de la séquence.
- "Highway" (*Figure 3.51*) : C'est une séquence mouvementée, subissant des perturbations liées à la vue prise via la caméra.
- "Mother-Daughter" (*Figure* 3.5K) : Séquence contenant beaucoup de détails spatiaux et peu de mouvements.
- "News" (*Figure 3.5L*) : C'est une séquence à mouvement modéré, très riche en détails spatiaux et ayant un arrière plan variant.
- "Salesman" (*Figure 3.5M*) : C'est une séquence moyennement mouvementée, très riche en détails spatiaux et ayant un arrière plan figé.
- "Silent" (Figure 3.5N) : C'est une séquence mouvementée ayant un arrière plan figé.
- "Suzie" (*Figure 3.50*) : Séquence mouvementée ayant peu de détails spatiaux et un arrière plan figé.

La base de test utilisée reflète une forte variabilité du contenu : Des séquences avec peu ou beaucoup de détails, à mouvements importants ou modérés, avec des arrière-plans figés ou variants. La performance d'une méthode proposée peut être pilotée par l'utilisation de ces séquences de test.

3.2.3 Logiciel utilisé

Dans le cadre de la norme H.264/AVC, un logiciel de référence a été développé afin d'expérimenter et de valider ses fonctionnalités. Ce logiciel se nomme JM (Joint Model), dont les sources sont librement téléchargeables sur la page Web des logiciels JVT [1]. Ce logiciel est fonctionnel, mais a reçu de nombreuses critiques concernant la lisibilité et la réutilisabilité de ses codes sources. Le manque de lisibilité du code vient du fait qu'il soit entièrement écrit en Langage C qui est orienté efficacité (peu de fonctions, réutilisation de variables, ...), et que la plupart des données utilisées sont stockées dans des variables globales. De plus, aucune documentation n'est fournie sur la structuration du modèle et sur la signification des variables utilisées, ce qui complique sa compréhension. L'encodeur fonctionne avec un fichier de configuration qui spécifie les options à utiliser, à savoir, "configfile.c". En plus de ce fichier de configuration, des options peuvent être passées en ligne de commande à l'encodeur.

Pour nos tests expérimentaux, nous avons utilisé l'une des dernières versions du logiciel de référence qui est JM15.0 [1]. Nous comparons nos approches proposées avec cette dernière ainsi qu'un algorithme du contrôle de débit amélioré qui est JVT-O016 [1] [106]. Les propositions faites dans ce travail nécessitent des modifications du code source, en particulier des sous-routines concernées par la partie modifiée. Nous avons donc commencé par analyser la structure du code, et en dégager les principales fonctions et structures. Ensuite, nous avons implémenté les approches proposées afin de comparer leurs performances avec la version du JM15.0 et l'algorithme du JVT-O016.

3.2.4 Résultats de simulations

Cette section présente les résultats relatifs à une analyse comparative de la performance de nos modèles R-Q proposés avec le logiciel de référence JM15.0. Les approches proposées ont été implémentées en utilisant la version JM10.2 [1] du logiciel de référence de H .264/AVC. Les séquences testées sont au format QCIF, avec un débit trame égal à 30fps. Les débits binaires ciblés sont bas et varient de 15 à 64Kbps. Le *Tableau 3.1* résume les conditions de test adoptées dans nos simulations.

TABLE $3.1 -$	Conditions de test.
Paramètres	Caractéristiques
-Profil utilisé	-Baseline
-Optimisation R-D	-Activée
-Nombre d'images	5
de référence	
-Saut de trames	-Désactivé
-Hadamard	-Activé
-Intra Période	0 ou 30
-Contrôle de débit	-Activé
-QP initial	0
-Type de canal	-CBR
-Structure GOP	-IPPPPou
	IPPPIPPPIPPP
-Nombre d'images	90

Le contrôle de débit dans le logiciel de référence JM15.0 [1] utilisé pour comparaison, a été exécuté selon plusieurs modes :

- Mode 0 : L'algorithme du contrôle de débit original (Présenté dans le *Chapitre I*);
- Mode 1 : L'algorithme du contrôle de débit est appliqué à toutes les unités de codage indépendamment de leurs types (I, P ou B);

- Mode 2 : Application du contrôle de débit original tout en effectuant une sélection intelligente du paramètre de quantification des unités de codage I et B;
- Mode 3 : Application du contrôle de débit original tout en employant des statistiques de débit binaire pour les unités de codage I et B.

Dans nos simulations, nous avons choisi le mode 2 puisqu'il permet de sélectionner intelligemment le paramètre de quantification des trames Intra. La comparaison avec le logiciel de référence JM15.0 est effectuée en calculant trois indicateurs :

• Différence entre les moyennes du *PSNR* par séquence (Entre notre approche : $\overline{PSNR_{Notre}}$ et l'algorithme du JM15.0 : $\overline{PSNR_{JM15.0}}$) :

$$\Delta PSNR = \overline{PSNR_{Notre}} - \overline{PSNR_{JM15.0}}$$
(3.11)

• Différence entre les moyennes du débit binaire généré suite au codage de chaque séquence (Entre notre approche : $\overline{R_{Notre}}$ et l'algorithme du JM15.0 : $\overline{R_{JM15.0}}$) :

$$\Delta R = \overline{R_{JM15.0}} - \overline{R_{Notre}} \tag{3.12}$$

• Différence entre les moyennes du temps généré suite au codage de chaque séquence (Entre notre approche : $\overline{T_{Notre}}$ et l'algorithme du JM15.0 : $\overline{T_{JM15.0}}$) :

$$\Delta T = \overline{T_{JM15.0}} - \overline{T_{Notre}} \tag{3.13}$$

Nous commençons par tester l'efficacité du modèle Intra R-Q utilisé pour déterminer le paramètre de quantification optimal des deux premières trames du GOP (I et P). Ensuite, nous testons l'efficacité du modèle Inter R-Q utilisé pour estimer le QP des unités de base P. Finalement, les deux modèles R-Q vont être combinés et mis en œuvre pour pouvoir prouver leur performance globale.

3.2.4.1 Performance du modèle Intra R-Q

Pour prouver la performance du modèle Intra R-Q proposé, nous l'avons testé sur des séquences vidéo appartenant à la base nommée *BaseTest*. Le Tableau 3.2 illustre les résultats de simulation obtenus après avoir appliqué le modèle Intra R-Q proposé en comparaison avec le logiciel de référence JM15.0. Notons que dans le tableau, une valeur moyenne positive signifie une augmentation et une valeur négative signifie une diminution.

TABLE 3.2 – Comparaison entre le modèle Intra R-Q et JM15.0 en termes du QP initial, des gains en PSNR et en débit binaire.

Séquences	Débit	JM15.0		Mo	dèle Ir	itra	Gain en	Gain en	
	Binaire	Débit	QP	PSNR	Débit	QP	PSNR	PSNR	$\mathbf{D}\mathbf{\acute{e}bit}$
	Cible	Réel	Initial		Réel	Initial		$(\Delta PSNR)$	(ΔR)
	(Kbps)	(Kbps)		(dB)	(Kbps)		(dB)	(dB)	(Kbps)
Akiyo	15	15.14	35	34.14	15.14	29	35.45	+1.31	0
	25	25.18	35	35.77	25.19	27	37.82	+2.05	-0.01
	35	35.22	35	37.22	35.23	26	39.12	+1.90	-0.01
	45	45.37	35	38.40	45.37	25	40.21	+1.81	0
	64	64.38	35	40.08	64.35	24	41.63	+1.55	+0.03
Bridge-Close	15	15.15	35	30.58	15.15	29	31.76	+1.18	0
	25	25.17	35	31.57	25.22	27	33.03	+1.46	-0.05
	35	35.32	35	32.40	35.20	26	33.78	+1.38	+0.12
	45	45.39	35	33.05	45.29	25	34.39	+1.34	+0.1
	64	64.80	35	33.97	64.60	24	35.21	+1.24	+0.2
Container	15	15.13	35	32.10	15.14	29	33.22	+1.12	-0.01
	25	25.14	35	33.90	25.17	27	35.49	+1.59	-0.03
	35	35.15	35	35.15	35.17	26	36.67	+1.52	-0.02
	45	45.19	35	36.09	45.19	25	37.48	+1.39	0
	64	64.26	35	37.46	64.15	24	38.38	+0.92	+0.11
Grandma	15	15.14	35	32.69	15.13	29	33.90	+1.21	+0.01
	25	25.18	35	34.16	25.16	27	35.75	+1.59	+0.03
	35	35.21	35	35.13	35.20	26	37.00	+1.87	+0.01
	45	45.21	35	36.09	45.38	25	37.90	+1.81	-0.17
	64	64.29	35	37.50	64.35	24	39.11	+1.61	-0.06
News	15	15.89	35	29.61	15.33	29	29.61	0	+0.56
	25	25.23	35	31.20	25.18	27	32.15	+0.95	+0.05
	35	35.31	35	32.16	35.28	26	34.03	+1.87	+0.03
	45	45.38	35	33.21	45.39	25	35.30	+2.09	-0.01
	64	64.37	35	35.30	64.58	24	37.15	+1.85	-0.21
Salesman	15	15.12	35	28.84	15.13	29	28.84	0	-0.01
	25	25.15	35	30.43	25.16	27	31.31	+0.88	-0.01
	35	35.20	35	31.63	35.18	26	33.04	+1.41	+0.02
	45	45.35	35	32.60	45.20	25	34.39	+1.79	+0.15
	64	64.37	35	34.27	64.22	24	36.21	+1.94	+0.15
\mathbf{Silent}	15	16.46	35	29.18	15.14	29	28.48	-0.70	+1.32
	25	25.18	35	30.51	25.19	27	31.04	+0.53	-0.01
	35	35.23	35	31.63	35.20	26	32.70	+1.07	+0.03
	45	45.29	35	32.50	45.33	25	33.96	+1.46	-0.04
	64	64.40	35	33.93	64.33	24	35.74	+1.81	+0.07

En comparaison avec le logiciel de référence JM15.0 qui fixe toujours le QP initial à une valeur élevée pour les faibles débits, le modèle proposé conduit à différentes valeurs de QP initial

en fonction du débit binaire fixé pour coder la totalité de la séquence. On peut noter également, que pour une précision du contrôle de débit (Différence entre le débit cible et réel est approximativement la même), le modèle Intra améliore les performances du codage en atteignant des valeurs élevées de la moyenne du PSNR (+2.05dB pour les séquences "Akiyo" et "News" codées à 25 et 45Kbps respectivement).

3.2.4.2 Performance du modèle Inter R-Q

Les séquences vidéo codées à des débits binaires variant de 15 à 64Kbps, ont été employées pour tester l'efficacité du modèle logarithmique R-Q proposé. Le Tableau 3.3 représente les gains obtenus en termes de moyenne de PSNR, de débit binaire et du temps de calcul entre JM15.0 et l'algorithme proposé.

Séquences	Débit	JM	15.0	Modèle	e Inter	Gain en	Gain en	Gain en		
	Binaire	Débit	PSNR	Débit	PSNR	PSNR	Débit	temps		
	Cible	Réel		Réel		$(\Delta PSNR)$	(ΔR)	(ΔT)		
	(Kbps)	$({ m Kbps})$	(dB)	(Kbps)	(dB)	(dB)	(Kbps)	(Sec)		
Akiyo	15	15.14	34.14	15.14	34.05	-0.09	0	+4.144		
	20	20.17	35.12	20.16	34.96	-0.16	+0.01	+6.159		
	25	25.18	35.77	25.17	35.77	0	+0.01	+5.685		
	30	30.26	36.56	30.28	36.39	-0.17	-0.02	+6.331		
	64	64.38	40.08	64.47	40.02	-0.06	-0.09	+7.428		
Bridge-Close	15	15.15	30.58	15.12	30.52	-0.06	+0.03	+3.725		
	20	20.21	31.15	20.16	31.11	-0.04	+0.05	+3.709		
	25	25.17	31.57	25.17	31.50	-0.07	0	+4.474		
	45	45.39	33.05	45.41	32.99	-0.06	-0.02	+6.84		
	64	64.80	33.97	64.55	33.92	-0.05	+0.25	+7.34		
Container	15	15.13	32.10	15.09	31.97	-0.13	+0.04	+0.407		
	20	20.13	32.97	20.12	32.90	-0.07	+0.01	+3.669		
	30	30.12	34.62	30.14	34.47	-0.15	-0.02	+2.931		
	35	35.15	35.15	35.15	35.04	-0.11	0	+4.249		
	45	45.19	36.09	45.23	36.03	-0.06	-0.04	+4.518		
Grandma	15	15.14	32.69	15.13	32.83	+0.14	+0.01	+5.796		
	20	20.23	33.45	20.15	33.55	+0.1	+0.08	+6.474		
	30	30.18	34.67	30.19	34.73	+0.06	-0.01	+5.794		
	45	45.21	36.09	45.28	36.04	-0.05	-0.07	+6.143		
	Suite du tableau à la page suivant									

TABLE 3.3 – Comparaison entre le modèle Inter R-Q et JM15.0 en termes des gains en PSNR, débit binaire et temps de calcul.

Suite du tableau précédent									
Séquences	Débit	JM1	5.0	Modèle Inter		Gain en	Gain en	Gain en	
	Binaire	Débit	PSNR	Débit	PSNR	PSNR	$\mathbf{D}\mathbf{\acute{e}bit}$	temps	
	Cible	\mathbf{R} éel		\mathbf{R} éel		$(\Delta PSNR)$	(ΔR)	(ΔT)	
	(Kbps)	(Kbps)	(dB)	$({ m Kbps})$	(dB)	(dB)	(Kbps)	(Sec)	
	64	64.29	37.50	64.32	37.41	-0.09	-0.03	+8.145	
News	15	15.89	29.61	15.15	29.42	-0.19	+0.74	+2.595	
	20	20.26	30.46	20.19	30.44	-0.02	+0.07	+0.408	
	25	25.23	31.20	25.25	31.14	-0.06	-0.02	+3.066	
	30	30.27	31.70	30.26	31.71	+0.01	+0.01	+2.688	
	35	35.31	32.16	35.30	32.18	+0.02	+0.01	+4.625	
Salesman	15	15.12	28.84	15.15	28.91	+0.07	-0.03	+6.252	
	20	20.14	29.64	20.14	29.59	-0.05	0	+4.676	
	25	25.15	30.43	25.17	30.31	-0.12	-0.02	+7.908	
	30	30.19	30.98	30.25	30.94	-0.04	-0.06	+6.727	
	35	35.20	31.63	35.22	31.48	-0.15	-0.02	+5.859	
Silent	20	20.25	29.73	20.20	29.70	-0.03	+0.05	+3.819	
	25	25.18	30.51	25.20	30.44	-0.07	-0.02	+8.122	
	35	35.21	31.58	35.24	31.63	+0.05	-0.03	+7.423	
	45	45.29	32.50	45.33	32.46	-0.04	-0.04	+6.054	
	64	64.33	33.94	64.31	33.87	-0.07	+0.02	+9.276	

Les données résumées dans le *Tableau 3.3* permettent d'établir la comparaison entre les performances du JM15.0 et le modèle logarithmique proposé pour différentes séquences de type IPPPP.... On remarque que la mise en œuvre du modèle Inter R-Q permet d'obtenir des moyennes de PSNR et de débit binaire, similaires à celles du JM15.0.

Le même comportement peut être observé au niveau des *Figures 3.6 et 3.7* qui représentent les variations du PSNR et du nombre de bits de codage générées à travers les trames P, des deux séquences "News" et "Silent" codées à 20 et 35Kbps respectivement. Nous constatons sur les deux figures que pour la majorité des trames P, les courbes du modèle logarithmique Inter et celles du JM15.0 se recouvrent. Ce recouvrement signifie que notre proposition permet d'obtenir les mêmes performances que l'algorithme du contrôle de débit appliqué dans JM15.0. Cette similarité entre les deux méthodes est accompagnée d'une diminution de la complexité calculatoire (Jusqu'à +9.276s).

La complexité calculatoire ainsi que l'occupation mémoire entre le schéma du contrôle de débit proposé et celui de H.264/AVC peuvent être comparées comme suit :

• Dans le contrôle de débit de H.264/AVC, chaque unité de base utilise deux régressions linéaires (Modèles linéaire et quadratique). Soit C_r la complexité calculatoire de chaque opération de régression. Afin de déterminer le paramètre de quantification (QP), une opération de racine carrée est requise pour chaque unité de base. De plus, le *MAD* est calculé pour chaque pixel dans la trame. Ainsi, l'ensemble des opérations au sein de chaque trame



FIGURE 3.6 – Comparaison des valeurs du PSNR/Trames P entre JM15.0 et le modèle Inter. Séquences manipulées : A- "Silent" codée à 35Kbps et B- "News" codée à 20Kbps.



FIGURE 3.7 – Comparaison du nombre de bits/Trames P entre JM15.0 et le modèle *Inter*. Séquences manipulées : A- "Silent" codée à 35Kbps et B- "News" codée à 20Kbps.

peut être formulé comme suit :

$$C_{H.264RC} = 2 \times C_r \times N_{bu} + C_{mad} + N_{bu} \times C_{sqrt} + C_{autres} \times N_{bu}$$
(3.14)

 N_{bu} est le nombre des unités de base appartenant à la trame; C_{sqrt} représente l'opération de la racine carrée; C_{autres} est l'ensemble de toutes les autres opérations effectuées au sein de l'unité de base et C_{mad} représente le calcul total des opérations du *MAD* qui est proportionnel au nombre de pixels dans une image.

Le schéma proposé du contrôle de débit utilise une seule régression linéaire liée à la mise à jour des coefficients du modèle logarithmique. Ainsi, l'Eq. 3.14 devient :

$$C_{RCPropos\acute{e}} = C_r \times N_{bu} + N_{bu} \times C_{Log} + C_{autres} \times N_{bu}$$

$$(3.15)$$

 C_{Log} représente l'opération du calcul du logarithmique.

• Dans le contrôle de débit, deux vecteurs sont demandés afin de stocker les valeurs du *MAD* pour chaque unité de base et deux autres sont utilisés pour sauvegarder les paramètres de quantification et les bits générés pour chaque unité de base dans la fenêtre d'estimation. Dans le schéma proposé, deux vecteurs uniquement sont employés pour stocker les paramètres de quantification et les débits binaires réels pour chaque unité de base. Ceci permet de minimiser la complexité de la mémoire en comparaison avec le contrôle de débit dans H.264/AVC.

3.2.4.3 Performance des modèles combinés Intra et Inter

Le modèle Intra utilisé pour déterminer le QP initial est combiné avec le modèle logarithmique employé pour estimer les paramètres de quantification (QPs) des unités de base P. Le *Tableau 3.4* présente les gains en termes du PSNR, du débit binaire et du temps de calcul des deux modèles combinés. La comparaison est effectuée avec l'algorithme JM15.0.

TABLE 3.4 – Comparaison entre les modèles combinés Intra&Inter et JM15.0 en termes des gains en PSNR, débit binaire et temps

Séquences	Débit	JM15.0		Intra&Inter		Gain en	Gain en	Gain en	
	Binaire	Débit	PSNR	Débit	PSNR	PSNR	Débit	\mathbf{temps}	
	Cible	\mathbf{R} éel		\mathbf{R} éel		$(\Delta PSNR)$	(ΔR)	(ΔT)	
	(Kbps)	$({ m Kbps})$	(dB)	$({ m Kbps})$	(dB)	(dB)	(Kbps)	(Sec)	
Akiyo	15	15.14	34.14	15.14	35.49	+1.35	0	+8.17	
	20	20.17	35.12	20.18	36.79	+1.67	-0.01	+13.337	
	25	25.18	35.77	25.20	37.83	+2.06	-0.02	+11.625	
	30	30.26	36.56	30.22	38.40	+1.84	+0.04	+10.011	
	64	64.38	40.08	64.31	41.60	+1.52	+0.07	+9.705	
Suite du tableau à la page suivante									

Suite du tableau précédent									
Séquences	Débit	JM15.0		Intra&Inter		Gain en	Gain en	Gain en	
	Binaire	Débit	PSNR	$\mathbf{D}\mathbf{\acute{e}bit}$	PSNR	PSNR	$\mathbf{D}\mathbf{\acute{e}bit}$	temps	
	Cible	Réel		Réel		$(\Delta PSNR)$	(ΔR)	(ΔT)	
	(Kbps)	(Kbps)	(dB)	$({ m Kbps})$	(dB)	(dB)	(Kbps)	(Sec)	
Bridge-Close	15	15.15	30.58	15.15	31.76	+1.18	0	+14.483	
	20	20.21	31.15	20.18	32.46	+1.31	+0.03	+13.629	
	25	25.17	31.57	25.21	33.03	+1.46	-0.04	+14.094	
	45	45.39	33.05	45.40	34.36	+1.31	-0.01	+14.939	
	64	64.80	33.97	64.45	35.16	+1.19	+0.35	+10.75	
Container	15	15.13	32.10	15.14	33.23	+1.13	-0.01	+12.028	
	20	20.13	32.97	20.15	34.67	+1.7	-0.02	+9.777	
	30	30.12	34.62	30.15	36.09	+1.47	-0.03	+6.878	
	35	35.15	35.15	35.15	36.65	+1.5	0	+7.898	
	45	45.19	36.09	45.19	37.52	+1.43	0	+7.546	
Grandma	15	15.14	32.69	15.13	33.91	+1.22	+0.01	+12.664	
	20	20.23	33.45	20.14	34.95	+1.5	+0.09	+13.208	
	30	30.18	34.67	30.17	36.28	+1.67	+0.03	+10.534	
	45.02	45.21	36.09	45.31	37.93	+1.84	-0.1	+10.627	
	64	64.29	37.50	64.34	39.05	+1.55	-0.05	+11.701	
News	15	15.89	29.61	15.33	29.61	0	+0.56	+12.91	
	20	20.26	30.46	20.22	30.96	+0.5	+0.04	+12.366	
	25	25.23	31.20	25.18	32.27	+1.07	+0.05	+13.476	
	35	35.31	32.16	35.28	34.00	+1.84	+0.03	+14.63	
	45	45.35	33.27	45.37	35.28	+2.01	+0.02	+12.404	
Salesman	15	15.12	28.84	15.13	28.76	-0.08	-0.01	+10.42	
	20	20.14	29.64	20.15	30.16	+0.52	-0.01	+17.447	
	25	25.15	30.43	25.16	31.35	+0.92	-0.01	+18.002	
	30	30.19	30.98	30.19	32.22	+1.24	0	+17.064	
	35	35.20	31.63	35.18	33.04	+1.41	+0.02	+16.498	
Silent	20	20.25	29.73	20.18	30.04	+0.31	+0.07	+15.396	
	25	25.18	30.51	25.19	31.11	+0.6	-0.01	+15.595	
	35	35.21	31.58	35.21	32.79	+1.21	0	+14.808	
	45	45.29	32.50	45.32	34.03	+1.53	-0.03	+15.826	
	64	64.33	33.94	64.46	35.68	+1.74	-0.13	+12.43	

Nous observons d'après le tableau ci-dessus que le modèle combiné améliore l'efficacité du codage en réduisant davantage, le temps de calcul (Jusqu'à +18.002s). Cette minimisation de la complexité calculatoire est accompagnée d'une amélioration de la moyenne du PSNR (Jusqu'à +2.06dB) avec une différence entre le débit cible et réel qui est approximativement nulle. Les Figures 3.8 et 3.9 montrent les performances RDO et le temps de calcul de plusieurs séquences vidéo codées à différents débits binaires.



FIGURE 3.8 – Comparaison de la qualité en termes de PSNR entre JM15.0 et les modèles Intra&Inter. Séquences manipulées : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grandma", E- "News" et F- "Salesman" codées à différents débits binaires.

3.3. CONCLUSION

En comparaison avec le logiciel de référence JM15.0, nous remarquons que la combinaison des modèles Intra et Inter R-Q permet d'améliorer la qualité mesurée en termes de PSNR pour toutes les séquences manipulées. Cette amélioration est aux alentours de +1.5dB sauf pour 'News' et "Salesman" codées aux débits binaires inférieurs à 25Kbps.

Nous observons également sur la *Figure 3.9* que l'écart en temps de codage de l'approche proposée par rapport au logiciel de référence JM15.0 est important sachant que les conditions de tests sont identiques.

Sur les *Figures 3.10 et 3.11* sont tracées les variations du PSNR et du nombre de bits à travers les trames P des séquences "Bridge-Close", "Grandma", "Salesman" et "Container" codées respectivement à 15, 20, 30 et 45Kbps.

La *Figure 3.10* montre que l'application de la combinaison des modèles Intra et Inter permet d'obtenir une variation des valeurs de PSNR des dernières trames inférieure à celle des premières trames. Toutefois, l'amélioration du PSNR des premières trames de chaque séquence vidéo est assez significative. Ceci rendra les variations non perceptibles à des débits binaires faibles et augmentera la qualité des dernières trames. Pour chaque séquence, la variation du PSNR est faible en comparaison avec le JM15.0 et conduit à une qualité consistante et plus élevée.

La *Figure 3.11* montre que la distribution réelle des bits à travers les trames P des quatre séquences précédemment décrites. Nous remarquons que la variation des bits/trame est plus faible pour l'approche proposée. De plus, la moyenne des bits/trame reste inférieure à la moyenne des bits/trame lorsque l'algorithme du JM15.0 est appliqué.

3.3 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle approche de modélisation débit - quantification (R-Q) pour les unités de base I et P. Cette dernière vise à pallier certains défauts de l'algorithme du contrôle de débit de H.264/AVC. Ainsi, nous résolvons tout d'abord le calcul inapproprié du paramètre de quantification initial (QP_I) utilisé pour coder la trame I et la première trame P, en proposant un modèle Intra R-Q non linéaire capable d'estimer efficacement le QP I. Ce modèle approxime la relation liant le débit global utilisé pour coder la totalité de la séquence, le nombre de bits cible pour coder la trame I et sa complexité sans avoir à déterminer cette dernière. Nous avons également résolu les problèmes provenant de l'utilisation du MAD comme mesure de complexité et de sa prédiction via un modèle linéaire utilisés pour prédire le paramètre de quantification des unités de base P. Un nouveau modèle R-Q logarithmique a été ainsi proposé, permettant de substituer les deux modèles linéaire et quadratique employés dans le schéma du contrôle de débit de H.264/AVC, et d'estimer le paramètre de quantification sans aucune mesure de complexité. Les résultats expérimentaux ont été menés sur un ensemble de séquences avec un style de GOP de type IPPPP..., codées à des débits binaires allant de 15 à 64Kbps. Nous avons testé l'efficacité du modèle Intra R-Q, puis celle du modèle Inter R-Q et finalement celle des modèles combinées. Nous avons comparé les résultats obtenus avec ceux issus du logiciel de référence JM15.0, l'une des versions récentes de la norme H.264/AVC. L'adoption du modèle Intra R-Q aide à mieux contrôler le débit de la trame I et à améliorer sa qualité et celle des trames suivantes dans le GOP. Le modèle Inter atteint la même performance (en termes de



FIGURE 3.9 – Comparaison du temps de codage entre JM15.0 et les modèles Intra&Inter. Séquences manipulées : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grandma", E- "News" et F- "Salesman" codées à différents débits binaires.



FIGURE 3.10 – Comparaison des valeurs du PSNR/Trames P entre JM15.0 et les modèles Intra&Inter combinés. Séquences manipulées : A- "Bridge-Close" codée à 15Kbps, B- "Container" codée à 45Kbps, C- "Grandma" codée à 20Kbps et D- "Salesman" codée à 30Kbps.



FIGURE 3.11 – Comparaison du nombre de bits/Trame-P entre JM15.0 et les modèles Intra&Inter combinés. Séquences manipulées : A- "Bridge-Close" codée à 15Kbps, B- "Container" codée à 45Kbps, C- "Grandma" codée à 20Kbps et D- "Salesman" codée à 30Kbps.

3.3. CONCLUSION

PSNR et de débit binaire) que celle du JM15.0 tout en minimisant la complexité calculatoire et celle de la mémoire. L'approche combinant les deux modèles Intra et Inter œuvre efficacement en améliorant les performances R-D en bas débits (Jusqu'à +1.5dB), tout en réduisant davantage le temps de codage. La qualité à travers la totalité de la séquence est améliorée et est moins variable. L'allocation de bits au niveau de l'unité de base P constitue également une autre défaillance du contrôleur de débit du standard H.264/AVC. Cette allocation de bits suppose que la séquence vidéo est stationnaire et que les unités de base adjacentes ont des caractéristiques similaires. Ainsi, le nombre de bits cible de chaque unité de base est estimé en divisant équitablement le débit binaire global (utilisé pour coder la totalité de la séquence) entre les unités non codées sans aucune considération de leurs caractéristiques spatio-temporelles. Cette allocation non uniforme est surement non appropriée surtout dans le cas de séquences vidéo à forts mouvements ou à changement de scènes. Dans le chapitre suivant, nous allons présenter deux approches d'allocation de bits pour les trames P. La première approche introduit deux mesures de complexités temporelles dans le processus d'allocation de bits. Ces mesures permettent de distribuer soigneusement le budget de bits selon l'importance de la trame P. La seconde approche consiste à réajuster le paramètre de quantification d'une trame donnée selon le nombre de ses pixels saillants. Les pixels saillants sont déterminés à partir d'une carte de saillance simulant l'attention visuelle humaine et détectant les régions importantes.

CHAPITRE 4 ALLOCATION DE BITS AU NIVEAU DES TRAMES P

Dans ce chapitre, nous allons introduire deux nouvelles approches d'allocation dynamique de bits au niveau des trames P. La première approche consiste à exploiter deux mesures de complexité dans le processus d'allocation de bits. Ces deux mesures sont capables de favoriser les trames P importantes en leur accordant un budget de bits adéquat pour leur codage. La seconde approche réajuste le paramètre de quantification des trames P tout en exploitant la notion de saillance visuelle représentée sous forme d'une carte. Cette dernière simule l'attention visuelle humaine en adoptant une approche dite "ascendante" (bottom-up). Le QP est ainsi ajusté en attribuant plus de bits aux zones saillantes. L'organisation de ce chapitre est la suivante : dans un premier temps, nous allons introduire les deux approches proposées pour améliorer la technique d'allocation de bits au niveau trame. Ensuite, les résultats de simulation mettant en œuvre les deux deux idées proposées vont être présentés et discutés.

4.1 Allocation de bits au niveau trame P

Dans H.264/AVC, l'allocation de bits au niveau trame reste inefficace car le nombre de bits est subdivisé équitablement entre les trames, sans aucune considération de leurs caractéristiques spatiales ou temporelles. Les solutions proposées dans la littérature et présentées dans le second chapitre; cherchent à allouer de manière appropriée plus de bits aux trames complexes. Ces techniques se diffèrent dans la définition de la mesure de complexité.

4.1.1 Allocation de bits basée sur des mesures de complexité

En considérant toutes les techniques précédemment décrites (*Chapitre 2*), une nouvelle méthode d'allocation de bits est proposée.Elle permet d'assigner les bits selon deux mesures temporelles. Nous rappelons que nous agissons dans le profil de base du codeur H.264/AVC. Ainsi, les trames B ne sont pas prises en considération car elles n'exploitent qu'une faible partie du budget
total alloué à la séquence vidéo. Cette petite portion peut être ignorée dans le cas d'un codage à bas débit. En se basant sur cette nouvelle considération, l'Eq. 1.46 (*Chapitre 1*) permettant de déterminer le nombre de bits cible nécessaire pour coder la $i^{\text{ème}}$ trame P deviendra :

$$T_{i,n} = \frac{B_{i,n}}{NP_n} \tag{4.1}$$

 $B_{i,n}$ est le nombre de bits restant après codage de la $i^{\text{ème}}$ trame, NP_n est le nombre des trames P restantes dans le $n^{\text{ème}}$ GOP. En prenant en compte les deux mesures de complexité temporelle, l'Eq. 4.1 deviendra :

$$\widehat{TB_{i,n}} = T_{i,n} \times C_1 \left(1, i - 1, n \right) \times C_2 \left(i, i - 1, n \right)$$
(4.2)

 $C_1(1, i-1, n)$ et $C_2(i, i-1, n)$ représentent respectivement l'activité temporelle de la $(i-1)^{\text{ème}}$ trame par rapport aux trames précédentes et celle de la $i^{\text{ème}}$ trame par rapport à son antécédente (i-1) dans le $n^{\text{ème}}$ GOP.

La détermination de ces deux meures de complexité peut être décrite comme suit :

• Le calcul de la mesure $C_1(1, i - 1, n)$ consiste à corriger le nombre de bits alloués à la trame courante en fonction de la complexité de mouvement entre la trame précédente (i-1) et toutes celles qui ont été codées précédemment :

$$C_1(1, i-1, n) = \frac{B_{i-1,n}^{reel}}{\frac{1}{i-2} \times \sum_{k=1}^{i-2} B_{k,n}^{reel}}$$
(4.3)

 $B_{i-1,n}^{reel}$ représente le nombre de bits réel utilisé pour coder la $(i-1)^{\text{ème}}$ trame du $n^{\text{ème}}$ GOP. C_1 prouves que la corrélation temporelle entre deux trames consécutives joue un rôle important dans l'estimation du nombre de bits utilisé pour coder chaque trame. Ainsi, le nombre de bits réel employé pour coder les trames précédentes peut précisément représenter la complexité de mouvement. Les *Figures 4.1A* et *4.1B* illustrent cette relation.

En comparant les courbes des deux figures, nous remarquons que les allures des courbes représentant la complexité de mouvement (*Figure 4.1A*) et la répartition des bits à travers les trames P (*Figure 4.1B*) sont corrélées. Ceci implique que $C_1(1, i - 1, n)$ est capable de prédire les contenus mouvementés et d'estimer avec précision le nombre de bits cible.

• L'idée de base supportée par la mesure $C_2(i, i - 1, n)$ consiste à exploiter à la fois, le niveau de différence entre deux trames consécutives en chaque pixel et la probabilité d'occurrence de ce niveau. Ainsi, $C_2(i, i - 1, n)$ suggère de pondérer chaque pixel de la trame de différence (calculée entre deux trames adjacentes), par sa valeur déterminée à partir de l'Histogramme de différences (HOD). $C_2(i, i - 1, n)$ est définie comme suit :

$$C_{2}(i, i-1, n) = \frac{\frac{1}{N \times M} \sum_{p=1}^{N} \sum_{q=1}^{M} I_{i,n}^{d}(p, q) \times HOD(I_{i,n}^{d}(p, q))}{\frac{1}{i-1} \times \sum_{k=1}^{i-1} FD(k, n)}$$
(4.4)



FIGURE 4.1 – Complexité de mouvement $C_1(1, i - 1, n)$, A- et répartition de bits/Trame, B- De la séquence "Foreman" codée à 20Kbps.

M et N sont les dimensions de la trame. L'entité $I_{i,n}^d$ représente la différence absolue entre la $i^{\text{ème}}$ trame courante et son antécédente (appartenant au $n^{\text{ème}}$ GOP), au pixel situé à la position (p,q). Elle est définie par :

$$I_{i,n}^d(p,q) = |I_{i,n}(p,q) - I_{i-1,n}(p,q)|$$
(4.5)

La mesure HOD est déterminé par :

$$HOD(I) = \frac{H(I)}{max(H(I))}$$
(4.6)

H(I) représente l'histogramme de l'image I. FD représente la moyenne des différences des trames adjacentes codées précédemment. Pour une trame donnée, cette entité est définie comme suit :

$$FD(k,n) = \frac{1}{N \times M} \times \sum_{p=1}^{N} \sum_{q=1}^{M} |I_{k,n}(p,q) - I_{k-1,n}(p,q)|$$
(4.7)

L'Eq. 1.47 (*Chapitre* 1) servant à déterminer le nombre de bits cible pour coder la trame courante sera modifiée comme suit :

$$TP_{i,n} = \beta \times \widehat{TB_{i,n}} + (1-\beta) \times \widetilde{T_{i,n}}$$
(4.8)

Une fois le nombre de bits cible $(TP_{i,n})$ est ajusté, de nouvelles considérations HRD (Hypothetical Reference Decoder) sont utilisées afin d'éviter le débordement (overflow) ou le manque de bits (underflow) du buffer du décodeur et la détérioration de la qualité des trames restantes. $(TP_{i,n})$ est délimité comme suit :

$$TP_{i,n} = max \left\{ L_{Bound}, min \left\{ TP_{i,n}, U_{Bound} \right\} \right\}$$

$$(4.9)$$

Avec L_{Bound} et U_{Bound} représentent respectivement, les limites inférieure et supérieure du buffer, elles sont définies par :

$$L_{Bound} = 0.2 \times \frac{RT}{Fr}$$

$$U_{Bound} = 1.5 \times \frac{RT}{Fr}$$
(4.10)

Pour prouver l'efficacité de la nouvelle allocation de bits à base de mesure de complexité, nous l'avons combinée avec le modèle débit-quantification proposé pour les trames Intra (Présenté dans le *Chapitre 3*). Nous avons ensuite testé l'ensemble sur les séquences appartenant à la base *BaseTest*. Deux logiciels de référence ont été considérés (JM15.0 et JVT-O016) pour effectuer la comparaison avec la nouvelle approche proposée. Dans la partie expérimentale, les courbes R-D, les variations du *PSNR*, du nombre de bits, de la qualité et du buffer vont être illustrées et étudiées.

4.1.2 Allocation de bits basée sur la saillance visuelle

Dans cette section, nous allons présenter une approche perceptuelle d'allocation de bits à travers les trames P. Cette dernière utilise un modèle d'attention visuelle exploitant certaines propriétés du Système Visuel Humain. L'idée de base consiste à quantifier finement les trames saillantes et grossièrement les trames non saillantes. La notion de saillance visuelle sera décrite dans la section suivante.

4.1.2.1 Présentation de la carte de saillance

Notre environnement visuel est constitué d'une quantité considérable d'informations devant être appréhendée et interprétée avec précision par notre système visuel. Ceci dans l'optique de sélectionner les régions d'intérêt de cet environnement. Pour effectuer cette sélection, deux mécanismes attentionnels sont considérés afin de traiter minutieusement l'information pertinente. Le premier mécanisme nommé "Top-down" (Descendant) est piloté par la tâche à accomplir à un instant donné, impliquant un contrôle volontaire et cognitif des mouvements oculaires. Cette lente stratégie nécessitant toutes les ressources attentionnelles est déployée pour effectuer une tâche : reconnaître un paysage sur une photo, chercher une personne portant une veste noire... Le second mécanisme exogène appelé "Bottom-up" (Ascendant) permet de sélectionner les informations visuelles selon leur degré de saillance. Ce dernier est calculé à partir des caractéristiques spatio-temporelles du flux lumineux visible dans notre champ visuel. Ce type de traitement est très rapide et automatique.

Plusieurs modèles d'attention visuelle ont été proposés dans la littérature. Pour des raisons de complexité, le modèle détaillé dans [15] est utilisé dans cette partie. Il est basé sur une approche

"Bottom-up". Ce mécanisme de la vision humaine est simulé par des approches multi-résolutions décomposant l'image en différents canaux perceptuels liés aux sélectivités angulaires et radiales du Système Visuel Humain (SVH). Certaines propriétés de ce dernier sont aussi considérées telles que l'intensité, la couleur et la sensibilité au contraste. La *Figure 4.2* illustre le fonctionnement de la génération de la carte de saillance.



FIGURE 4.2 – Algorithme de génération de la carte d'importance perceptuelle.

Dans l'étape de pré-traitement, une conversion de l'espace couleur est d'abord effectuée suivie de l'application de la fonction de sensibilité au contraste CSF (Contrast Function Sensitivity) sur chacune des trois composantes de luminance (Y) et des chrominances (U, V). Ainsi, les valeurs YUV sont transformées afin de simuler l'affichage des valeurs numériques sur écran, l'observation des grandeurs photométriques via les yeux et le traitement du SVH de l'information visuelle en signaux achromatique et chromatique.

Dans l'étape de génération de la carte de saillance, trois canaux sont créés pour traiter trois caractéristiques descendantes (Bottom-up) connues pour leur influence sur l'attention visuelle. Le canal d'intensité traite la composante achromatique de la trame d'entrée générée de la première étape du pré-traitement. Le canal de couleur considère les deux composantes de chrominance C_1 et C_2 .

Finalement, une technique intra de pooling est effectuée pour générer des cartes de points saillants. Les cartes d'importance finales IM_A (Pour la composante de la luminance), et IM_C (Pour les composantes de chrominance) sont obtenues en appliquant un filtre itératif gaussien sur les zones où il existe une forte concentration de points saillants. Ces zones possèdent une forte probabilité d'attirer le regard de l'observateur humain et le filtrage permettra de les mettre en évidence. La carte de saillance finale nommée dorénavant IM est une fusion des deux cartes IM_A et IM_C . La Figure 4.3 présente des exemples de carte d'importance perceptuelle sur des trames extraites de trois séquences différentes.

Les cartes de saillance ainsi que la composante de luminance des trames $75^{\text{ème}}$, $31^{\text{ème}}$ et $56^{\text{ème}}$ des séquences "Container", "Mother-Daughter" et "Salesman" sont illustrées sur la *Figure 4.3*. Les valeurs des pixels varient de 0 (Aucune importance perceptuelle) à 1 (pixels très importants visuellement). D'après les *Figures 4.3C1*, *4.3C2* et *4.3C3*, nous remarquons que les régions les plus claires sont celles ayant une importance perceptuelle très élevée.



FIGURE 4.3 – Cartes de saillance de la : 75^{ème} trame de la séquence "Container" : A1- Originale, B1- Luminance et C1- Carte de saillance, 31^{ème} trame de la séquence "Mother-Daughter" : A2-Originale, B2- Luminance et C2- Carte de saillance, et 56^{ème} trame de la séquence "Salesman" : A3- Originale, B3- Luminance et C3- Carte de saillance.

4.1.2.2 QP ajusté à base de carte de saillance

Dans l'algorithme du contrôle de débit du codeur H.264/AVC, la détermination du paramètre de quantification QP_{H264} se déroule comme suit : Une fois le *MAD* courant prédit via le modèle linéaire et décrit par l'*Eq. 1.46*, le paramètre de quantification de la trame courante est ensuite calculé à l'aide du modèle quadratique (*Eq. 1.54*). Pour maintenir une qualité visuelle consistante à travers les trames P successives, QP_{H264} est ajusté comme suit :

$$QP_{H264} = \min \left\{ QP_{Prev} + DQuant, \\ \max \left\{ QP_{Prev} - DQuant, QP_{H.264} \right\} \right\}$$

$$(4.11)$$

 QP_{Prev} est le paramètre de quantification de la trame précédente P. $DQuant \ge 0$ est un paramètre permettant de limiter la variation des paramètres de quantification des trames adjacentes. Ceci permettra de maintenir un lissage de la qualité à travers les trames de la séquence vidéo traitée. Le paramètre de quantification final est également borné comme suit :

$$QP_{H264} = \min\{51, \max\{QP_{H.264}, 1\}\}$$
(4.12)

Pour décrire comment l'information collectée à travers la carte de saillance peut être employée dans l'ajustement du paramètre de quantification, nous avons commencé par calculer le nombre de pixels saillants et non saillants contenus dans la carte de saillance de chaque trame P comme suit :

$$S_Pix = S_Pix + 1 \quad Si \quad IM(m,n) \ge c$$

$$NS_Pix = NS_Pix + 1 \quad Autrement$$
(4.13)

Pour une trame donnée, IM(m,n) est la valeur du pixel à la position (m,n) dans la carte de saillance IM. S_Pix et NS_Pix représentent respectivement le nombre des pixels importants et non importants visuellement dans la trame manipulée. Ils sont initialisés à 0 après le traitement de chaque trame. Nous supposons qu'un pixel est visuellement important lorsque sa valeur dans la carte de saillance est supérieure à une constante c initialisée empiriquement à la valeur 0.4 après plusieurs expérimentations.

Un facteur de pondération PW basé sur S_Pix et NS_Pix est ensuite déterminé comme suit :

$$PW = \begin{cases} 0, Si \quad S_Pix > \frac{M \times N}{2} \\ \frac{NS_Pix}{M \times N}, Autrement \end{cases}$$
(4.14)

Le facteur PW est égal à 0 lorsqu'il s'agit d'une trame visuellement importante (Nombre de ses pixels saillants dépasse la moitié du total de ses pixels). Dans le cas d'une trame non saillante, PW est égal au rapport $\frac{NS-Pix}{M\times N}$.

Finalement, le paramètre de quantification peut être ajusté de manière adaptative en incluant le facteur PW comme suit :

$$QP_{Ajust} = QP_{H264} \times (1 + PW) \tag{4.15}$$

Puisque l'objectif principal de cette approche est de réaliser des réductions du débit binaire sans aucune détérioration de la qualité visuelle; la valeur du paramètre de quantification ajusté (QP_{Ajust}) doit toujours être supérieure ou égale à la valeur déterminée par l'algorithme du contrôle de débit (QP_{H264}) .

4.2 Analyse de la performance des approches proposées

Pour analyser la performance des approches proposées, nous avons sauvegardé les mêmes conditions de test (décrites dans le *Paragraphe 3.3.3* du *Chapitre 3*). Dans ce chapitre, nous comparons la nouvelle approche d'allocation de bits à base de mesure de complexité avec le logiciel de référence JM15.0 et l'algorithme amélioré JVT-O016.

Nous rappelons que dans l'algorithme amélioré du contrôle de débit JVT-O016, la précision des modèles débit-distorsion est sensiblement améliorée. Ceci est réalisé en contournant le dilemme produit entre l'optimisation débit-distorsion et le contrôle de débit. Deux nouveaux modèles (quadratique et de distorsion linéaire) et une prédiction plus précise des bits d'entête ont été développés. Sur la base de ces modifications, une solution optimale d'allocation de bits est réalisée en utilisant une optimisation Lagrangienne.

Afin de tester l'efficacité de l'approche d'allocation de bits proposée, nous l'avons combiné avec le modèle Intra R-Q (présenté dans le *Chapitre 3*). Ces approches implémentées dans le logiciel de référence JM10.2 et nommées dorénavant "Notre+JM10.2", sont comparées au JM15.0. "Notre+JM10.2" est ensuite intégrée dans l'algorithme JVT-O016, et cette nouvelle implémentation désignée dorénavant par "Notre+JVT-O016" est comparée au JVT-O016. Les différences de la moyenne du PSNR et du débit binaire, les schémas R-D, les variations du niveau de buffer et celles de la qualité vont représenter l'ensemble des résultats expérimentaux.

4.2.1 Performance de l'allocation de bits à base des mesures de complexité

Dans cette section, le processus d'allocation de bits au niveau trame ainsi que le modèle Intra R-Q sont combinés et comparés aux deux algorithmes du contrôle de débit (JM15.0 et JVT-O016). Plusieurs tests expérimentaux ont été effectués sur les séquences vidéo appartenant à la base *BaseTest*. Les débits considérés varient toujours de 15 à 64Kbps.

• Schémas Débit-Distorsion (R-D) :

La comparaison avec le logiciel de référence JM15.0 est illustrée dans la *Figure 4.4* ou les courbes R-D de plusieurs séquences vidéo codées à des débits binaires variables sont tracées.

Nous remarquons que la moyenne du PSNR est améliorée pour toutes les séquences vidéo lorsque l'approche "Notre+JM10.2" est appliquée. Cette amélioration délivre un gain de +2.14dB pour la séquence "News" codée à 45Kbps et +2.01dB pour la séquence "Akiyo" codée à 25Kbps.

La comparaison de l'approche "Notre+JVT-O016" avec l'algorithme JVT-O016 est illustrée dans la *Figure 4.5*. Similaires à la *Figure 4.4*, les résultats obtenus mènent à la même conclusion et le gain de la moyenne du PSNR dans ce cas est de +1.96dB pour la séquence "News" codée à 35Kbps.

En outre, nous remarquons sur la *Figure 4.6* que le schéma proposé, lorsqu'il est combiné avec le JM10.2, dépasse à la fois le logiciel de référence JM15.0 et l'algorithme amélioré du JVT-O016 (Les mêmes séquences codées aux mêmes débits binaires).

• Différences des moyennes des débits binaires :

Les différences des moyennes des débits binaires de l'ensemble des implémentations, sont rapportées dans le *Tableau 4.1* (Voir le reste des tests à la fin de ce chapitre). La combinaison de notre algorithme avec le JVT-O016 représente les meilleurs résultats. Par rapport à ce dernier, l'amélioration est significative (Jusqu'à "+0.85Kbps" pour des débits binaires variant entre 15 et 64Kbps). Ce gain en débit génère également un gain en PSNR (+1.2dB, "Akiyo" codée à 64Kbps).

Pour mieux illustrer l'efficacité du processus d'allocation de bits, considérons un PSNR cible égal à 40dB. Pour la séquence "Akiyo", cette qualité (40.95dB) est atteinte avec un débit binaire de 64Kbps (JVT-O016), tandis qu'une valeur approximative de la même qualité (40.53dB) est obtenue à 45Kbps lorsque notre approche proposée est appliquée ("Notre+JVT-O016"). Nous remarquons également que la contrainte de bande passante est mieux satisfaite avec l'application du nouveau processus d'allocation de bits, le nombre de bits réel est toujours inférieur ou égal au nombre de bits actuel. La comparaison avec le logiciel de référence JM15.0 mène aux mêmes conclusions. Les différences entre les moyennes des débits sont également importantes mais restent légèrement plus faibles que dans le cas



FIGURE 4.4 – Courbes Débit-Distorsion de différentes séquences codées à des débits binaires variables lorsque l'approche "Notre+JM10.2" est appliquée et comparée avec l'algorithme du JM15.0 : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grandma", E- "News" et F- "Sa-lesman".



FIGURE 4.5 – Courbes Débit-Distorsion de différentes séquences codées à des débits binaires variables lorsque l'approche "Notre+JVT-O016" est appliquée et comparée avec l'algorithme du JVT-O016 : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grandma", E- "News" et F- "Salesman".



FIGURE 4.6 – Courbes Débit-Distorsion de différentes séquences codées à des débits binaires variables lorsque l'approche "Notre+JVT-O016" est appliquée et comparée à la fois avec les algorithmes du JM15.0 et du JVT-O016 : A- "Akiyo", B- "Bridge-Close", C- "Container", D- "Grandma", E- "News" et F- "Salesman".

précédent (Comparaison avec le JVT-O016).

TABLE 4.1 – Comparaison des moyennes de PSNR et des différences de débit binaire entre le logiciel de référence JM15.0 et l'approche proposée "Notre+JM10.2" et entre cette dernière implémentée au sein du JVT-O016 ("Notre+JVT-O016") et l'algorithme du JVT-O016.

Séquences	Débit	Différentes	Débit	PSNR	Gain en	Gain
	Binaire	Méthodes	Réel	Moyen	PSNR	En Débit
	Cible				Moyen	
	(Kbps)		(Kbps)	(dB)	(dB)	(Kbps)
Akiyo	15	JM15.0	15.14	34.14	+1.47	+0.05
		Notre+JM10.2	15.09	35.61		
		JVT-0016	15.10	34.29	+1.33	+0.08
		Notre+JVT-O016	15.02	35.62		
	25	JM15.0	25.18	35.77	+2.01	+0.13
		Notre+JM10.2	25.05	37.78		
		JVT-0016	25.12	36.38	+1.56	+0.14
		Notre+JVT-O016	24.98	37.94		
	35	JM15.0	35.22	37.22	+1.78	+0.26
		Notre+JM10.2	34.96	39.00		
		JVT-0016	35.17	37.90	+1.41	+0.38
		Notre+JVT-O016	34.79	39.31		
	45	JM15.0	45.37	38.40	+1.73	+0.43
		Notre+JM10.2	44.94	40.13		
		JVT-0016	45.18	39.15	+1.38	+0.57
		Notre+JVT-O016	44.61	40.53		
	64	JM15.0	64.38	40.08	+1.52	+0.89
		Notre+JM10.2	63.49	41.60		
		JVT-0016	64.15	40.95	+1.20	+0.85
		Notre+JVT-O016	63.30	42.15		
Bridge-Close	15	JM15.0	15.15	30.58	+1.19	-0.02
		Notre+JM10.2	15.17	31.77		
		JVT-0016	15.13	30.50	+1.39	+0.03
		Notre+JVT-O016	15.10	31.89		
	25	JM15.0	25.17	31.57	+1.42	+0.02
		Notre+JM10.2	25.15	32.99		
		JVT-0016	25.14	31.48	+1.57	+0.04
		Notre+JVT-O016	25.10	33.05		
	35	JM15.0	35.32	32.40	+1.39	+0.25
		Su	ite du t	ableau	à la pag	e suivante

4.2. ANALYSE DE LA PERFORMANCE DES APPROCHES PROPOSÉES

Séquences	Débit	Différentes	$\mathbf{D}\mathbf{\acute{e}bit}$	PSNR	Gain en	Gain
	Binaire	Méthodes	Réel	Moyen	PSNR	En Débit
	Cible				Moyen	
	(Kbps)		$({ m Kbps})$	(dB)	(dB)	(Kbps)
		Notre+JM10.2	35.07	33.79		
		JVT-0016	35.22	32.30	+1.52	+0.19
		Notre+JVT-O016	35.03	33.82		
	45	JM15.0	45.39	33.05	+1.38	+0.25
		Notre+JM10.2	45.14	34.43		
		JVT-0016	45.31	33.04	+1.41	+0.20
		Notre+JVT-O016	45.11	34.45		
	64	JM15.0	64.80	33.97	+1.20	+0.51
		Notre+JM10.2	64.29	35.17		
		JVT-0016	64.35	34.09	+1.18	+0.32
		Notre+JVT-O016	64.03	35.27		
Container	15	JM15.0	15.13	32.10	+1.12	+0.01
		Notre+JM10.2	15.12	33.22		
		JVT-0016	15.12	32.00	+1.37	0
		Notre+JVT-O016	15.12	33.37		
	25	JM15.0	25.14	33.90	+1.61	+0.09
		Notre+JM10.2	25.05	35.51		
		JVT-0016	25.09	33.80	+1.71	+0.06
		Notre+JVT-O016	25.03	35.51		
	35	JM15.0	35.15	35.15	+1.48	+0.26
		Notre+JM10.2	34.89	36.63		
		JVT-O016	35.03	35.21	+1.43	+0.08
		Notre+JVT-O016	34.95	36.64		
	45	JM15.0	45.19	36.09	+1.36	+0.23
		Notre+JM10.2	44.96	37.45		
		JVT-O016	45.13	36.34	+1.18	+0.12
		Notre+JVT-O016	45.01	37.52		
	64	JM15.0	64.26	37.46	+0.94	+0.45
		Notre+JM10.2	63.81	38.40		
		JVT-0016	64.15	37.72	+0.88	+0.35
		Notre+JVT-O016	63.80	38.60		
Grandma	15	JM15.0	15.14	32.69	+1.18	+0.04
		Notre+JM10.2	15.10	33.87		
		JVT-0016	15.14	32.92	+1.05	+0.06
		Notre+JVT-O016	15.08	33.97		
		Su	ite du t	ableau	à la pag	e suivante

Séquences	Débit	Différentes	Débit	PSNR	Gain en	Gain
Soquenees	Binaire	Méthodes	Réel	Moven	PSNR	En Débit
	Cible				Moven	
	(Kbps)		(Kbps)	(dB)	(dB)	(Kbps)
	25	JM15.0	25.18	34.16	+1.58	+0.06
		Notre+JM10.2	25.12	35.74		
		JVT-0016	25.11	34.47	+1.49	+0.06
		Notre+JVT-O016	25.05	35.96		
	35	JM15.0	35.21	35.13	+1.82	+0.12
		Notre+JM10.2	35.09	36.95		
		JVT-0016	35.14	35.47	+1.67	+0.19
		Notre+JVT-O016	34.95	37.14		
	45	JM15.0	45.21	36.09	+1.82	+0.17
		Notre+JM10.2	45.04	37.91		
		JVT-0016	45.21	36.44	+1.66	+0.23
		Notre+JVT-O016	44.98	38.10		
	64	JM15.0	64.29	37.50	+1.61	+0.29
		Notre+JM10.2	64.00	39.11		
		JVT-0016	64.20	38.05	+1.45	+0.35
		Notre+JVT-O016	63.85	39.50		
News	15	JM15.0	15.89	29.61	0	+0.78
		Notre+JM10.2	15.18	29.61		
		JVT-0016	15.10	29.55	+0.26	0
		Notre+JVT-O016	15.10	29.81		
	25	JM15.0	25.23	31.20	+0.98	+0.12
		Notre+JM10.2	25.11	32.18		
		JVT-0016	25.19	31.18	+1.29	+0.20
		Notre+JVT-O016	24.95	32.47		
	35	JM15.0	35.31	32.16	+1.89	+0.23
		Notre+JM10.2	35.08	34.05		
		JVT-0016	35.11	32.39	+1.92	+0.26
		Notre+JVT-O016	34.85	34.31		
	45	JM15.0	45.38	33.21	+2.14	+0.42
		Notre+JM10.2	44.96	35.35		
		JVT-0016	45.21	33.79	+1.71	+0.45
		Notre+JVT-O016	44.76	35.50		
	64	JM15.0	64.37	35.30	+1.83	+0.46
		Notre+JM10.2	63.91	37.13		
		JVT-O016	64.13	35.87	+1.39	+0.66
		Su	ite du t	ableau	à la pag	e suivante

4.2. ANALYSE DE LA PERFORMANCE DES APPROCHES PROPOSÉES 101

Séquences	Débit	Différentes	Débit	PSNR	Gain en	Gain
	Binaire	Méthodes	Réel	Moyen	PSNR	En Débit
	Cible				Moyen	
	(Kbps)		(Kbps)	(dB)	(dB)	(Kbps)
		Notre+JVT-O016	63.47	37.26		
Salesman	15	JM15.0	15.12	28.84	+0.39	-0.01
		Notre+JM10.2	15.13	29.23		
		JVT-0016	15.11	29.15	-0.19	+0.01
		Notre+JVT-O016	15.10	28.96		
	25	JM15.0	25.15	30.43	+0.83	+0.05
		Notre+JM10.2	25.10	31.26		
		JVT-0016	25.12	30.65	+1.13	+0.07
		Notre+JVT-O016	25.05	31.78		
	35	JM15.0	35.20	31.63	+1.35	+0.21
		Notre+JM10.2	34.99	32.98		
		JVT-0016	35.13	31.85	+1.64	+0.16
		Notre+JVT-O016	34.97	33.49		
	45	JM15.0	45.35	32.60	+1.76	+0.39
		Notre+JM10.2	44.96	34.36		
		JVT-0016	45.24	33.11	+1.85	+0.37
		Notre+JVT-O016	44.87	34.96		
	64	JM15.0	64.37	34.27	+1.90	+0.50
		Notre+JM10.2	63.93	36.17		
		JVT-0016	64.16	34.96	+1.78	+0.38
		Notre+JVT-O016	63.78	36.74		
Silent	15	JM15.0	16.46	29.18	-0.60	+1.31
		Notre+JM10.2	15.15	28.58		
		JVT-0016	15.08	28.82	-0.10	+0.01
		Notre+JVT-O016	15.07	28.72		
	25	JM15.0	25.18	30.51	+0.56	+0.06
		Notre+JM10.2	25.12	31.07		
		JVT-0016	25.06	30.50	+0.70	+0.09
		Notre+JVT-O016	24.97	31.20		
	35	JM15.0	35.23	31.63	+1.15	+0.17
		Notre+JM10.2	35.06	32.78		
		JVT-0016	35.14	31.70	+1.18	+0.25
		Notre+JVT-O016	34.89	32.88		
	45	JM15.0	45.29	32.50	+1.44	+0.27
		Notre+JM10.2	45.02	33.94		

Suite du ta	uite du tableau précédent									
Séquences	$\mathbf{D}\mathbf{\acute{e}bit}$	Différentes	$\mathbf{D}\mathbf{\acute{e}bit}$	PSNR	Gain en	Gain				
	Binaire	Méthodes	Réel	Moyen	PSNR	En Débit				
	Cible				Moyen					
	(Kbps)		$({ m Kbps})$	(dB)	(dB)	(Kbps)				
		JVT-0016	45.15	32.62	+1.60	+0.41				
		Notre+JVT-O016	44.74	34.22						
	64	JM15.0	64.40	33.93	+1.71	+0.65				
		Notre+JM10.2	63.75	35.64						
		JVT-0016	64.13	34.23	+1.57	+0.62				
		Notre+JVT-O016	63.51	35.80						

• Contribution du processus d'allocation de bits des trames P :

Une comparaison des gains obtenus lorsque le modèle Intra R-Q est appliqué (Tableau 3.2. Chapitre 3) et ceux obtenus lorsque ce modèle est combiné avec le processus d'allocation de bits (Approche "Notre+JM10.2") (Tableau 4.1), met en évidence le processus d'allocation de bits en montrant que ce dernier améliore les débits binaires réels tout en maintenant des moyennes de PSNR similaires. Cette amélioration, observée pour toutes les séquences et toute la gamme des débits choisis, est réalisée principalement grâce à la technique d'allocation qui répartit les bits en fonction de l'importance des trames.

L'efficacité de l'allocation de bits est clairement illustrée sur : les Figures 4.7A et 4.7C qui représentent les variations du PSNR et les Figures 4.7B et 4.7D illustrant la répartition des bits à travers les trames P des séquences "Bridge-Close" et "Silent" codées à 35 et 64Kbps respectivement. Nous remarquons que lorsqu'il y a besoin, plus de bits sont alloués aux trames ce qui provoque des gains locaux en PSNR (PSNR/Trame).

• Variations du PSNR et du nombre de bits par trame :

Les Figures 4.8 et 4.9 représentent respectivement, les variations du PSNR et du nombre de bits par trames P de quatre séquences vidéo : "Akiyo", "News", "Silent" et "Salesman" codées respectivement à 25, 35, 45 et 64Kbps.

La variation du PSNR par trame de l'algorithme proposé (Figure 4.8) reste plus importante que les deux algorithmes du contrôle de débit (JM15.0 et JVT-O016) et ce pour les quatre séquences. Cette amélioration est réalisée grâce à l'intégration du modèle Intra R-Q qui détermine les valeurs de QP des deux premières trames en fonction de leurs caractéristiques et par l'application du processus d'allocation de bits qui répartit, d'une manière appropriée, le budget de bits entre les trames P restantes. L'amélioration du PSNR des premières trames est assez significative, et la variation (jusqu'à +8dB pour la séquence "Salesman") principalement observée avec la mise en œuvre du JVT, n'est toujours pas perceptible même à faibles débits.

La Fiqure 4.9 illustre pour les mêmes séquences, la distribution réelle des bits à travers les



FIGURE 4.7 – Variations du PSNR (A, C) et Répartition des bits (B, D) à travers les trames P des séquences "Bridge-Close" et "Silent" codées à 35 et 64Kbps respectivement, lorsque l'approche d'allocation de bits combinée avec le modèle Intra R-Q est comparée avec ce dernier.



FIGURE 4.8 – Variations du PSNR à travers les trames P de quatre séquences vidéo codées à différents débits binaires dans deux cas : L'approche "Notre+JM10.2" est appliquée et comparée avec JM15.0. "Akiyo" codée à 25Kbps : A et "News" codée à 35Kbps : B. Et l'approche "Notre+JVT-O016" est appliquée et comparée avec le JVT-O016. "Silent" codée à 45Kbps : C et "Salesman" codée à "64kbps" : D.



FIGURE 4.9 – Répartition des bits à travers les trames P de quatre séquences vidéo codées à différents débits binaires dans deux cas : L'approche "Notre+JM10.2" est appliquée et comparée avec JM15.0. "Akiyo" codée à 25Kbps : A et "News" codée à 35Kbps : B. Et l'approche "Notre+JVT-O016" est appliquée et comparée avec le JVT-O016. "Silent" codée à 45Kbps : C et "Salesman" codée à "64kbps" : D.

trames P. Nous remarquons que le nombre de bits/Trame est plus faible lorsque nos approches sont adoptées. Ceci signifie que la répartition à base de caractéristiques de trames est plus efficace. De plus la moyenne des bits/Trame des approches proposées reste minimale en comparaison avec la moyenne des bits cibles/Trame.

• Séquences à 3GOPs :

L'efficacité de l'approche "Notre+JM10.2" est également étudiée en fonction de la longueur du GOP.



FIGURE 4.10 – Variations du niveau du buffer pour quatre séquences codées à différents débits binaires et ayant 3GOPs (Période Intra=30) dans deux cas : 1- Le JM15.0 est appliqué sur "Akiyo" codée à 35Kbps, "Grandma" codée à 45Kbps, "Bridge-Close" codée à 64Kbps et "Container" codée à "45Kbps" (A, B). 2- L'approche "Notre+JM10.2" est appliquée sur les mêmes séquences (C, D).

Les Figures 4.10A et 4.10B montrent les variations du buffer pour quatre séquences vi-

déo ("Akiyo", "Grandma", "Container" et "Bridge-Close") codées avec JM15.0 à 35, 45 et 64Kbps respectivement. Pour chaque séquence, 3GOPs (Période Intra=30) sont pris en considération. Les *Figures 4.10C* et 4.10D illustrent les mêmes courbes lorsque notre approche est adoptée. En attribuant aux trames I et à la première trame P des valeurs de paramètres de quantification inférieures à celles affectées par le logiciel de référence JM15.0, le modèle Intra R-Q, comme prévu, entraîne une augmentation du niveau du buffer au début de chaque GOP. Ensuite, le processus d'allocation de bits des trames P, incluant les mesures de complexité temporelle, réduit la variation du buffer observée lorsque le JM15.0 est appliqué. De plus, le niveau du buffer réel est très similaire au niveau de buffer cible. La *Figure 4.11* illustre les variations du PSNR des mêmes séquences ci-dessus codées aux mêmes débits binaires et ayant la même structure (IPPP...IPPP...).



FIGURE 4.11 – Variations du PSNR de quatre séquences codées à des différents débits binaires et ayant 3GOPs, lorsque l'approche "Notre+JM10.2" est appliquée et comparée avec JM15.0. A : "Akiyo" à 35Kbps, B : "Grandma" à 45Kbps, C : "Bridge-Close" à 64Kbps et D : "Container" à 45Kbps.

Nous pouvons remarquer que le même comportement est observé lorsqu'une séquence vidéo donnée, est composée d'un seul GOP. Le PSNR reste toujours élevé avec l'algorithme proposé et l'amélioration est significative pour les premières trames. Par exemple, les gains moyens en PSNR des séquences "Akiyo", "Bridge-Close", "Container" et "Grandma" sont égaux respectivement à : +1.61dB, +1.01dB, +1.18dB et +1.42dB. Nous pouvons déduire que l'approche proposée permet d'atteindre de bonnes performances en termes de débitdistorsion.

• Variation de la qualité :

Un dernier point concerne le lissage de la qualité temporelle (au fil des trames). Sous les contraintes du débit binaire cible et celles du buffer, un algorithme du contrôle de débit performant cherche à maximiser la qualité globale de la séquence vidéo tout en minimisant sa variation à travers toutes les trames. L'Eq.~4.16 est utilisée pour calculer la variation moyenne de la qualité mesurée par la métrique PSNR :

$$mdev = \frac{1}{N} \times \sum_{q=1}^{N} \left| PSNR_q - \overline{PSNR} \right|$$
(4.16)

N et \overline{PSNR} représentent respectivement le nombre de trames par séquence et la moyenne du PSNR.

La Figure 4.12 représente la variation moyenne du PSNR des séquences suivantes : "Akiyo", "Bridge-Close", "Container" et "Claire" codées à de faibles débits binaires allant de 15 à 64Kbps. Nous remarquons que pour les débits binaires compris entre 15 et 20Kbps, les algorithmes proposés garantissent la même qualité que les autres algorithmes du contrôle de débit (JM15.0 et JVT-0016). Pour les débits allant de 25 à 64Kbps, les séquences vidéo sont délivrées avec une qualité consistante et lisse lorsque nos approches sont adoptées.

En résumé, en comparaison avec les algorithmes du contrôle de débit (JM15.0 et JVT-O016), cette solution proposée ne réduit pas uniquement les débits binaires et réalise un gain décent en PSNR moyen, mais délivre également une qualité consistante à travers les séquences testées et codées suivant la gamme des faibles débits binaires sélectionnés.

4.2.2 Performance de l'allocation de bits basée sur la saillance visuelle

L'approche d'ajustement du paramètre de quantification à base de carte de saillance a été implémentée au niveau du logiciel de référence JM15.0. Ainsi, toutes les parties du codeur sont restées identiques sauf l'algorithme du contrôle de débit qui a été modifié par l'insertion de notre approche. Les mêmes conditions de test décrites au niveau du chapitre précédent ont été maintenues. Les séquences appartenant aux deux bases (BaseTr et BaseTest) ont été codées avec ou sans application de l'approche proposée. Quatre valeurs de débits binaires ont été considérées variant de 50 à 100Kbps. Pour mettre en évidence la performance de la méthode proposée, nous avons évalué la réduction de débit binaire ainsi que la qualité de la vidéo reconstruite.



FIGURE 4.12 – Variation du PSNR moyen de quatre séquences : "Akiyo" : A, "Bridge-Close" : B, "Claire" : C et "Container" : D codées à des différents débits binaires variant de 15 à 64Kbps lorsque les approches "Notre+JM10.2" et "Notre+JVT-O016" sont appliquées et comparées respectivement au JM15.0 et au JVT-O016.

La réduction du débit binaire est calculée comme suit :

$$Rate_reduction\,(\%) = 100 \times \left(1 - \frac{Rate_{prop}}{Rate_{H264}}\right) \tag{4.17}$$

 $Rate_{prop}$ et $Rate_{H264}$ représentent respectivement, le débit binaire réel obtenu après codage d'une trame donnée, par l'approche proposée et par l'algorithme du contrôle de débit dans H.264/AVC.

Deux mesures objectives ont été utilisées pour évaluer la qualité de la séquence vidéo reconstruite. Il s'agit du PSNR et du SSIM et décrit dans [97]. A la différence du PSNR, SSIM s'intéresse au contenu structurel de l'image et permet de détecter toute variation de ce contenu. Cette métrique a démontré sa forte corrélation avec le jugement humain.

• Calcul des rapports de débit binaire et des moyennes du PSNR et du SSIM :

TABLE 4.2 – Comparaison de la moyenne du PSNR et celle du SSIM et les ratios de débit binaire entre le logiciel de référence JM15.0 et l'approche proposée.

Séquences	Débit		J M15	.0	QP	Gain		
	cible	Débit	PSNR	Moyenne	Débit	PSNR	Moyenne	Débit
		\mathbf{R} éel		SSIM	\mathbf{R} éel		SSIM	
	(Kbps)	$({ m Kbps})$	(dB)		$({ m Kbps})$	(dB)		(%)
Akiyo	50	50.07	40.10	0.96	45.93	38.94	0.96	8.2684
	64	63.57	41.05	0.96	54.86	39.96	0.96	13.7014
	80	80.02	43.23	0.98	70.75	42.67	0.98	11.5846
	100	100.03	44.46	0.98	93.77	43.62	0.98	6.2581
Carphone	50	51.49	34.35	0.92	44.15	33.39	0.91	14.2552
	64	65.09	35.52	0.93	55.36	34.48	0.93	14.9485
	80	80.52	36.93	0.95	70.85	35.74	0.94	12.0000
	100	100.41	37.95	0.96	85.20	36.97	0.95	15.1479
Claire	50	50.79	42.03	0.97	41.57	40.62	0.96	18.1532
	64	64.72	43.17	0.97	49.47	41.71	0.97	23.5630
	80	80.71	44.77	0.98	67.24	43.63	0.98	16.6894
	100	101.06	45.59	0.98	94.77	45.13	0.98	6.2240
Container	50	50.17	37.11	0.9	44.13	36.00	0.9	12.0391
	64	64.34	37.71	0.91	52.87	36.34	0.9	17.8272
	80	80.62	39.39	0.92	64.59	38.53	0.92	19.8834
	100	99.83	39.83	0.93	82.49	39.24	0.93	17.3695
Grandma	50	50.49	37.34	0.93	42.95	36.06	0.91	14.9337
	64	63.53	38.36	0.94	51.87	37.18	0.93	18.3535
		-		Suite	e du tal	bleau à	la page s	uivante

Séquences	Débit		JM15	.0	Ql	P Percep	otuel	Gain
	\mathbf{cible}	Débit	PSNR	Moyenne	Débit	PSNR	Moyenne	Débit
		Réel		SSIM	Réel		SSIM	
	$({ m Kbps})$	$({ m Kbps})$	(dB)		$({ m Kbps})$	(dB)		(%)
	80	79.96	40.47	0.97	68.61	39.52	0.96	14.1946
	100	99.54	41.44	0.97	86.24	40.33	0.97	13.3615
Highway	50	50.07	36.89	0.89	45.71	36.81	0.9	8.7078
	64	64.03	37.63	0.9	58.00	37.53	0.91	9.4175
	80	80.04	38.83	0.93	72.91	38.38	0.93	8.9080
	100	100.47	39.40	0.93	90.78	39.00	0.93	9.6447
Miss-America	50	50.51	41.93	0.96	41.90	40.59	0.94	17.0461
	64	64.46	42.67	0.96	54.05	41.52	0.95	16.1496
	80	80.31	43.58	0.96	65.72	42.88	0.96	18.1671
	100	100.13	44.14	0.97	82.71	43.62	0.97	17.3974
Mother-daughter	50	50.11	37.74	0.94	44.75	36.77	0.93	10.6965
	64	63.89	39.02	0.95	56.00	37.66	0.94	12.3494
	80	79.76	40.63	0.97	70.85	40.06	0.96	11.1710
	100	99.97	41.55	0.97	93.14	41.32	0.97	6.8320
News	50	50.58	34.63	0.93	37.17	32.85	0.91	26.5125
	64	65.04	35.80	0.94	56.83	35.40	0.94	12.6230
	80	81.14	38.41	0.97	70.86	37.61	0.96	12.6695
	100	101.12	39.33	0.97	92.90	38.88	0.97	8.1290

Suite du tableau précédent

Le Tableau 4.2 montre la comparaison en termes de réduction de débit binaire et de moyennes du PSNR et du SSIM entre l'approche proposée et le logiciel de référence JM15.0. Nous remarquons que l'approche proposée permet d'avoir des réductions variables du débit binaire (de 6 à 26%) en comparaison avec JM15.0. Ces réductions, qui sont importantes par rapport aux débits cibles considérés, sont réalisées grâce au paramètre de quantification adaptatif qui permet d'allouer moins de bits aux trames contenant peu ou pas de régions saillantes. L'allocation d'un faible budget de bits à ce type de trames conduit sans doute à une réduction du PSNR. D'après le Tableau 4.2, nous pouvons constater que les chutes en PSNR varient de 0.08dB (Séquence"Highway" codée à 50Kbps) à 1.78dB (Séquence "News" codée aussi à 50Kbps). Ces chutes sont toutefois non significatives par rapport aux niveaux du PSNR de la totalité de la séquence vidéo traitée. Les baisses en PSNR ne génèrent pas visuellement des différences entre les séquences vidéos codées par les deux approches, puisque les trames saillantes sont codées soigneusement et avec plus de bits lorsque notre approche est appliquée. Finalement, nous remarquons que le même score SSIM est obtenu pour les deux algorithmes.

• Évaluation de la qualité visuelle :

Pour illustrer davantage le résultat de similarité de la qualité visuelle des deux algorithmes (JM15.0 et le nôtre); les *Figures 4.13, 4.14*, et *4.15* donnent un aperçu visuel de deux trames reconstruites des séquences "Container", 'Miss-America", et "Mother-Daughter" codées respectivement à 50, 64, et 80Kbps lorsque l'approche proposée et le logiciel de référence JM15.0 sont appliqués.



FIGURE 4.13 – Comparaison des trames reconstruites de la séquence "Container" codée à 50Kbps. La 62^{ème} trame : A- JM15.0 est appliqué (PSNR=38.03dB, 3240Bits/Pic), B- l'approche proposée est appliquée (PSNR=36.81dB, 520Bits/Pic). La 71^{ème} trame : C- JM15.0 est appliqué (PSNR=38.05dB, 3272Bits/Pic), D- l'approche proposée est appliquée (PSNR=38.13dB, 896Bits/Pic)

A partir de la *Figure 4.13*, nous remarquons que la qualité perçue des deux trames reconstruites (*Figure 4.13B* et 4.13D) de la séquence "Container" lorsque notre approche est appliquée est similaire à celle des trames codées par le JM15.0 (*Figure 4.13A* et 4.13C. Cette similarité de qualité visuelle est accompagnée de réductions de débit binaire (83.95% et 72.62%) et une chute de PSNR (1.2dB).

Les mêmes observations peuvent être faites des deux Figures 4.14 et 4.15 des séquences "Miss-America" et "Mother-Daughter". Comparée à l'algorithme du contrôle de débit JM15.0 (Figure 4.14A et 4.14C et Figure 4.15A et 4.15C), notre approche atteint la même qualité visuelle (Figure 4.14B et 4.14D et Figure 4.15B et 4.15D), malgré les dégradations du PSNR (allant jusqu'à 3.81dB). Les réductions en débit binaire sont respectivement égales à : 87,66% et 72,82% ($25^{\text{ème}}$ et $77^{\text{ème}}$ trames de la séquence "Miss-America"), et 97.98% et 88.72% ($11^{\text{ème}}$ et $74^{\text{ème}}$ trames de la séquence "Mother-Daughter").



FIGURE 4.14 – Comparaison des trames reconstruites de la séquence "Miss-America" codée à 64Kbps. La 25^{ème} trame : A- JM15.0 est appliqué (PSNR=43.63dB, 3760Bits/Pic), B- l'approche proposée est appliquée (PSNR=41.15dB, 464Bits/Pic). La 77^{ème} trame : C- JM15.0 est appliqué (PSNR=43.08dB, 3680Bits/Pic), D- l'approche proposée est appliquée (PSNR=41.72dB, 1000Bits/Pic)



FIGURE 4.15 – Comparaison des trames reconstruites de la séquence "Mother-Daughter" codée à 80Kbps. La 11^{ème} trame : A- JM15.0 est appliqué (PSNR=44.22dB, 13088Bits/Pic), B- l'approche proposée est appliquée (PSNR=40.41dB, 264Bits/Pic). La 74^{ème} trame : C- JM15.0 est appliqué (PSNR=40.31dB, 6952Bits/Pic), D- l'approche proposée est appliquée (PSNR=37.93dB, 784Bits/Pic)

4.3 Conclusion

Dans ce chapitre, nous avons présenté deux nouvelles approches permettant d'améliorer le processus d'allocation de bits de l'algorithme du contrôle de débit du standard H.264/AVC. Les deux techniques proposées sont appliquées au niveau trame. La première approche intègre deux mesures de complexité temporelle dans le calcul du nombre de bits cible utilisé pour coder une trame donnée. La première mesure est un rapport de mouvement déterminé à base du nombre de bits généré après codage d'une trame et celui des trames précédentes. La seconde mesure exploite la différence entre trames adjacentes et l'histogramme de cette différence. Par rapport au logiciel de référence JM15.0 et l'algorithme du contrôle de débit JVT-O016, la combinaison de la nouvelle allocation de bits à base de mesures de complexité et du modèle Intra R-Q (présenté dans le *Chapitre 3*), améliore considérablement les performances en termes de débit-distorsion (R-D), de moyenne de PSNR et de réduction du débit binaire. Une qualité plus consistante à travers les trames des séquences traitées est également atteinte.

La seconde approche consiste à ajuster de manière adaptative le paramètre de quantification en introduisant des informations perceptuelles obtenues à partir des cartes de saillance. Ainsi, le paramètre de quantification d'une trame donnée est pondéré par un facteur déterminé à partir du nombre de pixels saillants et non saillants contenus dans la carte d'attention visuelle. Les résultats de simulation montrent que l'approche proposée réduit considérablement les débits binaires tout en obtenant une qualité perçue équivalente à celle obtenue avec l'algorithme du contrôle du JM15.0.

Il est à noter que le processus d'allocation de bits au niveau macrobloc reste également non optimal. Pour remédier à cette insuffisance, nous allons étendre l'approche de saillance visuelle jusqu'au niveau macrobloc. L'intégration d'une telle approche procédera à une répartition de bits adaptative permettant de promouvoir les macroblocs les plus importants dans une trame donnée, et améliorera davantage la qualité visuelle.

CHAPITRE

5

VERS UN CONTRÔLE PERCEPTUEL DU DÉBIT H.264/AVC

Dans ce chapitre, nous présentons quatre nouvelles approches perceptuelles pour améliorer aussi bien le contrôle de débit que la qualité visuelle des séquences codées H.264/AVC. La première approche est une extension au niveau macrobloc, de l'approche d'ajustement perceptuel du paramètre de quantification présentée dans le chapitre précédent et appliquée au niveau trame. Dans la deuxième approche, le nombre de bits cible de chaque macrobloc est pondéré par son importance visuelle définie par une carte de saillance. La troisième approche consiste à combiner l'algorithme d'ajustement du paramètre de quantification avec le modèle Intra présenté dans le Chapitre 3. Dans la quatrième et dernière approche enfin, la sensibilité au contraste du SVH est utilisée pour pondérer une mesure de complexité classique telle que le MAD. La première approche permet de réduire le débit binaire cible tout en maintenant une qualité visuelle de la vidéo, similaire à celle obtenue par un codage H.264/AVC. La seconde approche permet d'améliorer la qualité visuelle tout en respectant la contrainte de débit cible. La troisième approche fournit des gains importants à la fois en débit et en qualité. L'intérêt de la dernière approche enfin est qu'elle permet d'atteindre des performances intéressantes avec une complexité calculatoire réduite.

5.1 Approches basées perception au niveau macrobloc

5.1.1 Ajustement perceptuel des QPs des macroblocs P

Dans le chapitre précédent, l'attention visuelle humaine est modélisée par une carte de saillance propre à chaque trame. Le nombre de pixels saillants et non saillants est utilisé pour déterminer un facteur de pondération. Ce facteur sert ensuite à ajuster le paramètre de quantification d'une trame. L'idée de base étant de réduire le débit sans perte de qualité visuelle, le paramètre de quantification ajusté est défini pour être toujours supérieur ou égal à celui déterminé par l'algorihtme du contrôle de débit adopté par H.264/AVC. Dans ce paragraphe, cet ajustement au niveau trame est étendu au niveau macrobloc.

Dans H.264/AVC, le paramètre de quantification de chaque macrobloc P est déterminé selon les trois cas suivants :

♦ **Cas 1** : Si le macrobloc considéré est le premier dans la trame courante, alors son paramètre de quantification $(QP_{MB,H264})$ est fixé comme suit :

$$QP_{MB,H264} = QP_{apf} \tag{5.1}$$

 QP_{apf} est la valeur moyenne des paramètres de quantification de tous les macroblocs de la trame précédente.

◊ Cas 2 : Si le nombre de bits cible de la trame courante est négatif. Dans ce cas, le paramètre de quantification devrait être supérieur à celui du macrobloc précédent, c'est-à-dire :

$$Q\widetilde{P_{MB,H264}} = QP_{MB,prev} + DQuant \tag{5.2}$$

 $QP_{MB,prev}$ représente le paramètre de quantification du macrobloc précédent. Pour réduire les artefacts d'effet de blocs, DQuant doit être égal à 1 si le nombre des macroblocs est supérieur à 8, sinon il est égal à 2. Pour maintenir une meilleure qualité visuelle, le paramètre de quantification doit être encore borné par :

$$QP_{MB,H264} = max\left\{1, QP_{apf} - \Delta, min\left\{51, QP_{apf} + \Delta, Q\widehat{P_{MB,H264}}\right\}\right\}$$
(5.3)

où Δ est égal à 3 si le nombre de macroblocs dans une unité de base est inférieur au nombre total des macroblocs constituants la largeur d'une trame, sinon, il est égal à 6.

♦ **Cas** 3 : Le paramètre de quantification $Q\bar{P}_{MB,H264}$ du macrobloc courant est calculé via le modèle quadratique (*Chapitre I, Eq. 1.54*). Ensuite il est borné comme suit :

$$Q\widetilde{P}_{MB,H264} = max \left\{ QP_{MB,prev} - DQuant, \\ min \left\{ Q\widetilde{P}_{MB,H264}, QP_{MB,prev} + DQuant \right\} \right\}$$
(5.4)

Pour réduire les artefacts d'effet de blocs et améliorer la qualité visuelle perçue, le paramètre de quantification estimé par l'Eq. 5.4 est borné par :

$$QP_{MB,H264} = max\left\{1, QP_{apf} - \Delta, min\left\{51, QP_{apf} + \Delta, Q\widetilde{P_{MB,H264}}\right\}\right\}$$
(5.5)

L'intégration de l'information collectée à travers la carte de saillance dans le processus d'ajustement du paramètre de quantification est similaire à celle présentée dans la Section 4.1.2 du Chapitre 4. Après avoir calculé le nombre de pixels saillants S_Pix et non saillants NS_Pix par macrobloc, un facteur de pondération PW est ensuite déterminé comme suit :

$$PW = \begin{cases} 0, Si \quad \left(S_Pix > \frac{16 \times 16}{2}\right) \\ \frac{NS_Pix}{16 \times 16}, Autrement \end{cases}$$
(5.6)

Le facteur PW est égal à 0 lorsqu'il s'agit d'un macrobloc visuellement important (Nombre de ses pixels saillants dépasse la moitié de ses 256 pixels). Dans le cas d'un macrobloc non saillant, PW est égal au rapport $NS_Pix/256$. Finalement, le paramètre de quantification peut être ajusté de manière adaptative en incluant le facteur PW comme suit :

$$QP_{MB,Ajust} = QP_{MB,H264} \times (1 + PW) \tag{5.7}$$

Ainsi définie, la valeur du paramètre de quantification ajusté $(QP_{MB,Ajust})$ est toujours supérieure ou égale à la valeur déterminée par l'algorithme du contrôle de débit $(QP_{MB,H264})$.

5.1.2 Allocation de bits au niveau macrobloc P basée saillance visuelle

Une allocation de bits adéquate consiste à répartir convenablement le budget de bits entre l'ensemble des images et des macroblocs constituants la séquence vidéo tout en considérant leurs importances visuelles. Dans l'algorithme du contrôle de débit adopté par H.264/AVC, l'allocation de bits se fait à trois niveaux : GOP, trame et macrobloc. Au niveau trame, le budget de bits est divisé équitablement entre l'ensemble des images constituant une séquence vidéo donnée. Au niveau macrobloc, le nombre de bits cible de l'image comportant le macrobloc est pondéré par le rapport des MADs des macroblocs précédents, afin de déterminer le nombre de bits cible du macrobloc traité (Eq. 1.51 du Chapitre I).

Cette répartition présente deux défauts : Le nombre de bits cible de la trame est déterminé sans aucune considération de ses caractéristiques spatio-temporelles et l'utilisation du rapport des MADs suppose qu'un macrobloc ayant une large valeur du MAD doit être codé avec plus de bits. En considérant les solutions bibliographiques décrites dans le *Chapitre 2* pour remédier à ces défauts, une nouvelle approche d'allocation de bits est proposée. Elle cherche à répartir les bits entre les macroblocs selon leur importance visuelle calculée à partir de la carte de saillance présentée dans le *Chapitre 4*.

Pour ce faire, cette approche exploite l'efficacité, largement présentée dans le chapitre précédent, du débit cible ajusté TP d'une trame P (Eq. 4.9). A partir de ce débit cible ajusté, l'allocation se fait selon l'Eq. 5.8

$$TB_{MB}(k,n) = TP(n) \times \begin{cases} \frac{\sum_{i=1}^{16} \sum_{j=1}^{16} ROI(i,j,k,n)}{\sum_{k=1}^{N_{MB}} \sum_{i=1}^{16} \sum_{j=1}^{16} ROI(i,j,k,n)}, & Si & \overline{ROI(k,n)} \ge \overline{ROI(n)} \\ \frac{1}{3} \times \left(\frac{\overline{ROI(n)} + \sum_{i=1}^{16} \sum_{j=1}^{16} ROI(i,j,k,n)}{\sum_{k=1}^{N_{MB}} \sum_{i=1}^{16} \sum_{j=1}^{16} ROI(i,j,k,n)} \right), & Sinon \end{cases}$$
(5.8)

 $TB_{MB}(k, n)$ représente le débit cible du $k^{\text{ème}}$ macrobloc appartenant à la $n^{\text{ème}}$ trame P. ROI(i,j,k,n) est la valeur de la saillance du pixel situé à la position (i,j) du $k^{\text{ème}}$ macrobloc appartenant à la $n^{\text{ème}}$ trame P. ROI(n) et ROI(k, n) représentent respectivement, la moyenne de la carte de saillance de la $n^{\text{ème}}$ trame et celle du $k^{\text{ème}}$ macrobloc. N_{MB} est le nombre total des macroblocs. Cette allocation de bits permet à un macrobloc donné dont la moyenne de ses coefficients saillants est supérieure à la moyenne de la saillance de la trame à laquelle il appartient, d'avoir plus de bits que dans le cas d'une répartition équitable $(TP(n)/N_{MB})$. Ce surplus de bits est adaptatif et varie en fonction de l'importance du macrobloc. Pour un macroboc non saillant dont la ROI(k, n) est inférieure à la moyenne de la saillance de la trame, le nombre de bits cible du macrobloc s'adapte également à sa faible importance. Au mieux, quand la moyenne du macrobloc est proche de la moyennes de la trame, le débit cible atteint 2/3 du rapport $TP(n)/N_{MB}$. Pour un macrobloc dont la saillance est proche de zéro, le nombre de bits cible sera égal au tiers du rapport $TP(n)/N_{MB}$.

5.1.3 Ajustement des QPs des macroblocs I et P

L'algorithme du contrôle de débit adopté par le standard H.26/AVC est employé afin d'estimer le paramètre de quantification de chaque trame ou macrobloc. Ce paramètre a un grand impact sur les performances du codeur en termes de débit et de qualité, car il régularise la quantité des détails spatiaux pouvant être sauvegardée. Lorsque le paramètre de quantification a une grande valeur, certains détails sont éliminés. Ceci génère une diminution du débit binaire mais engendre une dégradation de la qualité visuelle. Dans cette section, et pour réaliser des gains en termes de débit et de qualité visuelle, l'approche d'ajustement du paramètre de quantification à base de carte visuelle est associé au modèle Intra R-Q présenté dans le *Chapitre 3*. L'idée est d'ajuster le paramètre de quantification des macroblocs I et P. Pour ce faire, le modèle Intra R-Q sert dans un premier temps à déterminer efficacement le paramètre de quantification initial des deux premières trames. L'algorithme du contrôle de débit étant appliqué au niveau macrobloc, chaque macrobloc appartenant aux deux premières trames est alors codé avec ce paramètre de quantification initial. Pour les macroblocs appartenant au reste des trames P, leurs paramètres de quantification sont ajustés suivant la carte d'attention visuelle.

5.1.4 MAD basé sur les propriétés du Système Visuel Humain

Dans le codeur H.264/AVC, le MAD est habituellement mis en œuvre pour mesurer la distorsion de l'unité de codage. Il est, généralement défini comme la moyenne des différences entre les pixels du macrobloc original et ceux du macrobloc prédit.

$$MAD = \frac{\sum_{p=1}^{16} \sum_{q=1}^{16} |MB_{Orig}(p,q) - MB_{Pred}(p,q)|}{256}$$
(5.9)

 $MB_{Orig}(p,q)$ et $MB_{Pred}(p,q)$ représentent respectivement les valeurs du pixel original et prédit situé à la position (p,q).

Une manière de réduire les variations de la qualité est d'utiliser des estimateurs de qualité plus appropriés que le MAD. Dans ce contexte, trois modifications sont apportées au calcul du MAD:

• La différence absolue pour chaque pixel dans un macrobloc donné est calculée sur la base de l'image reconstruite et non pas celle prédite.

$$MAD_{l} = \frac{\sum_{p=1}^{16} \sum_{q=1}^{16} |MB_{Orig}(p,q) - MB_{Rec}(p,q)|}{256}$$
(5.10)

• Comme le SVH est sensible aux changements de contraste plutôt qu'aux changements de luminance, un contraste local C(p,q) est calculé pour chaque différence absolue du pixel situé à la position (p,q).

$$C(p,q) = \frac{|MB_{Orig}(p,q) - MB_{Rec}(p,q)|}{MAD_{l}}$$
(5.11)

• Le contraste local C(p,q) est pondéré par la fonction de sensibilité au contraste (CSF).

$$C_w(p,q) = C(p,q) \times CSF(p,q)$$
(5.12)

CSF exprime la variation de la sensibilité du SVH au contraste en fonction de la fréquence spatiale. Plusieurs modèles de CSF existent dans la littérature. La fonction utilisée ici est achromatique et isotrope [83] (*Figure 5.1*).



FIGURE 5.1 – Fonction de Sensibilité au Contraste (CSF 2D) Isotrope et achromatique.

• La probabilité d'occurrence des erreurs générées après pondération du contraste par la CSF (C_w) est calculée $(Occ(C_w))$ selon :

$$Occ(k) = card\left(\left\{C_w(p,q) = k, (p,q) \in \{1, \dots, 16\}^2\right\}\right)$$
(5.13)

La mesure finale (MAD_{CSF}) pondère le contraste de chaque erreur par sa probabilité d'occurrence.

$$MAD_{CSF} = \frac{1}{256} \times \sum_{p=1}^{16} \sum_{q=1}^{16} \left(C_w(p,q) \times Occ(C_w(p,q)) \right)$$
(5.14)

Ainsi défini, le MAD_{CSF} intègre les erreurs de faible valeur qui quand elles sont nombreuses, peuvent altérer la qualité. L'efficacité du MAD pondéré par la CSF va être évaluée subjectivement par des tests expérimentaux élaborés sur un ensemble de séquences vidéo.

5.2 Analyse de la performance des approches proposées

Les approches à base de carte de saillance ont été implémentées au niveau du logiciel de référence JM15.0. L'algorithme du contrôle de débit ainsi modifié, est exécuté au niveau macrobloc. Les mêmes conditions de test décrites au niveau du *Chapitre 3* ont été sauvegardées. Les séquences appartenant aux deux bases (*BaseTr* et *BaseTest*) ont été codées avec et sans application de ces approches. Les débits binaires considérés varient de 15 à 80Kbps. La performance des techniques proposées, est évaluée en calculant le taux de réduction de débit binaire (*Eq. 4.17*) ainsi que la qualité de la vidéo reconstruite mesurée en termes de moyennes de PSNR et de SSIM. Nous utilisons également la mesure du PSNR pondéré par certaines propriétés du Système Visuel Humain (PSNR-HVS) [30] [68] [71]. Le PSNR-HVS est connu par sa forte corrélation avec le MOS (Mean Opinion Score). Finalement, la performance du *MAD* pondéré par la CSF est évaluée subjectivement en utilisant la méthode d'échelle de dégradation à double stimulus (DSIS : Double Stimulus Impairment Scale) qui est détaillée ultérieurement.

5.2.1 Performance d'ajustement perceptuel des QPs des macroblocs P

• Calcul du débit binaire et des moyennes du PSNR et du SSIM :

Le Tableau 5.1 illustre le débit binaire et les moyennes du PSNR et du SSIM de l'approche d'ajustement du paramètre de quantification au niveau macrobloc et de l'algorithme de référence JM15.0. La technique proposée permet d'avoir des réductions variables du débit binaire allant de 0.17 à 38.8% en comparaison avec JM15.0. Ces réductions qui sont significatives, sont rendues possibles grâce à l'adaptativité du paramètre de quantification qui alloue beaucoup moins de bits aux macroblocs non saillants. Cette adaptativité, qui est accompagnée d'une légère variation du PSNR (allant de +0.09 pour la séquence "Akiyo" à 30Kbps à -1.3dB pour "Silent" à 80Kbps), permet également de conserver la même qualité visuelle que JM15.0 puisque les deux algorithmes obtiennent le même score SSIM.

TABLE 5.1 – Comparaison de la moyenne du PSNR et celle du SSIM et du débit binaire entre le JM15.0 et l'approche d'ajustement du QP appliquée au niveau macrobloc.

Séquences	Débit		JM15.0)	(Gain		
	cible	Débit	PSNR	SSIM	Débit	PSNR	SSIM	Débit
		Réel				Réel		
	(Kbps)	$({ m Kbps})$	(dB)		(Kbps)	(dB)		(%)
Akiyo	30	30.26	36.56	0.93	28.43	36.66	0.93	6.0475
	45	45.37	38.40	0.95	39.93	38.20	0.95	11.9903
	60	60.58	39.87	0.96	50.79	39.31	0.96	16.1604
	75	75.46	41.05	0.97	58.71	40.15	0.96	22.1971
	80	80.54	42.36	0.98	60.81	41.80	0.98	24.4971
				Sui	te du ta	bleau à la	a page s	uivante

c Carphone Carphone Claire Claire Container Grandma Miss-America	cible 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30	Débit Réel (Kbps) 30.21 45.34 60.42 75.45 80.51 30.49 45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	PSNR (dB) 31.54 33.28 34.60 35.68 36.47 38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00 24.67	0.89 0.91 0.92 0.94 0.95 0.96 0.97 0.98 0.88 0.90 0.91 0.92	Débit (Kbps) 30.16 43.14 56.37 67.78 71.51 28.21 36.39 43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42	PSNR Réel (dB) 31.69 33.19 34.35 35.22 35.69 38.90 40.06 40.91 42.03 42.19 34.54 36.91 37.86 38.93	SSIM 0.89 0.91 0.92 0.94 0.95 0.96 0.97 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.92	Débit (%) 0.1655 4.8522 6.703 10.1656 11.1787 7.4778 20.4242 28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Carphone (K Carphone (Claire (Claire (Container (Grandma (Miss-America (Kbps) 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30	Réel (Kbps) 30.21 45.34 60.42 75.45 80.51 30.49 45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	(dB) 31.54 33.28 34.60 35.68 36.47 38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00 24.67	0.89 0.91 0.92 0.94 0.95 0.95 0.96 0.97 0.97 0.97 0.98 0.88 0.88 0.90 0.91 0.92 0.93	 (Kbps) 30.16 43.14 56.37 67.78 71.51 28.21 36.39 43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42 	Réel (dB) 31.69 33.19 34.35 35.22 35.69 38.90 40.06 40.91 42.03 42.19 34.54 36.94 37.86 38.93	0.89 0.91 0.92 0.94 0.95 0.96 0.96 0.97 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.92	 (%) 0.1655 4.8522 6.703 10.1656 11.1787 7.4778 20.4242 28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Carphone (K Carphone (Claire (Claire (Container (Grandma (Miss-America (Kbps) 30 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 30	(Kbps) 30.21 45.34 60.42 75.45 80.51 30.49 45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	(dB) 31.54 33.28 34.60 35.68 36.47 38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00 24.67	0.89 0.91 0.92 0.94 0.95 0.95 0.96 0.97 0.97 0.98 0.88 0.88 0.90 0.91 0.92 0.93	(Kbps) 30.16 43.14 56.37 67.78 71.51 28.21 36.39 43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42	(dB) 31.69 33.19 34.35 35.22 35.69 38.90 40.06 40.91 42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.89 0.91 0.92 0.94 0.95 0.96 0.96 0.97 0.97 0.97 0.98 0.88 0.88 0.90 0.91 0.92 0.92	 (%) 0.1655 4.8522 6.703 10.1656 11.1787 7.4778 20.4242 28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Carphone Carphone Service Serv	30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 30 30	$\begin{array}{c} 30.21 \\ 45.34 \\ 60.42 \\ 75.45 \\ 80.51 \\ 30.49 \\ 45.73 \\ 61.03 \\ 75.89 \\ 81.02 \\ 30.12 \\ 45.19 \\ 60.13 \\ 75.15 \\ 80.22 \\ 30.18 \end{array}$	31.54 33.28 34.60 35.68 36.47 38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00 24.67	0.89 0.91 0.92 0.94 0.95 0.95 0.96 0.97 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.93	$\begin{array}{c} 30.16\\ 43.14\\ 56.37\\ 67.78\\ 71.51\\ 28.21\\ 36.39\\ 43.42\\ 56.08\\ 49.56\\ 28.97\\ 43.83\\ 57.86\\ 72.02\\ 76.42\\ \end{array}$	$\begin{array}{c} 31.69\\ 33.19\\ 34.35\\ 35.22\\ 35.69\\ 38.90\\ 40.06\\ 40.91\\ 42.03\\ 42.19\\ 34.54\\ 36.01\\ 36.94\\ 37.86\\ 38.93\\ \end{array}$	0.89 0.91 0.92 0.94 0.95 0.96 0.96 0.97 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.92	$\begin{array}{c} 0.1655\\ 4.8522\\ 6.703\\ 10.1656\\ 11.1787\\ 7.4778\\ 20.4242\\ 28.8546\\ 26.1035\\ 38.8299\\ 3.818\\ 3.0075\\ 3.7751\\ 4.165\\ \end{array}$
Claire Claire Grandma Miss-America	45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 30	45.34 60.42 75.45 80.51 30.49 45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	33.28 34.60 35.68 36.47 38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00	0.91 0.92 0.94 0.95 0.96 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.93	$\begin{array}{r} 43.14\\ 56.37\\ 67.78\\ 71.51\\ 28.21\\ 36.39\\ 43.42\\ 56.08\\ 49.56\\ 28.97\\ 43.83\\ 57.86\\ 72.02\\ 76.42\\ \end{array}$	33.19 34.35 35.22 35.69 38.90 40.06 40.91 42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.91 0.92 0.94 0.95 0.96 0.96 0.97 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.92	4.8522 6.703 10.1656 11.1787 7.4778 20.4242 28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Claire Container Grandma Miss-America	60 75 80 30 45 60 75 80 30 45 60 75 80 30	60.42 75.45 80.51 30.49 45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	34.60 35.68 36.47 38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00 24.67	0.92 0.94 0.95 0.95 0.96 0.97 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.93	56.37 67.78 71.51 28.21 36.39 43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42	34.35 35.22 35.69 38.90 40.06 40.91 42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.92 0.94 0.95 0.96 0.97 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.92	6.703 10.1656 11.1787 7.4778 20.4242 28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Claire Claire Container Grandma Miss-America	75 80 30 45 60 75 80 30 45 60 75 80 30	75.45 80.51 30.49 45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	35.68 36.47 38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00	0.94 0.95 0.96 0.97 0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.93	67.78 71.51 28.21 36.39 43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42	35.22 35.69 38.90 40.06 40.91 42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.94 0.95 0.96 0.97 0.97 0.98 0.98 0.88 0.90 0.91 0.92 0.92	10.1656 11.1787 7.4778 20.4242 28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Claire Claire Container Container	80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30 45 60 75 80 30	80.51 30.49 45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	36.47 38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00 24.67	0.95 0.95 0.96 0.97 0.98 0.88 0.90 0.91 0.92 0.93	71.51 28.21 36.39 43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42	35.69 38.90 40.06 40.91 42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.95 0.96 0.97 0.97 0.98 0.88 0.90 0.91 0.92	11.1787 7.4778 20.4242 28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Claire	30 45 60 75 80 30 45 60 75 80 30	30.49 45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	38.68 40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00 24.67	0.95 0.96 0.97 0.98 0.88 0.90 0.91 0.92 0.93	28.21 36.39 43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42	38.90 40.06 40.91 42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.96 0.97 0.97 0.98 0.88 0.90 0.91 0.92	7.4778 20.4242 28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Container Grandma Miss-America	45 60 75 80 30 45 60 75 80 30	45.73 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18	40.57 41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00	0.96 0.97 0.98 0.88 0.90 0.91 0.92 0.93	36.39 43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42	40.06 40.91 42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.96 0.97 0.98 0.88 0.90 0.91 0.92 0.92	$20.4242 \\28.8546 \\26.1035 \\38.8299 \\3.818 \\3.0095 \\3.7751 \\4.165 \\$
Container Grandma Miss-America	60 75 80 30 45 60 75 80 30	 61.03 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18 	41.91 43.06 43.83 34.62 36.09 37.20 38.01 39.00	0.97 0.98 0.88 0.90 0.91 0.92 0.93	43.42 56.08 49.56 28.97 43.83 57.86 72.02 76.42	40.91 42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.97 0.97 0.98 0.88 0.90 0.91 0.92 0.92	28.8546 26.1035 38.8299 3.818 3.0095 3.7751 4.165
Container Grandma Miss-America	75 80 30 45 60 75 80 30	 75.89 81.02 30.12 45.19 60.13 75.15 80.22 30.18 	43.06 43.83 34.62 36.09 37.20 38.01 39.00	0.97 0.98 0.88 0.90 0.91 0.92 0.93	56.08 49.56 28.97 43.83 57.86 72.02 76.42	42.03 42.19 34.54 36.01 36.94 37.86 38.93	0.97 0.98 0.88 0.90 0.91 0.92 0.02	26.1035 38.8299 3.818 3.0095 3.7751 4.165
Container Container	80 30 45 60 75 80 30	 81.02 30.12 45.19 60.13 75.15 80.22 30.18 	43.83 34.62 36.09 37.20 38.01 39.00	0.98 0.88 0.90 0.91 0.92 0.93	49.56 28.97 43.83 57.86 72.02 76.42	42.19 34.54 36.01 36.94 37.86 38.93	0.98 0.88 0.90 0.91 0.92	38.8299 3.818 3.0095 3.7751 4.165
Container Grandma Miss-America	30 45 60 75 80 30	30.12 45.19 60.13 75.15 80.22 30.18	34.62 36.09 37.20 38.01 39.00	0.88 0.90 0.91 0.92 0.93	28.97 43.83 57.86 72.02 76.42	34.54 36.01 36.94 37.86 38.93	0.88 0.90 0.91 0.92	3.818 3.0095 3.7751 4.165
Grandma Miss-America	45 60 75 80 30	45.19 60.13 75.15 80.22 30.18	36.09 37.20 38.01 39.00	0.90 0.91 0.92 0.93	43.83 57.86 72.02 76.42	36.01 36.94 37.86 38.93	0.90 0.91 0.92	3.0095 3.7751 4.165
Grandma Miss-America	60 75 80 30	60.13 75.15 80.22 30.18	37.20 38.01 39.00	0.91 0.92 0.93	57.86 72.02 76.42	36.94 37.86 38.93	0.91 0.92	3.7751 4.165
Grandma Miss-America	75 80 30	75.15 80.22 30.18	38.01 39.00	$0.92 \\ 0.93$	72.02 76.42	37.86 38.93	0.92	4.165
Grandma Miss-America	80 30	80.22 30.18	39.00	0.93	76.42	38.93	0.02	
Grandma Miss-America	30	30.18	34.67		10.42	00.00	0.95	4.7369
Miss-America			54.07	0.88	28.46	34.58	0.88	5.6991
Miss-America	45	45.21	36.09	0.92	41.76	35.85	0.92	7.631
Miss-America	60	60.38	37.14	0.93	55.44	36.91	0.92	8.1815
Miss-America	75	75.42	38.26	0.94	68.96	37.70	0.93	8.5653
Miss-America	80	80.41	39.77	0.96	74.01	39.54	0.96	7.9592
	30	30.28	39.18	0.93	28.78	39.23	0.93	4.9537
	45	45.43	40.80	0.95	42.85	40.56	0.95	5.679
	60	60.46	41.85	0.95	55.58	41.57	0.95	8.0714
	75	75.37	42.66	0.96	67.27	41.98	0.95	10.7469
	80	80.54	43.33	0.97	67.17	42.78	0.96	16.6004
Mother-Daughter	30	30.49	35.07	0.91	29.31	34.97	0.90	3.8701
	45	45.74	36.78	0.93	42.35	36.39	0.92	7.4114
	60	60.84	38.11	0.95	55.43	37.47	0.94	8.8921
	75	75.89	39.20	0.96	68.35	38.28	0.95	9.9354
	80	80.89	40.27	0.97	71.07	39.69	0.96	12.1399
News	30	30.27	31.70	0.90	30.13	31.89	0.90	0.4625
	45	45.38	33.21	0.92	42.26	33.39	0.92	6.8752
	60	60.45	34.99	0.94	53.68	34.51	0.93	11.1993
	75	75.43	36.24	0.95	67.22	35.57	0.94	10.8842
		80.40	38.08	0.97	71.23	37.29	0.96	11.4054

Suite du tableau précédent										
Séquences	$\mathbf{D}\mathbf{\acute{e}bit}$	JM15.0			QI	P Ajust	é	Gain		
	cible	Débit	PSNR	\mathbf{SSIM}	Débit	PSNR	\mathbf{SSIM}	Débit		
		Réel				Réel				
	(Kbps)	$({ m Kbps})$	(dB)		(Kbps)	(dB)		(%)		
Silent	30	30.18	31.07	0.84	29.77	30.89	0.84	1.3585		
	45	45.29	32.50	0.87	41.38	31.92	0.86	8.6332		
	60	60.28	33.77	0.90	52.64	32.78	0.88	12.6741		
	75	75.51	34.80	0.92	60.87	33.59	0.90	19.3881		
	80	80.45	36.74	0.95	65.24	35.44	0.94	18.9061		

• Évaluation de la qualité visuelle :

Les Figures 5.2, 5.3 et 5.4 représentent trois séquences : "Grandma", "Silent" et "Claire", codées respectivement à 60, 75 et 80Kbps par le JM15.0 et la technique proposée.



FIGURE 5.2 – Comparaison des trames reconstruites de la séquence "Grandma" codée à 60Kbps. La 25^{ème} trame : A- JM15.0 est appliqué (PSNR=38.57dB, 2016Bits/Pic); B- l'approche proposée est appliquée (PSNR=36.64dB, 464Bits/Pic). La 84^{ème} trame : C- JM15.0 est appliqué (PSNR=37.55dB, 1904Bits/Pic); D- l'approche proposée est appliquée (PSNR=38.65dB, 912Bits/Pic)

D'après la Figure 5.2, la qualité visuelle des deux trames reconstruites (Figure 5.2B et 5.2D de la séquence "Grandma" lorsque notre approche est appliquée, est similaire à celle des trames codées par le logiciel de référence JM15.0 (Figure 5.2A et 5.2C. Cette similarité de qualité visuelle est accompagnée d'une réduction en débit binaire (76.98% et 52.10%) et d'une variation en PSNR (-1.93dB et +1.1dB). Les mêmes conclusions peuvent être déduites à partir des Figures 5.3 et 5.4.

L'ajustement adaptatif du paramètre de quantification permet d'avoir la même qualité visuelle (*Figure 5.3B* et 5.3D et *Figure 5.4B* et 5.4D) malgré les chutes en PSNR. Les réductions en débit binaire sont respectivement égales à : 90.28% et 56.70% ($25^{\text{ème}}$ et $84^{\text{ème}}$ trames de la séquence "Silent"), et 76.61% et 69.94% ($18^{\text{ème}}$ et $68^{\text{ème}}$ trames de la séquence "Claire").



FIGURE 5.3 – Comparaison des trames reconstruites de la séquence "Silent"codée à 75Kbps. La 5^{ème} trame : A- JM15.0 est appliqué (PSNR=32.93dB, 2800Bits/Pic); B- l'approche proposée est appliquée (PSNR=32.93dB, 280Bits/Pic). La 62^{ème} trame : C- JM15.0 est appliqué (PSNR=34.93dB, 3344Bits/Pic); D- l'approche proposée est appliquée (PSNR=35.66dB, 1448Bits/Pic)

Pour améliorer davantage la performance de l'approche présentée dans ce paragraphe, nous l'avons combinée avec le modèle Intra présenté dans le *Chapitre 3*. Le but à atteindre est de réaliser des gains à la fois en débit et en qualité.

5.2.2 Performance de l'allocation de bits basée saillance visuelle

L'approche d'allocation de bits au niveau macrobloc P exploitant la notion de saillance, est combiné avec les modèles Intra et Inter R-Q (*Chapitre 3*) et l'approche d'allocation de bits effectuée au niveau trame (*Chapitre 4*). L'algorithme résultant nommé dorénavant "Algo_global", est comparé à l'algorithme de référence JM15.0.

• Différences des moyennes des débits binaires, PSNR et PSNR-HVS :

Le *Tableau 5.2*, récapitule l'ensemble des résultats obtenus avec les deux algorithmes (JM15.0 et "Algo_global").


FIGURE 5.4 – Comparaison des trames reconstruites de la séquence "Claire"codée à 80Kbps. La 18^{ème} trame : A- JM15.0 est appliqué (PSNR=43.36dB, 3728Bits/Pic); B- l'approche proposée est appliquée (PSNR=39.76dB, 872Bits/Pic). La 68^{ème} trame : C- JM15.0 est appliqué (PSNR=44.93dB, 3912Bits/Pic); D- l'approche proposée est appliquée (PSNR=43.50dB, 1176Bits/Pic)

TABLE 5.2 – Différences des moyennes des débits binaires, PSNR et PSNR-HVS entre le logiciel de référence JM15.0 et l'algorithme proposé "Algo_global".

Séquences	Débit	Différentes	Débit	Moyenne	Moyenne	Gain en	Gain en	Gain en
	binaire	Méthodes	Réel	PSNR	PSNR-HVS	PSNR	PSNR-HVS	$\mathbf{D}\mathbf{\acute{e}bit}$
	Cible							
	$({ m Kbps})$		$({ m Kbps})$	(dB)	(dB)	(dB)	(dB)	(Kbps)
Akiyo	15	JM15.0	15.14	34.14	33.22			
		Algo_global	15.09	35.46	35.37	1.32	2.15	0.05
	30	JM15.0	30.26	36.56	36.49			
		Algo_global	30.04	38.38	39.63	1.82	3.14	0.22
	45	JM15.0	45.37	38.40	39.16			
		Algo_global	44.93	40.04	42.16	1.64	3.00	0.44
	60	JM15.0	60.58	39.87	41.36			
		Algo_global	59.97	41.29	44.11	1.42	2.75	0.61
	75	JM15.0	75.46	41.05	43.07			
		Algo_global	74.79	42.03	44.99	0.98	1.92	0.67
		<u> </u>			Suite	du table	au à la page	suivante

5.2. ANALYSE DE LA PERFORMANCE DES APPROCHES PROPOSÉES

Séquences	Débit	Différentes	Débit	Moyenne	Moyenne	Gain en	Gain en	Gain en
	binaire	Méthodes	Réel	PSNR	PSNR-HVS	PSNR	PSNR-HVS	$\mathbf{D}\mathbf{\acute{e}bit}$
	Cible							
	(Kbps)		$({ m Kbps})$	(dB)	(dB)	(dB)	(dB)	(Kbps)
Bridge-Close	15	JM15.0	15.15	30.58	32.02			
		Algo_global	15.43	31.83	34.53	1.25	2.51	-0.28
	30	JM15.0	30.22	32.01	34.00			
		Algo_global	30.02	33.30	36.89	1.29	2.89	0.2
	45	JM15.0	45.39	33.05	35.61			
		Algo_global	45.23	34.33	38.70	1.28	3.09	0.16
	60	JM15.0	60.63	33.85	37.22			
		Algo_global	60.22	34.96	39.80	1.11	2.58	0.41
	75	JM15.0	75.80	34.41	38.02			
		Algo_global	75.50	35.44	40.56	1.03	2.54	0.3
Container	15	JM15.0	15.13	32.10	34.22			
		Algo_global	15.14	33.19	34.78	1.09	0.56	-0.01
	30	JM15.0	30.12	34.62	37.16			
		Algo_global	29.98	36.05	38.67	1.43	1.51	0.14
	45	JM15.0	45.19	36.09	38.94			
		Algo_global	45.02	37.49	40.07	1.4	1.13	0.17
	60	JM15.0	60.13	37.20	40.13			
		Algo_global	59.88	38.28	40.83	1.08	0.70	0.25
	75	JM15.0	75.15	38.01	41.08			
		Algo_global	74.86	38.85	41.53	0.84	0.45	0.29
Grandma	15	JM15.0	15.14	32.69	31.82			
		Algo_global	15.10	33.92	33.59	1.23	1.77	0.04
	30	JM15.0	30.18	34.67	34.50			
		Algo_global	30.03	36.25	37.27	1.58	2.77	0.15
	45	JM15.0	45.21	36.09	36.95			
		Algo_global	45.14	37.84	39.63	1.75	2.68	0.07
	60	JM15.0	60.38	37.14	38.59			
		Algo_global	59.99	38.86	41.18	1.72	2.59	0.39
	75	JM15.0	75.42	38.26	40.17			
		Algo_global	75.05	39.55	42.22	1.29	2.05	0.37
\mathbf{News}	15	JM15.0	15.89	29.61	28.50			
		Algo_global	15.99	29.61	28.50	0	0	-0.1
	30	JM15.0	30.27	31.70	32.34			
		Algo_global	30.07	33.17	33.21	1.47	0.87	0.2
	45	JM15.0	45.38	33.21	34.65			
					Suite	du table	au à la page	suivante

Suite du tableau précédent

Suite du ta	bleau pi	récédent						
Séquences	Débit	Différentes	Débit	Moyenne	Moyenne	Gain en	Gain en	Gain en
	binaire	Méthodes	Réel	PSNR	PSNR-HVS	PSNR	PSNR-HVS	$\mathbf{D}\mathbf{\acute{e}bit}$
	Cible							
	$({ m Kbps})$		(Kbps)	(dB)	(dB)	(dB)	(dB)	(Kbps)
		Algo_global	45.13	35.26	36.59	2.05	1.94	0.25
	60	JM15.0	60.45	34.99	37.09			
		Algo_global	60.05	36.70	39.10	1.71	2.01	0.4
	75	JM15.0	75.43	36.24	38.70			
		Algo_global	74.76	37.68	40.51	1.44	1.81	0.67
Salesman	15	JM15.0	15.12	28.84	26.89			
		Algo_global	15.14	28.84	26.84	0	-0.05	-0.02
	30	JM15.0	30.19	30.98	30.55			
		Algo_global	30.06	32.12	31.47	1.14	0.92	0.13
	45	JM15.0	45.35	32.60	33.12			
		Algo_global	45.09	34.23	34.77	1.63	1.65	0.26
	60	JM15.0	60.55	33.98	35.25			
		Algo_global	59.98	35.72	37.37	1.74	2.12	0.57
	75	JM15.0	75.34	35.09	37.05			
		Algo_global	74.97	36.75	39.14	1.66	2.09	0.37
Silent	15	JM15.0	16.46	29.18	26.68			
		Algo_global	15.16	28.56	25.32	-0.62	-1.36	1.3
	30	JM15.0	30.18	31.07	29.65			
		Algo_global	30.14	32.01	30.21	0.94	0.56	0.04
	45	JM15.0	45.29	32.50	31.97			
		Algo_global	45.08	33.98	33.08	1.48	1.11	0.21
	60	JM15.0	60.28	33.77	34.05			
		Algo_global	59.96	35.35	35.25	1.58	1.20	0.32
	75	JM15.0	75.51	34.80	35.58			
		Algo_global	74.57	36.30	36.88	1.5	1.30	0.94

D'après le Tableau 5.2, l'algorithme proposé présente de meilleurs résultats en termes de qualité et de débit binaire en comparaison avec le JM15.0. Par rapport à ce dernier, les gains en débit sont significatifs (Jusqu'à +0.94Kbps) pour des débits binaires variant entre 15 et 75Kbps. De plus, les débits réels obtenus par l'approche "Algo_global" sont dans la majorité des cas inférieurs ou proches des débits cibles. Ces gains en débits illustrent aussi bien l'efficacité du processus d'allocation de bits au niveau trame, que celui au niveau macrobloc. Les moyennes du PSNR et du PSNR-HVS sont également importantes pour presque toute la gamme des débits binaires cibles choisie. Les gains PSNR atteignent +2.05dB (Séquence "News" codée à 45Kbps) et ceux du PSNR-HVS s'élèvent à +3.14dB (Séquence "Akiyo" codée à 30Kbps). Ces gains qui génèrent une meilleure qualité visuelle des séquences codées, sont induits par une répartition qui s'adapte bien à la saillance du macrobloc.

• Évaluation de la qualité visuelle

Les *Figures 5.5*, *5.6* et *5.7* illustrent l'amélioration de la qualité visuelle en termes de préservation de détails, de trois séquences reconstruites :"Akiyo", "News" et "Mother-Daughter", codées respectivement à 30, 45 et 60Kbps par le JM15.0 et la technique proposée ("Algo_golbal").



FIGURE 5.5 – La 6^{ème} trame de la séquence "Akiyo" codée à 30Kbps. (Gauche) Trame Originale, (Milieu) Notre approche proposée et (Droit) JM15.0.



FIGURE 5.6 – La $10^{\text{ème}}$ trame de la séquence "News" codée à 45Kbps. (Gauche) Trame Originale, (Milieu) Notre approche proposée et (Droit) JM15.0.

Sur la Figure 5.5, la 6^{ème} trame de la séquence "Akiyo" est illustrée. La qualité visuelle de la trame est nettement améliorée lorsque l'approche "Algo_global" est employée. Les altérations perçues au niveau du visage, du cou et sur la veste de la présentatrice, causées par le JM15.0 (Figure 5.5C) et qui nuisent à la qualité sont totalement absentes avec l'approche proposée (Figure 5.5B). Les mêmes observations peuvent être formulée pour les séquences "News" et "Mother-Daughter".



FIGURE 5.7 – La 90^{ème} trame de la séquence "Mother-Daughter" codée à 60Kbps. (Gauche) Trame Originale, (Milieu) Notre approche proposée et (Droit) JM15.0.

5.2.3 Performance de l'approche d'ajustement des QPs des macroblocs I et P

L'approche d'ajustement du paramètre de quantification combiné avec le modèle Intra R-Q nommée dorénavant : "RQ $+QP_{Ajust}$ " est comparée avec l'algorithme de référence JM15.0.

• Calcul des ratios de débit binaire et des moyennes du PSNR, PSNR-HVS et SSIM :

Le Tableau 5.3 donne les gains en termes de débit binaire et de qualité, des deux implémentations (JM15.0 et "RQ+ QP_{Ajust} ").

TABLE 5.3 – Comparaison des gains en termes de débit binaire, PSNR, PSNR-HVS et SSIM entre le logiciel de référence JM15.0 et l'algorithme proposé "RQ+ $QP_{Ajust\acutee}$ ".

Séquences	Débit	Différentes	Débit	Moyenne	Moyenne	Moyenne	Gain en	Gain en	Gain en
	binaire	Méthodes	Réel	PSNR	PSNR-HVS	SSIM	débit	PSNR	PSNR-HVS
	Cible								
	(Kbps)		(Kbps)	(dB)	(dB)		(%)	(dB)	(dB)
Akiyo	30	JM15.0	30.26	36.56	36.53	0.93			
		$RQ+QP_{Ajust}$	27.81	38.38	39.69	0.96	8.0964	1.82	3.16
	45	JM15.0	45.37	38.40	39.18	0.95			
		$RQ+QP_{Ajust}$	39.03	40.00	42.18	0.97	13.9739	1.6	3.00
	60	JM15.0	60.58	39.87	41.36	0.96			
		$RQ+QP_{Ajust}$	49.77	41.00	43.57	0.97	17.8441	1.13	2.21
	75	JM15.0	75.46	41.05	43.07	0.97			
		$RQ+QP_{Ajust}$	61.66	41.89	44.94	0.98	18.2878	0.84	1.87
Bridge-Close	30	JM15.0	30.22	32.01	34.00	0.86			
						Suite	du table	au à la p	oage suivante

Séquences	Débit	Différentes	Débit	Moyenne	Moyenne	Moyenne	Gain en	Gain en	Gain en
	binaire	Méthodes	Réel	PSNR	PSNR	SSIM	débit	PSNR	PSNR-HVS
	Cible								
	(Kbps)		$({ m Kbps})$	(dB)	(dB)		(%)	(dB)	(dB)
		$RQ+QP_{Ajust}$	30.22	33.35	36.91	0.91	0	1.34	2.91
	45	JM15.0	45.39	33.05	35.61	0.88			
		$RQ+QP_{Ajust}$	45.26	34.42	38.88	0.93	0.2864	1.37	3.27
	60	JM15.0	60.63	33.85	37.22	0.9			
		$RQ+QP_{Ajust}$	60.03	34.95	39.87	0.94	0.9896	1.1	2.65
	75	JM15.0	75.80	34.41	38.02	0.91			
		$RQ+QP_{Ajust}$	73.56	35.39	40.45	0.94	2.9551	0.98	2.43
Container	30	JM15.0	30.18	34.67	37.16	0.88			
		$RQ+QP_{Ajust}$	29.07	36.47	38.89	0.90	3.6779	1.8	1.73
	45	JM15.0	45.19	36.09	38.94	0.90			
		$RQ+QP_{Ajust}$	43.13	37.36	40.24	0.91	4.5585	1.27	1.30
	60	JM15.0	60.13	37.20	40.13	0.91			
		$RQ+QP_{Ajust}$	57.67	38.21	41.24	0.92	4.0911	1.01	1.11
	75	JM15.0	75.15	38.01	41.08	0.92			
		$RQ+QP_{Ajust}$	72.65	38.86	42.08	0.93	3.3266	0.85	1.00
Grandma	30	JM15.0	30.18	34.67	34.50	0.88			
		$RQ+QP_{Ajust}$	29.07	36.47	37.96	0.93	3.6779	1.8	3.46
	45	JM15.0	45.21	36.09	36.95	0.92			
		$RQ+QP_{Ajust}$	42.46	38.09	40.63	0.95	6.0827	2	3.68
	60	JM15.0	60.38	37.14	38.59	0.93			
		$RQ+QP_{Ajust}$	55.70	39.01	42.11	0.96	7.7509	1.87	3.52
	75	JM15.0	75.42	38.26	40.17	0.94			
		$RQ+QP_{Ajust}$	68.98	39.61	43.20	0.96	8.5388	1.35	3.03
News	30	JM15.0	30.27	31.70	32.34	0.90			
		$RQ+QP_{Ajust}$	30.21	33.46	33.60	0.93	0.1982	1.76	1.26
	45	JM15.0	45.38	33.21	34.65	0.92			
		$RQ+QP_{Ajust}$	43.57	35.26	36.34	0.95	3.9885	2.05	1.69
	60	JM15.0	60.45	34.99	37.09	0.94			
		$RQ+QP_{Ajust}$	54.95	36.18	37.71	0.96	9.0984	1.19	0.62
	75	JM15.0	75.43	36.24	38.70	0.95			
		$RQ+QP_{Ajust}$	65.86	37.08	39.09	0.96	12.6872	0.84	0.39
Salesman	30	JM15.0	30.19	30.98	30.55	0.85			
		$RQ+QP_{Ajust}$	30.16	32.40	32.03	0.91	0.0993	1.42	1.48
	45	JM15.0	45.35	32.60	33.13	0.89			
		$RQ+QP_{Ajust}$	44.33	34.83	36.29	0.94	2.2491	2.23	3.16

Suite du ta	Suite du tableau précédent								
Séquences	$\mathbf{D}\mathbf{\acute{e}bit}$	Différentes	Débit	Moyenne	Moyenne	Moyenne	Gain en	Gain en	Gain en
	binaire	Méthodes	Réel	PSNR	PSNR	SSIM	débit	PSNR	PSNR-HVS
	Cible								
	(Kbps)		(Kbps)	(dB)	(dB)		(%)	(dB)	(dB)
	60	JM15.0	60.55	33.98	35.25	0.91			
		$RQ+QP_{Ajust}$	58.31	36.22	38.61	0.96	3.6994	2.24	3.36
	75	JM15.0	75.34	35.09	37.05	0.93			
		$RQ+QP_{Ajust}$	72.76	37.10	40.40	0.96	3.4244	2.01	3.35
Silent	30	JM15.0	30.18	31.07	29.65	0.84			
		$RQ+QP_{Ajust}$	30.23	32.09	30.38	0.88	-0.1656	1.02	0.73
	45	JM15.0	45.29	32.50	31.97	0.87			
		$RQ+QP_{Ajust}$	42.57	33.71	32.60	0.91	6.0057	1.21	0.63
	60	JM15.0	60.28	33.77	34.05	0.90			
		$RQ+QP_{Ajust}$	52.99	34.77	34.24	0.93	12.0935	1	0.19
	75	JM15.0	75.51	34.80	35.15	0.92			
		$RQ+QP_{Ajust}$	62.08	35.34	35.58	0.94	17.7857	0.54	0.43

Nous remarquons que la technique proposée améliore d'une manière très significative le débit binaire (jusqu'à 18% de réduction) et la qualité (+2.24dB en PSNR, +3.68dB en PSNR-HVS et 5% en SSIM). La réduction du débit binaire est obtenue grâce à l'utilisation de valeurs larges du paramètre de quantification pour coder les macroblocs non importants visuellement. Les gains en qualité proviennent de l'utilisation du modèle Intra R-Q qui permet de choisir efficacement le paramètre de quantification initial pour coder les deux premières trames ce qui améliore la qualité de la totalité de la séquence.

5.2.4 Performance du MAD utilisant les propriétés du SVH

Dans cette section, nous jugeons la performance du MAD pondéré par la CSF en utilisant une mesure d'évaluation subjective. Cette dernière consiste à mener une campagne de tests subjectifs faisant appel à un panel d'observateurs. Ces tests doivent impérativement se dérouler dans une salle équipée spécialement à cet effet (*Figure 5.8*) et suivre des protocoles normalisés dans le but d'assurer une fiabilité des résultats et une reproductibilité des campagnes et des paradigmes. Les méthodes d'évaluation subjective de la qualité restent à l'heure actuelle les seules solutions unanimement reconnues pour déterminer de façon fiable la qualité perçue d'une image ou d'une vidéo.

Ces méthodes comportent deux phases :

- Une phase expérimentale : présentation d'images ou de séquences vidéo à des groupes d'observateurs qui doivent juger la qualité perçue sur une échelle discrète ou continue.
- Une phase analytique : traitement statistique des données recueillies pour obtenir la note de qualité, appelée Mean Opinion Score (MOS), de l'image ou de la vidéo. L'obtention de



FIGURE 5.8 – Salle psycho-visuelle 4/E209 du laboratoire SIC.

cette note permet par la suite, soit d'optimiser les métriques objectives en servant de base d'apprentissage, soit d'évaluer les performances des métriques objectives en comparant les notes prédites par le modèle avec les MOS.

a) Conditions d'expérience

Effectuer des tests subjectifs demande de suivre des protocoles bien spécifiques décrits dans les normes sous forme de recommandations de l'Union Internationale des Télécommunications (ITU). Dans ces recommandations, sont spécifiées les conditions d'observation (normalisation des environnements de tests), la méthodologie à suivre, les caractéristiques des stimuli à utiliser et la manière dont sont sélectionnés les observateurs [11]. Dans ces conditions, l'observateur visualise le test sans aucune gêne due à son environnement. L'intérêt de normaliser les conditions d'observations et les protocoles est de réduire tout biais. Les notes ainsi obtenues sont censées être stables et peuvent être comparées, échangées et exploitées par d'autres études. Ces recommandations sont largement utilisées par l'ensemble des laboratoires de recherche car elles garantissent l'inter-connectivité des résultats.

b) Choix des observateurs

Il est recommandé d'avoir un panel d'observateurs le plus large possible, l'ITU préconise un groupe constitué d'au moins 15 observateurs pour valider un test. Ces derniers pouvant être soit experts ou de préférence novices dans le domaine de la campagne d'évaluation. Pour que les personnes puissent cependant passer l'évaluation, il faut s'assurer qu'elles n'aient pas de gêne visuelle majeure et répondent aux critères suivants :

- Une vision normale des couleurs, vérifiée avec le test d'Ishihara ou équivalent.
- Une acuité visuelle normale ou rendu normale par correction (i.e. 10/10 aux deux yeux avec ou sans lunettes ou lentilles) cette acuité sera vérifiée avec le test de Snellen.
- Une sensibilité au contraste normale vérifiée par Pelli-Robson ou l'équivalent.

Le panel d'observateurs choisi dépend à la fois des types de traitement à évaluer et de la nature de l'étude à mener. Il est préférable que le groupe d'observateurs représente une distribution

statistique de la population. On peut ainsi être amené à prendre en compte, lors de la sélection de ce panel, plusieurs critères comme l'âge et le sexe mais aussi l'origine ethnique, socioculturelle, professionnelle, ... etc.

c) Séance d'évaluation

Si l'on veut qu'une séance soit efficace, elle ne doit pas excéder 30 minutes, en incluant les explications et les exemples. Au-delà de ce temps, l'observateur commence à fatiguer et/ou s'adapter, son jugement ne sera plus fiable ce qui pourrait rendre ses résultats inutilisables. En règle générale, une séance de tests débute par une description de la campagne et de l'objectif de cette évaluation. Ces différentes explications permettent une meilleure implication de l'observateur, en étant directement concerné, il sera plus attentif et donnera le meilleur de son jugement durant le test. La négligence de cette phase peut porter préjudice aux résultats. Les explications d'ordre technique notamment celles concernant les échelles de notation, les protocoles de présentation et de vote ne sont expliqués que dans la suite de la séance d'évaluation. Comme nous l'avons déjà évoqué précédemment, nous effectuons dans un premier temps une séquence d'essais avec chaque participant (i.e. exemples typiques d'images dégradées et leurs notes de qualité respectives, généralement 4 ou 5 sont suffisants) afin de vérifier si le sujet a bien compris les recommandations. Après cette séquence d'essais, nous répondons aux éventuelles questions ou remarques de l'observateur afin d'éviter toute incompréhension du travail d'évaluation à faire. Dans un second temps, une séquence dite de fausses présentations permet à l'observateur de stabiliser son jugement. Ce n'est seulement qu'après ces deux séquences que le test commence, l'expérimentateur se retire de la salle psychophysique pour ne pas perturber l'observateur lors des présentations. Il convient de choisir des stimuli de sorte que toute la plage de notation soit utilisée par une majorité des observateurs. Entre chaque présentation, un temps de pause (affichage d'un gris moyen) permet d'inhiber la mémoire courte de l'observateur et d'éviter toute confusion avec l'image / séquence d'images suivante.

d) Mesure d'évaluation subjective

Nous avons choisi la méthode à double stimulis (DSIS). Cette dernière est cyclique en ce sens que l'on propose à l'observateur une paire de stimuli affichés successivement, la référence étant toujours présentée avant le stimulus à évaluer. L'échelle de notation est discrète à 5 niveaux, elle va de "très mauvais" à "très bonne".

Cette méthode est très largement utilisée lorsque l'on peut détecter facilement les artefacts causés par la compression/transmission sur les stimuli à évaluer. Les observateurs seront priés de regarder l'image pendant toute la durée de T1 et T3. Le vote sera autorisé uniquement pendant T4.

e) Résultats de simulation subjective

Le Tableau 5.4 illustre les résultats de simulations obtenus après avoir implémenté le MAD pondéré par la CSF.



FIGURE 5.9 – Méthode DSIS.

TABLE 5.4 – Comparaison entre l'algorithme de référence JM15.0 et l'approche utilisant le MAD_{CSF} en termes de différence des débits binaires et des scores subjectifs.

Séquences	Débit	Différentes	Débit	Gain de	Score	Gain
	Binaire	Méthodes	\mathbf{R} éel	débit	DSIS	MOS
	Cible				(MOS)	
	(Kbps)		$({ m Kbps})$	(Kbps)		
Akiyo	15	JM15.0	15.14	+0.02	3.07	-0.14
		MAD_{CSF} +JM10.2	15.17		2.93	
	20	JM15.0	20.17	0	3.07	+0.64
		MAD_{CSF} +JM10.2	20.17		3.71	
	30	JM15.0	30.26	+0.04	3.57	+0.79
		MAD_{CSF} +JM10.2	30.22		4.36	
	45	JM15.0	45.37	+0.07	4.43	-0.22
		MAD_{CSF} +JM10.2	45.30		4.21	
	80	JM15.0	80.54	+3.06	4.71	0
		MAD_{CSF} +JM10.2	77.48		4.71	
Bridge-Close	15	JM15.0	15.15	-0.02	2.5	+0.14
		MAD_{CSF} +JM10.2	15.17		2.64	
	20	JM15.0	20.21	-0.03	2.71	+0.22
		MAD_{CSF} +JM10.2	20.24		2.93	
	30	JM15.0	30.28	+0.06	3.07	+0.07
		MAD_{CSF} +JM10.2	30.24		3.14	
	45	JM15.0	45.39	+0.08	3.43	+0.07
		MAD_{CSF} +JM10.2	45.31		3.5	
	80	JM15.0	80.55	-0.04	4	+0.36
		Suite o	lu table	eau à la p	bage su	ivante

Séquences	Débit	Différentes	$\mathbf{D}\mathbf{\acute{e}bit}$	Gain de	Score	Gair
	Binaire	Méthodes	Réel	débit	DSIS	MOS
	Cible				(MOS)	
	(Kbps)		(Kbps)	(Kbps)		
		MAD_{CSF} +JM10.2	80.59		4.36	
Claire	15	JM15.0	15.30	0	3	+0.2
		MAD_{CSF} +JM10.2	15.30		3.29	
	20	JM15.0	20.38	+0.06	3.93	-0.29
		MAD_{CSF} +JM10.2	20.32		3.64	
	30	JM15.0	30.53	+0.18	4	-0.14
		MAD_{CSF} +JM10.2	30.35		3.86	
	45	JM15.0	45.73	+5.08	4.36	0
		MAD_{CSF} +JM10.2	40.65		4.36	
	80	JM15.0	81.02	+10.63	4.14	0
		MAD_{CSF} +JM10.2	70.39		4.14	
Container	15	JM15.0	15.13	+0.02	3.57	+0.1
		MAD_{CSF} +JM10.2	15.11		3.71	
	20	JM15.0	20.13	-0.09	3.64	-0.14
		MAD_{CSF} +JM10.2	20.22		3.5	
	30	JM15.0	30.12	-0.02	3.71	0
		MAD_{CSF} +JM10.2	30.14		3.71	
	45	JM15.0	45.19	+0.04	4	+0.1
		MAD_{CSF} +JM10.2	45.15		4.14	
	80	JM15.0	80.22	+0.04	4.21	+0.0
		MAD_{CSF} +JM10.2	80.18		4.29	
Grandma	15	JM15.0	15.14	+0.01	2.71	+0.0
		MAD_{CSF} +JM10.2	15.13		2.79	
	20	JM15.0	20.23	+0.09	3.36	-0.1
		MAD_{CSF} +JM10.2	20.14		3.21	
	30	JM15.0	30.18	+0.01	3.79	-0.1
		MAD_{CSF} +JM10.2	30.17		3.69	
	45	JM15.0	45.26	+0.07	3.57	0
		MAD_{CSF} +JM10.2	45.19		3.57	
	80	JM15.0	80.41	+0.36	4.64	+0.0
		MAD_{CSF} +JM10.2	80.05		4.71	
/liss-America	15	JM15.0	15.25	+0.09	2.21	+0.2
		MAD_{CSF} +JM10.2	15.16		2.43	
	20	JM15.0	20.23	+0.07	2.79	+0.1
		MAD_{CSF} +JM10.2	20.16		2.93	
		Suite o	lu table	au à la 1	bage sur	ivant

5.3. CONCLUSION

Séquences	Débit	Différentes	Débit	Gain de	Score	Gain
	Binaire	Méthodes	Réel	débit	DSIS	MOS
	Cible				(MOS)	
	(Kbps)		(Kbps)	(Kbps)		
	30	JM15.0	30.29	+0.12	3.57	-0.14
		MAD_{CSF} +JM10.2	30.17		3.43	
	45	JM15.0	45.40	+0.34	3.5	+0.29
		MAD_{CSF} +JM10.2	45.06		3.79	
	80	JM15.0	80.54	+0.77	4.79	0
		MAD_{CSF} +JM10.2	79.77		4.79	
Salesman	15	JM15.0	15.12	0	1.64	-0.21
		MAD_{CSF} +JM10.2	15.12		1.43	
	20	JM15.0	20.14	+0.01	2	0
		MAD_{CSF} +JM10.2	20.13		2	
	30	JM15.0	30.19	-0.04	2.43	+0.36
		MAD_{CSF} +JM10.2	30.23		2.79	
	45	JM15.0	45.32	+0.11	3.36	+0.14
		MAD_{CSF} +JM10.2	45.21		3.5	
	80	JM15.0	80.24	+0.03	4.64	+0.21
		MAD_{CSF} +JM10.2	80.21		4.43	

Nous rappelons que plus le MOS est élevé, meilleure est la qualité visuelle. Les MOS obtenues sont généralement supérieures lorsque l'approche de pondération du MAD par la CSF est adoptée. Cette approche est visuellement plus efficace puisqu'elle génère un gain subjectif allant jusqu'à +0.64 en comparaison avec l'algorithme de référence JM15.0.

5.3 Conclusion

Ce Chapitre présente un ensemble d'approches exploitant la notion de carte de saillance. Elles sont appliquées au niveau macrobloc de l'algorithme du contrôle de débit. La première contribution consistait à adapter le paramètre de quantification d'un macrobloc P donné selon son importance visuelle. Cette dernière est déterminée en fonction du nombre de ses pixels importants et non importants. La notion de saillance a été également intégrée au niveau du processus d'allocation de bits utilisé pour estimer les bits pour coder un macrobloc. A base de la carte d'attention visuelle, une classification des macroblocs Inter en saillants et non saillants a été effectuée. Cette classification a été employée pour ajuster minutieusement le budget habituel calculé par l'algorithme du contrôle de débit adopté par H.264/AVC. Les macroblocs saillants ont été codés avec plus de bits que ceux déterminées par l'algorithme du H.264/AVC. Ce supplément de bits a été déduit des macroblocs non saillants codés avec moins de bits. Dans une troisième approche, nous avons combiné le modèle Intra proposé dans le *Chapitre 3* avec la technique d'ajustement

du paramètre de quantification à base de saillance. L'idée est d'ajuster efficacement le paramètre de quantification pour tous types de macroblocs I et P. Finalement, nous avons amélioré la performance de la mesure MAD en exploitant certaines propriétés du Système Visuel Humain. Les approches proposées ont permis d'obtenir trois types de résultats :

- Réduction significative du débit binaire (38%) sans aucune dégradation de la qualité visuelle (Même SSIM).
- \circ Amélioration de la qualité visuelle (+2.05 dB en PSNR et +3.14dB en PSNR-HVS) tout en obtenant un débit binaire réel inférieur ou égal au débit cible.
- Réduction significative du débit binaire (18%) et amélioration de la qualité visuelle (+2.24 dB en PSNR et +3.68dB en PSNR-HVS et 5% en SSIM).

Ces résultats prouvent l'efficacité des approches proposées en comparaison avec l'algorithme du contrôle de débit adopté par H.264/AVC. Ils montrent également que l'intégration de la carte de saillance est bénéfique que cela soit au niveau du processus d'allocation de bits ou au niveau du détermination du paramètre de quantification.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

L'objectif principal de cette thèse était de concevoir et de développer de nouvelles techniques visant à améliorer les performances du contrôle de débit de la dernière née des normes de compression qu'est H.264/AVC. Cette norme est certes plus efficace que les normes précédentes mais souffre néanmoins au niveau du contrôle de débit, de certaines insuffisances. Chercher à pallier ces limitations présente donc un intérêt certain pour la multitude d'applications visée par cette norme.

Pour répondre au mieux à l'objectif de cette thèse, nous avons orienté notre travail vers trois directions. La première direction, plutôt expérimentale, visait à déterminer d'une manière plus appropriée les paramètres de quantification utilisés par les unités de codage (trames ou macroblocs) I et P. La deuxième direction visait à proposer une allocation de bits plus fine pour les trames P en exploitant et les propriétés statistiques du signal d'images et les modèles d'attention visuelle. La troisième direction enfin s'intéressait à l'intégration au niveau macrobloc des cartes d'attention visuelle au sein du processus du contrôle de débit.

• Contributions majeures :

La première orientation a été présentée dans le *Chapitre 3*. Les *Chapitres 1* et 2 ont été dédiés à la description des algorithmes du contrôle de débit des différents standards de codage vidéo. Ils se sont focalisés ensuite sur celui de la norme H.264/AVC pour mettre en évidence des insuffisances importantes liées essentiellement : à la détermination du paramètre de quantification initial pour coder la trame I et la première trame P, à l'utilisation du *MAD* comme mesure de complexité et son modèle de prédiction et au processus d'allocation de bits au niveau d'une unité de codage.

En exploitant ces insuffisances, le *Chapitre 3* a présenté une approche de modélisation Débit-Quantification (R-Q) permettant de déterminer efficacement les paramètres de quantification des unités de base. Deux modèles ont été proposés. Le premier, dédié aux trames I, permet d'estimer de manière appropriée le paramètre de quantification initial. Ce modèle a été élaboré à partir d'un ensemble de tests expérimentaux. Il a permis d'établir une relation non linéaire entre le débit global pour coder la totalité de la séquence, le débit cible et la complexité de la trame I. Le particularité de ce modèle réside dans la prise en considération de la mesure de complexité de la trame I sans avoir à la calculer (Contrairement à la totalité des études proposées dans la littérature). Le second modèle R-Q est destiné à la détermination du paramètre de quantification d'une unité de base P. Basé également sur une étude expérimentale, ce modèle est logarithmique et il relie le nombre de bits cible au paramètre de quantification d'une unité P. La particularité de ce second modèle est qu'il permet de remplacer les deux modèles linéaire et quadratique de l'algorithme du contrôle de débit de H.264/AVC. L'intégration de ces deux modèles dans le codage de plusieurs séquences vidéo à des débits bas variant de 10 à 80Kbps a permis, outre une réduction de la complexité calculatoire, un gain significatif et simultané en qualité et en débit binaire.

La deuxième orientation, traitant l'allocation de bits au niveau trame, a été abordée dans le Chapitre 4. L'allocation de bits au niveau trame du profil de base du standard H.264/AVC reste basique. Pour l'affiner, deux approches ont été proposées. La première, basée sur les caractéristiques du signal image, exploite deux mesures de complexité. La première est un rapport de mouvement déterminé à partir de bits réels utilisés pour coder les trames précédentes. La seconde mesure exploite la différence entre trames adjacentes et l'histogramme de cette différence. La deuxième approche décrite dans ce chapitre, consiste à exploiter la notion de carte de saillance pour ajuster le paramètre de quantification au niveau trame P. Un modèle d'attention visuelle basé sur une approche "Bottom-Up", permettant de détecter les zones saillantes est utilisé. Les tests expérimentaux ont prouvé l'efficacité des deux approches proposées en comparaison avec les algorithmes JM15.0 et JVT-O016. L'approche d'allocation de bits à base de mesure de complexité a permis d'améliorer la qualité visuelle de plus de 2dB tout en fournissant un débit réel inférieur ou égal au débit cible de la bande passante du canal. L'ajustement du paramètre de quantification à base de carte de saillance a réduit considérablement le débit binaire jusqu'à 26%, tout en conservant la même qualité visuelle que celle fournie par l'algorithme du contrôle de débit de H.264/AVC.

La dernière orientation enfin a été décrite dans le Chapitre 5. Elle visait, compte tenu de l'importance des gains obtenus avec l'utilisation des modèles d'attention visuelle au niveau trame, à étendre le principe au niveau macrobloc. La carte de saillance a ainsi été exploitée et dans l'ajustement du paramètre de quantification de chaque macrobloc P et dans le processus d'allocation de bits. La paramètre de quantification a été ajusté par un facteur de pondération déterminé à partir du nombre des pixels saillants dans le macrobloc considéré. Pour le processus d'allocation de bits, le nombre de bits cible de chaque macrobloc a été adapté selon son importance visuelle (définie par la carte de saillance). Les macroblocs saillants ont alors été codés avec un nombre de bits toujours plus grand que celui utilisé par H.264/AVC. Le surplus de bits ainsi induit, est déduit du budget de bits alloué aux macroblocs non saillants. Les résultats obtenus là aussi militent fortement pour une exploitation des modèles d'attention visuelle dans le contrôle de débit; un gain de plus de 3dB sous la contrainte de débit cible et une réduction du débit allant jusqu'à 38% sous la contrainte d'une qualité vidéo similaire à celle de H.264/AVC. Quand le modèle Intra R-Q permettant de déterminer le paramètre de quantification initial est combiné avec l'approche d'ajustement du paramètre de quantification d'un macrobloc P, le gain est à la fois en qualité (+3dB) et en débit (réduction allant jusqu'à 18%).

• Perspectives :

Les perspectives sont multiples et relatives aux différentes approches abordées et au sujet traité dans cette thèse.

- 1 : Les évaluations expérimentales des approches proposées dans cette thèse ont ciblé principalement les vidéos en format QCIF et ont permis une modélisation Débit-Qantification plus fine. La même démarche pourrait être envisagée pour la compression de la vidéo HD, une des applications phare de H.264/AVC.
- 2 : Dans les approches d'allocation de bits proposées, nous nous sommes intéressés à l'amélioration de la répartition du débit binaire cible en tenant compte de l'importance visuelle de l'unité de base traitée. Dans le codeur H.264/AVC, le flux binaire est principalement composé de bits utilisés pour le codage de la texture plus les bits d'entête incluant le mode de codage d'un macrobloc, les vecteurs de mouvements et les paramètres de quantification. La majorité des algorithmes du contrôle de débit considèrent que le budget de bits d'entête est toujours fixe. Cependant, la structure compliquée du codeur H.264/AVC peut conduire à un changement important du nombre de bits nécessaire pour coder les informations secondaires. Adopter la méthode conventionnelle du nombre de bits d'entête fixe ne permet pas de réaliser une prédiction précise. Toutes ces observations militent fortement pour une étude permettant de proposer un modèle capable de prédire les bits d'entête de l'unité de codage courante en fonction de ses caractéristiques spatio-temporelles et de l'historique des unités codées précédemment.
- 3 : L'un des objectifs hautement souhaitable d'un algorithme du contrôle de débit est l'obtention d' une qualité consistante tout au long la séquence vidéo traitée. Pour éviter les grands changements des paramètres de quantification de deux unités de base voisines, l'algorithme du contrôle de débit adopté par H.264/AVC introduit un limiteur (ΔQP) égal à ±2 permettant de garantir le lissage de qualité à travers les trames voisines (ou macroblocs voisins). Ce limiteur peut introduire une importante variation de la qualité visuelle après le codage de plusieurs unités de base. Dans l'une de nos approches, nous avons utilisé une carte d'attention visuelle pour nous informer sur les trames ou macroblocs saillants. Cette information qui a été employée pour adapter le paramètre de quantification, peut aussi être exploitée pour pondérer le limiteur ΔQP afin de réduire davantage les variations perceptibles de la qualité vidéo et garantir la stabilité.
- 4 : L'analyse, l'optimisation et le contrôle de la relation débit-distorsion (R-D) pour améliorer les performances d'un codeur vidéo, représentent des issues difficiles puisque les vidéo sources sont à contenus différents et à caractéristiques non uniques. De plus, les différentes structures du codage vidéo entraînent des difficultés à la procédure d'analyse, d'optimisation et du contrôle débit-distorsion. Il est crucial de comprendre que la relation entre le débit et la distorsion d'un signal vidéo compressé ne dépend pas uniquement des caractéristiques inhérentes à la vidéo source, mais dépend également de la structure des codeurs eux-mêmes puisque chacun d'eux possède des capacités de compression différentes. Ainsi, l'exploitation de la structure hybride du codeur peut constituer aussi un champ très fertile pour davantage améliorer les performances des codeurs vidéo.

LISTE DES PUBLICATIONS

◊ Best Paper Award in the European Conférence on Visual Image Processing (EUVIP'11), Paris, France

- ◊ Revue Internationale avec comité de lecture :
 - M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "Efficient bit allocation using new intra and Inter-Frame modeling for H.264/AVC", Journal of Electronic Imaging (A paraître)

◊ Conférences Internationales avec Comité de Lecture :

- M. Hrarti, A. Tamtaoui, "Digital Video Quality Measures For H.263 And MPEG-2 Coded Video ", 3rd IEEE International Symposium on Image Video Communications over fixed and mobile networks (ISIVC2006), Yasmine Hammamet-Tunisia, September 2006 (Session Poster).
- M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "New H.264 Intra-Rate Estimation And Inter-Rate Control Driven By improved MAD-Based Contrast Sensitivity", IEEE International Conference on Image Processing (ICIP'09), 7-11 Novembre 2009, Caire, Egypte (Session Poster).
- M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "A New approach of Rate-Quantization Modeling for Intra and Inter Frames in H.264 Rate Control", IEEE International Conference on Signal and Image Processing Applications (ICSIPA'09), 18-19 Novembre 2009, Kuala Lumpur, Malaysia (Session Présentation).
- M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "A macroblockbased perceptually adaptive bit allocation for H264 rate control ", 5rd IEEE International Symposium on Image Video Communications over fixed and mobile networks (ISIVC2010), 30 September, 1&2 October 2010, Rabat- Maroc (Session Poster).

- M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "Improving the quality of H.264/AVC by using a new Rate-Quantization model ", IS&T/SPIE Electronic Imaging (IQSP'11), 23-27 January, 2011, California, UNITED STATES (Session Présentation).
- M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "A Perceptual Optimization of H.264/AVC Bit Allocation at the Frame and Macroblock Levels", IS&T/SPIE Electronic Imaging (IQSP'12), 22-26 January, 2012, Burlingame, California, UNITED STATES (Session Présentation, à paraître).
- M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "Adaptive Quantization Based on Saliency Map at Frame Level of H.264/AVC Rate Control Scheme ", European Workshop on Visual Image Processing (EUVIP'11), 4-6 Juillet 2011 Paris, France (Session Présentation).

◊ Conférence Nationale avec Comité de Lecture :

 M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "Contrôle de débit au niveau macrobloc du codeur H.264 conduit par un MAD amélioré à base de sensibilité au contraste", Ecole d'Hiver sur l'Image Numérique Couleur (EHINC'09), 13-16 Janvier 2009, Toulon, France (Session Poster).

◊ Workshops Internationaux et nationaux avec Comité de Lecture :

- M. Hrarti, A. Saadane, M. C. Larabi, A. Tamtaoui, D. Aboutajdine, "A Novel Inter Rate-Quantization Model in H.264/AVC Rate Control Scheme ", Workshop sur les Technologies de l'Information et de la Communication (WOTIC'09), 24-25 Décembre 2009, Agadir, Maroc (Session Présentation).
- M. Hrarti, A. Saadane, A. Tamtaoui, D. Aboutajdine, "Etude et Simulation du Contrôle de Débit dans le Codeur H.264/AVC ", Workshop sur les Technologies de l'Information et de la Communication (WOTIC'07), 4-5 Juillet 2007, à l'ENSIAS, Rabat, Maroc (Session Présentation).

ANNEXE STANDARD DE CODAGE VIDÉO H.264/AVC

Depuis quelques années, les standards de codage vidéo se développent considérablement. La variété de programmes, la bonne qualité de réception, l'Interactivité proposée expliquent certainement cet engouement populaire. Les fournisseurs de codeurs conscients de l'enjeu économique s'efforcent de rester concurrentiels en proposant des solutions toujours meilleures. En effet, l'augmentation du nombre de programmes, des services proposés aux téléspectateurs sans nuire à la qualité de l'image et sans augmentation de débit font partie des principaux objectifs à atteindre. Le nouveau standard International de codage vidéo H.264/MPEG-4 part10 vise à exaucer ces objectifs en améliorant significativement l'efficacité du codage et la robustesse d'erreurs contrairement aux normes précédentes telles que MPEG-2, H.263, MPEG-4 part2. Cette annexe présente une vue d'ensemble du profil de base du H.264/MPEG-4 part10. En commençant par la structure globale de ce système de codage, nous explicitons ensuite les différents éléments le constituant, en suivant l'ordre du traitement effectué.

A1 Standard H.264/AVC

A1.1 Introduction

Les groupes d'étude Internationaux, VCEG (Video Coding Experts Group) de l'ITU-T (International Telecommunication Union - Telecommunication) et MPEG (Moving Picture Experts Group) de l'ISO/IEC, ont recherché depuis le début des années 1990, des techniques de codage vidéo pour différents types d'applications. Ainsi, l'ITU-T a développé le H.261 comme première norme de codage vidéo pour les applications de vidéoconférence. La norme MPEG-1 a été accomplie pour le stockage sur le disque compact et le standard MPEG-2 (l'ITU-T l'a adopté comme H.262) a été destiné pour la télévision numérique. D'autre part, l'ITU a développé le H.263 afin d'améliorer la performance du H.261, et le modèle de base de codage du H.263 a été adopté comme noyau dans quelques parties du MPEG-4 telles que la partie 2. En 1998, le groupe VCEG s'est lancé dans un projet appelé H.26L en vue de multiplier par deux l'efficacité du codage par rapport à n'importe quelle norme de codage vidéo existante. En décembre 2001, soucieuses de mettre au point un système de compression plus efficace, les instances de normalisation (ISO/IEC, l'IUT) conjuguent leurs efforts au sein du groupe de travail JVT (Joint Video Team) chargé de développer le système de codage AVC (Advanced Video Coding). En 2003, le système AVC est intégré en tant que partie 10 à la norme MPEG-4 (ISO/IEC14496-10) et repris sous l'appellation H.264 à l'UIT.

La norme H.264/AVC a les mêmes éléments fonctionnels de base que les autres standards précédents tels que la transformation pour la réduction de la corrélation spatiale, la quantification pour le contrôle de débit binaire, la prédiction compensée en mouvement pour diminuer la corrélation temporelle et le codage entropique pour réduire la corrélation statistique. Cependant, afin d'accomplir une meilleure performance de codage, les changements importants dans H.264/AVC se produisent dans les détails de chaque élément fonctionnel en incluant la prédiction Intra-image, une nouvelle transformation DCT entière de blocs 4x4, de multiples images de référence, des tailles variables de blocs et une compensation de mouvement jusqu'au quart de pixel, un filtre de déblocage et un codage entropique amélioré.

Améliorer l'efficacité du codage vient aux dépens de la complexité calculatoire au sein du codeur/décodeur. *H.264/AVC* adopte des méthodes pour réduire la complexité d'implantation. Les conditions des canaux à bruit tels que les réseaux sans fil, obstruent la réception parfaite du train binaire codé par le décodeur. Un décodage incorrect dégrade la qualité subjective de l'image. Ainsi, le H.264/AVC quelques techniques telles que la résilience d'erreur dans les réseaux à bruits.

Pour des applications particulières telles que la vidéo conférence et la vidéo téléphonie, le codage haute qualité pour la transmission et le vidéo streaming sur les réseaux par paquets, H.264/AVC définit des profils et des niveaux spécifiant des restrictions appliquées sur les trains binaires. Sept profils sont définis afin de couvrir la gamme diversifiée des applications allant des réseaux sans fil jusqu' au cinéma numérique [51].

A1.2 Couche codage vidéo

Le concept de la couche de codage vidéo H.264/AVC est semblable à celui des autres normes telles que le MPEG-2. Il s'agit d'un hybride de prédiction temporelle et de prédiction spatiale, allié à un codage par transformée. La Figure A1 illustre le fonctionnement de la couche de codage pour un macrobloc. Chaque image de la vidéo d'entrée est découpée en des macroblocs. Chacun est codé par le mode Intra ou Inter, ensuite une prédiction est formée. La prédiction Intra de chaque échantillon d'un bloc a recours aux échantillons dans les blocs voisins ayant été déjà codés, décodés et reconstruits. Le processus de codage choisit quels échantillons voisins prendre et comment en tirer profit pour la prédiction Intra menée simultanément au niveau du codeur et du décodeur en se servant de l'information connexe transmise pour cette prédiction. L'Inter-codage (prédictif ou Biprédictif) emploie une prédiction (compensation et estimation de mouvement) sur la base d'images précédemment décodées. Le processus de codage dans le cas d'une prédiction Inter consiste à choisir des données sur le déplacement, englobant l'image de référence, et un déplacement appliqué à tous les échantillons du bloc. Les données sur le déplacement transmises en tant qu'informations connexes sont disposées par le codeur et le décodeur pour fournir simultanément le signal de prédiction Inter-image. L'erreur résiduelle (*Intra* ou *Inter*-image) qui n'est autre que la différence entre le bloc de départ et le bloc prédit, est transformée. Les coefficients de la transformée sont rééchelonnés et quantifiés et, par suite soumis à un codage entropique puis transmis avec les autres informations pour une prédiction Intra-image ou *Inter*-image.

Le codeur effectue également un décodage (une reconstruction) des macroblocs afin de créer les références nécessaires à la prédiction. Les coefficients quantifiés subissent un rééchelonnement et une transformée inverses similaires à ceux du décodeur, aboutissant au reliquat de prédiction décodé. Ce dernier est ajouté à la prédiction. Le résultat de cette opération est transmis au filtre de "déblocage" qui fournit la vidéo décodée après avoir éliminé les effets de bloc [73] [76].



FIGURE A1 – Principe du codage H.264/AVC d'un macrobloc.

Une description plus détaillée des caractéristiques techniques de la norme H.264 va suivre. Il est à noter qu'il existe différentes manières d'envisager les traitements au sein du codeur, ce qui est défini dans les profils et niveaux.

A1.3 Profils et niveaux

Les profils et les niveaux définissent les points de conformité. Ces derniers sont conçus pour faciliter l'Interopérabilité entre plusieurs applications axées sur la norme H.264/AVC, qui requièrent une configuration semblable. Un profil définit un ensemble d'outils de codage ou d'algorithmes servant à générer un flux compatible tandis qu'un niveau impose des contraintes à certains paramètres clés du flux. Les décodeurs conformes à un profil spécifique doivent prendre en charge toutes ses fonctionnalités. Les codeurs ne sont pas tenus d'user un ensemble particulier de fonctionnalités prises en charge dans un profil, mais doivent fournir des flux compatibles. Il existe trois profils dans la première version du H.264/AVC : basique, principal et étendu. Le profil de base est applicable aux services conversationnels en temps réel tels que : la vidéo-téléphonie, la vidéoconférence et les communications sans fil. Le profil principal est destiné à la télévision et au stockage alors que le profil étendu est plus adapté aux services multimédias à travers Internet. En outre, il existe quatre profils supérieurs (High profiles) destinés à d'autres types d'applications.



FIGURE A2 – Les parties spécifiques des profils dans H.264/AVC.

- La Figure A2 illustre les différentes parties communes et spécifiques des profils ci-dessus [76].
 - Le profil de base prend en charge toutes les fonctionnalités de la norme H.264/AVC à savoir le codage Intra et Inter (tranches I et P) et le codage entropique CAVLC (Context Adaptive Variable Length Coding), à l'exception des deux ensembles suivants :
 - * *Ensemble 1* : Tranches *B*, prédiction pondérée, *CABAC* (Context Adaptive Binary Arithmetic Coding), codage de champ et commutation adaptative en termes de macroblocs entre le codage en mode images et en mode champ.
 - * Ensemble 2 : Tranches SI et SP.
 - Le **profil principal** inclut le support de la vidéo entrelacée tout en prenant en charge le premier ensemble de fonctionnalités, mais contrairement au profil de base, ne reconnaît pas la fonctionnalité FMO (Flexible Macroblock Ordering).
 - Le **profil étendu** prend en charge les deux ensembles de fonctionnalités en plus des fonctionnalités du profil de base à l'exception du CABAC et de la commutation adaptative entre le codage en mode images et en mode champ. Ce profil permet d'améliorer la résilience d'erreur (Data Partionning) [76].
 - Les **profils supérieurs** : La norme *H.264/AVC* porte principalement sur la vidéo de "qualité divertissement" basée sur 8 bits/échantillon et un échantillonnage de la chrominance 420. Pour répondre aux besoins des applications les plus exigeantes - conçues par exemple pour la contribution et la distribution de contenus, le montage studio et la postproduction - le projet conjoint a été étendu en vue d'augmenter les capacités de la norme standard. Au départ appelées professional extensions, elles ont été rebaptisées fidelity range extensions ou *FRExt*. Lors de sa mise au point, le projet FRExt a abouti à quatre nouveaux profils appelés collectivement profils supérieurs :
 - * *Profil supérieur (HP)* Prend en charge la vidéo 8 bits et l'échantillonnage 420, et répond aux besoins du consommateur faisant usage du matériel haut de gamme et des applications performantes qui exigent une vidéo haute résolution sans devoir recourir à des formats de chrominance ou une précision d'échantillonnage étendus.
 - * Profil supérieur 10 (Hi10P) Prend en charge la vidéo 420, jusqu'à 10 bits par échantillon.

- * *Profil supérieur 422 (H422P)* Prend en charge l'échantillonnage de la chrominance jusqu'à 422 et 10 bits par échantillon.
- * *Profil supérieur 444 (H444P)* Prend en charge l'échantillonnage de la chrominance jusqu'à 444 et 12 bits par échantillon. En outre, il permet un codage zonal sans pertes ainsi qu'une transformée des couleurs résiduelles pour le codage de vidéo RVB évitant les erreurs de transformation de l'espace couleur.

Ces profils prennent tous en charge les fonctions du profil principal et ainsi qu'une taille de bloc pour transformées adaptatives et des matrices adaptées à la quantification perceptuelle [88].

La norme H.264/AVC emprunte le même jeu de définitions de niveau pour tous les profils, mais des applications individuelles peuvent prendre en charge un niveau différent pour chaque profil compatible. Onze niveaux sont définis spécifiant les limites supérieures de la taille d'images (dans les macroblocs), la vitesse de traitement du décodeur(en macroblocs par seconde), la taille des mémoires multi-image, le débit binaire vidéo et la taille de la mémoire vidéo. Dans ce présent chapitre, nous détaillons la majorité des fonctionnalités du profil de base.

A1.4 Description technique

Dans cette section, nous allons détailler la majorité des composantes de la Figure A1, à savoir :

- * La prédiction Intra-Image,
- * La prédiction Inter-Image,
- * La transformation et la quantification,
- * Le codage entropique,
- * Le filtre de suppression des effets de blocs.

A1.5 Prédiction Intra-Image

La prédiction Intra-image permet d'exploiter la redondance spatiale présente dans l'image courante sans avoir besoin d'aucune image de référence. Contrairement aux normes de codage vidéo précédentes où la prédiction Intervenait dans le domaine transformé, la prédiction spécifique à la norme H.264/AVC Intervient toujours dans le domaine spatial en se référant aux échantillons voisins de blocs déjà codés. Un bloc prédit est donc construit à partir de ses voisins déjà codés, décodés et construits. Il est ensuite soustrait au bloc original et le bloc de différence est codé et transmis.

Chaque macrobloc peut être transmis en choisissant un type de codage parmi d'autres mais en fonction de celui désigné pour le codage de la tranche. Dans tous les types de codage de tranches, deux classes de codage Intra sont pris en charge : Intra-4x4 et Intra-16x16. Dans le mode Intra-4x4, chaque bloc 4x4 de luminance recourt à un des neuf modes de prédiction. Outre la prédiction DC, huit modes de prédiction directionnelle sont spécifiés. Dans le mode Intra-16x16, une prédiction uniforme est réalisée pour l'ensemble de la luminance d'un macrobloc. Quatre modes de prédiction coexistent. La prédiction des échantillons de chrominance d'un macrobloc repose toujours sur une technique semblable à celle utilisée pour la luminance des macroblocs Intra-16x16. La prédiction Intra au-delà des limites de la tranche n'est pas autorisée afin de maintenir l'in-

dépendance des tranches les unes par rapport aux autres. Dans ce qui suit, nous présentons les procédures de la prédiction Intra-image, suivant la taille de bloc considérée [39] [73] [76].

a) Prédiction des blocs 4x4 de luminance

Ce type de prédiction est plus employé dans les régions de détails, sachant que la précision est nécessaire. Ainsi, on a défini neuf modes de prédiction différents (8 modes directionnels et un mode DC (*Figure A3B*). Chaque bloc 4x4 (pixels [a-p]) est prédit spatialement à partir des pixels au-dessus et à gauche (pixels [A-M]) du bloc courant qui ont été déjà codés et reconstruits (*Figure A3A*).



FIGURE A3 – A- Labellisation des échantillons de prédiction (4x4), B- Huit des neuf modes de prédiction Intra-4x4.

La *Figure A4* ainsi que le *Tableau A1* présentent les neuf modes de prédiction. Les flèches de la figure indiquent la direction de la prédiction de chaque mode. Pour les modes [3-8], les pixels prédits sont calculés par une moyenne pondérée des pixels [A-M].



FIGURE A4 – Modes de prédiction des blocs 4x4 de luminance.

Par exemple, les échantillons a et d sont prédits respectivement par : round (I/4+M/2+A/4) et round (B/4+C/2+D/4) par le biais du mode 4. Les mêmes échantillons sont prédits respectivement par : round (I/2+J/2) et round (J/4+K/2+L/4) en employant le mode 8. L'encodeur peut choisir le mode de prédiction qui minimise le résiduel entre le bloc à coder et sa prédiction [39] [73].

b) Prédiction des blocs 16x16 de luminance

Modes	Signification
Mode 0 (Vertical)	[A-D] Extrapolés verticalement
Mode 1 (Horizontal)	[I-L] Extrapolés horizontalement
Mode 2 (Composante continue : DC)	[a-p] Prédits en moyennant [A-D] et [I-L]
Mode 3 (Diagonal bas-gauche)	[a-p] Interpolés à un angle de 45°bas-gauche
Mode 4 (Diagonal bas-droit)	[a-p] Interpolés à un angle de 45° bas-droit
Mode 5 (Vertical-droit)	[a-p] Interpolés à un angle de 22,5° vertical-droit
Mode 6 (Horizontal-bas)	[a-p] Interpolés à un angle de 22,5° Horizontal-bas
Mode 7 (Vertical-gauche)	[a-p] Interpolés à un angle de 22,5° Vertical-gauche
Mode 8 (Horizontal-haut)	[a-p] Interpolés à un angle de 22,5° Horizontal-haut

TABLE A1 – Signification des modes de prédiction des blocs 4x4 de luminance.

Ce type de prédiction est recommandé dans le cas des zones d'image régulières. Quatre modes sont disponibles. Ils sont représentés sur la *Figure A5* et explicités sur le *Tableau A2*.



FIGURE A5 – Modes de prédiction des blocs 16x16 de luminance.

TABLE A2 – Signification d	es modes de	prédiction des	blocs $16x16$	de luminance.
----------------------------	-------------	----------------	---------------	---------------

Modes	Signification
Mode 0 (Vertical)	Extrapolation à partir de H
Mode 1 (Horizontal)	Extrapolation à partir de V
Mode 2 (Composante continue : DC)	Moyenne de H et V
Mode 3 (Plan)	Fonction linéaire plane ajustée à partir de H et V

c) Prédiction des blocs 8x8 de chrominance

Dans ce type de prédiction, les deux composantes de chrominance (Cr et Cb) utilisent le même mode de prédiction. Chaque composante 8x8 d'un macrobloc est prédite à partir des

échantillons situés en haut et à gauche du bloc courant et qui ont été précédemment codés et reconstruits. Il existe quatre modes de prédiction similaires à ceux employés dans la prédiction des blocs 16x16 de luminance, sauf que l'ordre de numérotation est différent : DC (Mode0), Horizontal (Mode1), Vertical (Mode2) et Plan (Mode 3) [73].

d) Sélection de mode Intra dans H.264/AVC

H.264/AVC offre un ensemble riche de modèles de prédiction Intra (9 modes de prédiction Intra-4x4, 4 modes de prédiction Intra-16x16 et 4 modes de prédiction des blocs de chrominance). Avec l'emploi de ces techniques avancées de prédiction Intra, l'efficacité du codage a été significativement améliorée. Pour tirer profit des différents modes décrits ci-dessus, le codeur H.264/AVC peut sélectionner le meilleur mode en profitant de la procédure d'optimisation débit-distorsion (RDO) qui cherche exhaustivement le meilleur mode minimisant le coût suivant :

$$J(s, c, mode/QP, \lambda_{mode}) = D(s, c, mode/QP) + \lambda_{mode} * R(s, c, mode/QP)$$
(A1)

Avec QP est le paramètre de quantification du macrobloc, λ_{mode} est le multiplicateur de Lagrange, D est la distorsion (SSD, MAD, MSE...) entre le bloc original s et sa reconstruction c, R est le nombre de bits pour coder le bloc s et mode représente le potentiel mode de prédiction. Pour mieux expliciter le processus de sélection de mode, nous avons pris à titre d'exemple le codeur JM10.2, au sein duquel l'algorithme de recherche exhaustive (FS : Full Search) est introduit pour examiner tous les modes possibles de prédiction Intra afin de choisir le meilleur. Les étapes majeures de ce processus peuvent être résumées comme suit :

• Etape 1 : Trouver le meilleur mode de prédiction Intra-16x16

Générer les quatre prédictions de blocs selon les quatre modes Intra-16x16 ensuite calculer leurs macroblocs résiduels. Pour chacun de ces macroblocs :

- ♦ 1.1 : Exécuter la transformée d'Hadamard pour chaque bloc 4x4.
- \diamond **1.2** : Extraire toutes les composantes continues (DC) des seize blocs 4x4 et les mettre dans un autre bloc 4x4 auquel la transformée d'Hadamard sera exécutée une seconde fois.
- ◇ 1.3 : Sommer les valeurs absolues de tous les coefficients transformés par Hadamard (à l'exception des coefficients DC de l'étape 1.1) et utiliser cette sommation comme coût.
 Le mode possédant le plus petit coût sera choisi comme meilleur mode de prédiction Intra-16x16.

• Etape 2 : Trouver le meilleur mode de prédiction Intra4x4

Diviser le macrobloc en seize blocs 4x4. Pour chacun d'eux, trouver le meilleur mode de prédiction comme suit :

- $\diamond 2.1$: Générer les neuf blocs en se basant sur les neuf modes de prédiction Intra-4x4.
- ♦ **2.2** : Pour chaque mode, calculer le coût définit par : $Cost_{4\times4} = MSE_{4\times4} + R \times \lambda(Q_p)$ Avec *MSE* est l'erreur quadratique moyenne encodée, *R* est le nombre de bits pour coder le bloc 4x4 et $\lambda(Q_p)$ est une fonction exponentielle du paramètre de quantification.
- $\diamond~2.3$: Le meilleur mode est celui qui minimise le coût.
- Etape3 : Trouver le meilleur mode de prédiction

Calculer le coût global : $Cost = MSE + R \times \lambda(Q_p)$ de ce macrobloc en tirant profit du meilleur mode de prédiction Intra-16x16 et des seize meilleurs modes de prédiction Intra-4x4. Le mode ayant le faible coût est choisi comme mode optimal [100].

Il existe également un processus de choix du mode optimal des deux composantes de chrominance. Finalement, nous pouvons remarquer que l'application de la procédure d'optimisation débit-distorsion dans la sélection du mode optimal dans le cas de la prédiction *Intra*-image, apporte une complexité calculatoire accrue (le nombre de combinaisons pour les composantes de luminance et chrominances d'un macrobloc est $N_{8\times8} \times (N_{4\times4} \times 16 + N_{16\times16})$, avec désignent $N_{8\times8}$, $N_{4\times4}$ et $N_{16\times16}$ respectivement, le nombre des modes des blocs 8x8 de chrominance, des blocs 4x4 et des macroblocs de luminance). Ceci veut dire que pour un macrobloc, 592 calculs de RDO sont exécutés, ce qui impose une difficulté de réaliser des implantations en temps réel. Actuellement, il y a proposition de méthodes rapides (Fast Method) représentant un domaine de recherche actif et permettant d'alléger la complexité calculatoire en exploitant les caractéristiques des trames, les propriétés des macroblocs [96].

e) Codage des modes de prédiction

Le choix du mode de prédiction Intra de chaque bloc 4x4 doit être signalé au décodeur et ceci peut potentiellement exiger un large nombre de bits. Cependant, au voisinage de blocs 4x4, les modes sont fortement corrélés. Par exemple, si les deux blocs 4x4 (A et B de la Fi-gure A6) ont été prédits par le mode 2, il est probable que le meilleur mode du bloc courant C soit aussi le mode 2.

Pour chaque bloc courant C, le codeur et le décodeur calculent le mode de prédiction le plus probable. Si les deux blocs A et B sont codés en Intra-4x4 et appartiennent à la même tranche courante, le mode de prédiction le plus probable est le minimum des modes de prédiction de A et B. Le codeur envoie pour chaque bloc 4x4, un indicateur nommé "use-most-probable-mode" pour indiquer si le mode de prédiction le plus probable est employé. Si c'est le cas, cet indicateur est égal à "1", sinon il est égal à "0" et un autre paramètre "remaining-mode-selector" est plus petit que le mode de prédiction le plus probable courant, alors le mode de prédiction prend la valeur du "remaining-mode-selector", sinon, le mode de prédiction devient (remaining-mode-selector+1) [73].



FIGURE A6 – Blocs adjacents codés en Intra-4x4.

Dans cette section nous avons présenté toute la procédure de la prédiction Intra-image suivant la taille de bloc considérée. En effet, nous pouvons remarquer que la norme H.264/AVC apporte quelques différences par rapport aux autres standards à savoir : l'usage de blocs de taille 4x4 alors que les normes antécédentes travaillaient généralement avec des blocs de taille 8x8, ainsi que l'introduction d'un ensemble riche de modes de prédiction. Dans ce qui suit, nous allons présenter le second type de prédiction nommée "Inter-image".

A1.6 Prédiction Inter-Image

Outre les types de codage Intra des macroblocs, plusieurs types de codage prédictif ou à compensation de mouvement sont définis pour les macroblocs d'une tranche P. La prédiction Inter crée un modèle de prédiction à partir d'une ou plusieurs trames vidéo codées précédemment au moyen de l'estimation et de la compensation de mouvement afin de réduire les corrélations temporelles. Parmi les principales différences qu'apporte H.264/AVC par rapport à ses antécédents, nous citons : une large gamme de tailles de blocs (jusqu'à 4x4) et un sous échantillonnage plus fin des vecteurs de mouvement (résolution au quart d'échantillon pour la luminance). Dans ce qui suit, nous allons présenter l'ensemble des mécanismes de la prédiction Inter tels que : le processus de compensation de mouvement, la prédiction des vecteurs de mouvement et le choix du mode de codage optimal.

a) Compensation de mouvement à structure d'arbre

H.264/AVC admet un large support de taille de blocs allant de 16x16 à 4x4. La composante de luminance de chaque macrobloc peut être fractionnée de quatre manières comme le montre la *Figure A7A* : 16x16, 16x8, 8x16 ou 8x8. Chacune des régions subdivisées est une partition de macrobloc. Si le mode 8x8 est choisi, les quatre sous-macroblocs peuvent à leur tour être divisés en quatre manières comme le montre la *Figure A7B*. Ces partitions et sous partitions génèrent un grand nombre de combinaisons au sein d'un macrobloc. Cette méthode de partitionnement de macroblocs en tailles variables est connue sous le nom de "compensation de mouvement à structure d'arbre".



FIGURE A7 – A- Partitions de macrobloc : 16x16, 8x16, 16x8, 8x8. B- Sous-partitions de macrobloc : 8x8, 4x8, 8x4, 4x4.

Chaque partition ou sous-partition nécessite un vecteur de mouvement. Chaque vecteur doit être codé et transmis; de plus, le choix de partitionnement retenu doit être inclus dans le train binaire codé. Choisir une partition de grande taille (16x16, 16x8, 8x16) implique qu'un petit nombre de bits est nécessaire pour signaler le choix de vecteur(s) de mouvement et le type de partition; cependant, le résiduel compensé en mouvement peut contenir une quantité d'énergie significative dans les zones d'image très détaillées. Choisir une partition de petite taille (8x4, 4x8...) engendre un résidu à faible énergie après le processus de compensation de mouvement mais nécessite un large nombre de bits pour coder les vecteurs de mouvement et le choix de partition(s). Le choix de la taille de partition a un impact significatif sur la performance de la compression. En général, une partition à taille large est plus appropriée pour les zones d'image homogènes et une petite partition peut être bénéfique pour les zones d'intérêt.

La résolution de chaque composante de chrominance (Cr et Cb) est moitié de celle de la composante de luminance. Chaque bloc de chrominance est partitionné de la même manière que la composante de luminance, sauf que la taille de partition est divisée par deux horizontalement et verticalement (une partition de luminance 8x16 correspond à une partition 4x8 de chrominance). Les composantes horizontale et verticale de chaque vecteur de mouvement sont divisées en deux pour les blocs de chrominance [73] [76].

b) Vecteurs de mouvement

Chaque partition d'un macrobloc codé en Inter est prédite à partir d'une zone de même taille d'une image de référence. La différence entre les deux zones (le vecteur de mouvement) a une résolution au quart de pixel (pour la composante de luminance). Les échantillons de luminance et de chrominance aux positions sous échantillonnées n'existent pas dans l'image de référence, il est donc nécessaire de les créer par Interpolation à partir d'échantillons voisins. La *Figure A8* illustre un exemple d'Interpolation entre échantillons. Une sous partition de taille 4x4 dans la trame courante (*Figure A8A*) doit être prédite à partir d'une région voisine de l'image de référence. Si les composantes horizontale et verticale du vecteur de mouvement sont des valeurs entières (*Figure A8B*), les échantillons appropriés dans le bloc de référence existent réellement (points gris). Si l'une ou les deux composantes du vecteur sont des valeurs fractionnelles (*Figure A8C*), les échantillons de prédiction (points gris) sont générés par Interpolation entre les échantillons adjacents dans la trame de référence (points blancs) [73].



FIGURE A8 – Exemple de la prédiction entière et la prédiction sous-pixel.

c) Prédiction des vecteurs de mouvement

Encoder un vecteur de mouvement pour chaque partition peut prendre un nombre significatif

de bits, spécialement si des partitions de petite taille sont choisies. Les vecteurs de mouvement des partitions voisines sont souvent fortement corrélés, ainsi, chaque vecteur de mouvement est prédit à partir des vecteurs des partitions voisines déjà codées. Un vecteur prédit (MVp) est formé en se basant sur les vecteurs de mouvement précédemment calculés. La différence entre le vecteur de mouvement courant et celui prédit (MVD) est codée et transmise. La méthode de formation de la prédiction (MVp) dépend de la taille de la partition, de la compensation de mouvement et de la disponibilité des vecteurs voisins [73].

d) Modes de prédiction des vecteurs de mouvement

Quatre modes de prédiction sont introduits pour prédire le vecteur de mouvement du bloc courant : *Prédiction médiane* (Median Prediction), *Prédiction Uplayer* (Uplayer Prediction), *Prédiction Correspondante* (Corresponding Prediction) et *Prédiction voisine des trames de référence* (Neighboring Ref-frame Prediction). La prédiction de base la plus appliquée est la prédiction médiane.

En effet, le standard H.264/AVC applique un codage prédictif des vecteurs de mouvement, et le résiduel de ce vecteur MVD est donné par la formule suivante : MVD = MV - MVp, avec MV est le vecteur de mouvement actuel, MVp est le médian de trois vecteurs voisins représentés dans la Figure A9 (MVa, MVb, MVc).

MVd	МУЬ	MVc			
MVa	MV				

FIGURE A9 – Localisation des vecteurs spatiaux utilisés pour la prédiction.

La prédiction médiane est calculée selon la disponibilité ou non des vecteurs voisins. Ainsi, si MVb n'est pas disponible, sa valeur sera égale à celle du vecteur MVd. Si un ou plusieurs des vecteurs voisins ne sont pas disponibles, MVp est égal, en fonction des disponibilités, à MVa ou MVb ou MVc ou "0" [53] [73] [84].

e) Décision de mode en H.264/AVC

Le standard H.264/AVC adopte un ensemble très riche de modes de prédiction en le comparant à ses antécédents. Pour une trame P, un macrobloc peut être codé de différentes manières : SKIP, 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4, Intra-4x4 et Intra-16x16. Nous rappelons que le mode Intra-4x4 est composé de neuf Interpolations directionnelles tandis que le mode Intra-16x16 de quatre directions d'Interpolation. Le codeur examinera tous ces modes afin d'obtenir le mode optimal (Notons que le processus d'optimisation débit distorsion est effectué également à ce niveau là pour sélectionner le mode minimisant le coût de Lagrange). Brièvement, la décision du mode d'un macrobloc d'une trame P peut être résumée comme suit :

 \diamond Tester le mode SKIP en premier lieu : Le mode SKIP représente le cas ou la taille du bloc est 16x16 mais aucun vecteur de mouvement ou information résiduelle ne sont codés. Ainsi,

c'est le mode le plus simple du fait qu'il possède la plus faible complexité calculatoire.

- ◇ Tester les modes Inter/Intra : Dans le cas du mode Inter, toutes les combinaisons 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, 4x4 sont testées et en appliquant le processus de RDO, le mode Inter ayant le coût minimal est sélectionné. La même procédure est appliquée dans le cas du mode Intra avec ses combinaisons Intra-4x4 (9 modes) et Intra-16x16 (4modes), le mode Intra minimisant le coût de Lagrange est choisi.
- ◊ Finalement, une comparaison est effectuée entre le mode SKIP, Intra et Inter, et celui ayant le faible coût est sélectionné comme mode optimal de codage [24].

Nous remarquons que le processus de choix du mode optimal apporte une complexité calculatoire accrue, ainsi, des méthodes rapides ont été récemment proposées pour alléger le surcroît de calcul (Mêmes remarques que dans le cas de la prédiction intra-image).

Le principe de la prédiction Inter-image utilisé dans H.264/AVC est similaire à celui des autres standards. Les différences résident dans le partitionnement fin des macroblocs et la résolution des vecteurs de mouvement. Une fois que l'erreur de prédiction est obtenue, elle va être transformée et quantifiée avant d'être transmise.

A1.7 Transformation et Quantification

À l'instar des normes de codage vidéo précédentes, la norme H.264/AVC a également recours au codage par transformée du reliquat de la prédiction. Toutefois, dans le cas de H.264/AVC, la transformée est appliquée à des blocs de 4x4. Le profil de base du H.264/AVC procède avec trois sortes de transformations selon le type de données résiduelles à coder. Le premier type de transformation opère sur tous les échantillons de tous les blocs de l'erreur résiduelle des composantes de luminance et de chrominance (Cr, Cb) indépendamment du type de la prédiction Intra/Inter choisi. La taille de la matrice de transformation est 4x4 (H1 de la Figure A10). Si le macrobloc est prédit en mode Intra-16x16, une seconde transformée d'Hadamard de taille 4x4 (H2 de la Figure A10) est appliquée sur les composantes continues DC placées dans une matrice 4x4. La troisième transformée d'Hadamard de taille 2x2 (H3 de la Figure A10) est appliquée sur les composantes continues DC (placées dans une matrice 2x2) des deux composantes de chrominance.

FIGURE A10 – Matrices H1, H2 et H3 des trois transformées différentes Appliquées dans H.264/AVC.

L'ordre de transmission de tous les coefficients est montré sur la *Figure A11*. Si le macrobloc est prédit en mode Intra-16x16, le bloc labellisé "-1" est transmis en premier. Ce bloc contient les coefficients DC de tous les blocs de la composante de luminance. Ensuite, les blocs résiduels

de luminance [0-15] sont transmis dans le même ordre (les coefficients DC d'un macrobloc codé en Intra-16x16 sont mis à "0"). Les blocs 16 et 17 contenant les coefficients DC des deux composantes de chrominance Cb et Cr respectivement sont envoyés. Finalement, les blocs résiduels de chrominance [18-25] (sans les coefficients DC) sont transmis [73].



FIGURE A11 – Ordre de transmission des blocs résiduels dans un macrobloc .

Dans ce qui suit, nous allons détailler les deux processus de transformation et quantification pour les blocs résiduels 4x4 et les blocs contenant les coefficients DC des composantes de luminance et de chrominance lorsqu'un macrobloc est codé en mode Intra-16x16.

a) Transformation et quantification des blocs résiduels 4x4 (blocs : [0-15], [18-25])

Les blocs résiduels (labellisés [0-15] et [18-25] dans la *Figure A11*) sont transformés en appliquant une transformée en cosinus discrète entière (DCT entière ou ICT). L'ICT est basée sur la DCT classique utilisée dans les standards précédents avec quelques différences fondamentales :

- ◊ C'est une transformation entière (Toutes les opérations peuvent être effectuées en arithmétique entière sans perte d'exactitude).
- $\diamond\,$ La transformée inverse est entièrement spécifié dans le standard H.264/AVC; ainsi, si cette spécification est suivie correctement, la différence entre le codeur et le décodeur ne devrait pas se produire.
- ◊ Le cœur de la transformée nécessite uniquement des additions et des décalages.
- ◊ Une multiplication échelonnée (partie de la transformée complète) est intégrée à l'étape de quantification (réduisant le nombre total de multiplications).

Le processus entier de transformation et quantification peut être effectué à l'aide de l'arithmétique entière à 16 bits sans aucune perte d'exactitude [64] [73].

• Développement de la transformation à partir de la DCT 4x4 classique :

L'équation de la transformation utilisée peut être factorisée sous la forme suivante :

$$Y = (CXC^{T}) \otimes E$$

$$= \left(\begin{bmatrix} 1 & 1 & 1 & d \\ 1 & d & -1 & -1 \\ 1 & -d & -1 & 1 \\ 1 & -1 & 1 & -d \end{bmatrix} [X] \begin{bmatrix} 1 & 1 & 1 & d \\ 1 & d & -1 & -1 \\ 1 & -d & -1 & 1 \\ 1 & -1 & 1 & -d \end{bmatrix} \right) \otimes \begin{bmatrix} a^{2} & ab & a^{2} & ab \\ ab & b^{2} & ab & b^{2} \\ a^{2} & ab & a^{2} & ab \\ ab & b^{2} & ab & b^{2} \end{bmatrix}$$
(A2)

avec $a=1/2, b=\sqrt{\frac{1}{2}}$ et d=1/2.

 CXC^{T} est le cœur de la transformation 2D, E est une matrice d'échelonnage et le symbole \otimes indique une multiplication terme à terme. Pour simplifier l'implantation de la transformée, d est approximé par "0.5". Pour assurer l'orthogonalité de la transformation, la valeur de b est devenue $\sqrt{\frac{2}{5}}$ (car la valeur de b dans le cas de la DCT classique a été égale à $sqrt\frac{1}{2}cos\left(\frac{\pi}{8}\right)$).

Pour éviter les multiplications par un demi dans le cœur de la transformation CXC^T , ce qui provoque une perte d'exactitude, les matrices C et E ont été encore modifiées. La transformation finale devient donc :

$$Y = \left(C_f X C_f^T\right) \otimes E_f$$

$$= \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} [X] \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \right) \otimes \begin{bmatrix} a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \\ a^2 & ab/2 & a^2 & ab/2 \\ ab/2 & b^2/4 & ab/2 & b^2/4 \end{bmatrix}$$
(A3)

Cette transformation est une approximation de la DCT 4x4. En raison du changement des deux facteurs b et d, le résultat obtenu par la nouvelle transformée n'est pas toutefois égal à celui donné par la DCT classique [14] [73].

La transformation inverse est donnée par :

$$X = C_i^T (Y \otimes E_i) C_i$$

$$= \begin{pmatrix} 1 & 1 & 1 & 1/2 \\ 1 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{pmatrix} \begin{bmatrix} (X) \begin{pmatrix} a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \\ ab & b^2 & ab & b^2 \end{pmatrix} \otimes \begin{bmatrix} 1 & 1 & 1 & 1/2 \\ 1 & 1/2 & -1 & -1 \\ 1 & -1/2 & -1 & 1 \\ 1 & -1 & 1 & -1/2 \end{bmatrix}$$
(A4)

Cette fois Y est pré échelonné en multipliant chaque coefficient par le facteur approprié de pondération de la matrice E_i . Il est à noter que la transformation et son inverse sont orthogonaux, c'est-à-dire :

$$T^{-1}\left(T\left(X\right)\right) = X\tag{A5}$$

• Quantification H.264/AVC utilise une quantification scalaire [14] [73] dont l'opération de base est la suivante :

$$Z_{ij} = round\left(\frac{Y_{ij}}{Q_{step}}\right) \tag{A6}$$

Avec Y_{ij} est un coefficient de la transformation décrite ci-dessus, Q_{step} est la taille du pas de quantification et Z_{ij} est le coefficient quantifié. L'opération *round* donne l'entier le plus proche.

Au total, 52 valeurs du Q_{step} sont supportées par le standard, indexées par un paramètre de quantification QP (*Tableau A3*).

TABLE A3 – Tailles du pas de quantification dans H.264/AVC.

\mathbf{QP}	0	1	2	3	4	5	6	7	8	9	10	11	12	• • •
Q_{step}	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	2.25	2.5	
QP		18		24		30		36		42		48		51
Q_{step}		5		10		20		40		80		160		224

Il est à noter que Q_{step} double à chaque incrément de "6" de QP; et il augmente de "12.5%" à chaque incrément de "1" de QP. Ce large éventail de tailles de pas de quantification permet au codeur de contrôler avec exactitude et flexibilité la différence entre le débit binaire et la qualité. Les valeurs de QP de la luminance et de la chrominance peuvent être différentes; les deux paramètres varient entre "0" et "51" mais le QP_C est dérivé de QP_Y .

Les facteurs a^2 , ab/2 et $b^2/4$ (Eq. A3) sont incorporés dans le quantificateur. Tout d'abord, un bloc d'entrée X est transformé en un bloc de coefficients non quantifiés $W = CXC^T$. Ensuite, chaque coefficient W_{ij} est quantifié et échelonné en une seule opération :

$$Z_{ij} = round\left(W_{ij} \bullet \frac{PF}{Q_{step}}\right) \tag{A7}$$

PF prend l'une des valeurs suivantes : a^2 , ab/2 ou $b^2/4$, selon la position (i,j) :

TABLE A4 – Valeurs du PF selon la position de pixels.

Position	\mathbf{PF}
(0,0),(2,0),(0,2),(2,2)	a^2
(1,1),(1,3),(3,1),(3,3)	$b^{2}/4$
Autres	ab/2

Pour éviter toute opération de division, le facteur (PF/Q_{step}) est considéré comme une multiplication par un facteur MF (Facteur de multiplication) et un décalage à droite.

$$Z_{ij} = round \left(W_{ij} \bullet \frac{MF}{2^{qbits}} \right) \tag{A8}$$

Avec :

$$\frac{MF}{2^{qbits}} = \frac{PF}{Q_{step}} \tag{A9}$$

Et

$$qbits = 15 + floor\left(\frac{QP}{6}\right) \tag{A10}$$

En arithmétique entière, l'Eq. A8 peut être implantée comme suit :

$$|Z_{ij}| = (|W_{ij}| \bullet MF + f) >> qbits$$

signe (Z_{ij}) = signe (W_{ij}) (A11)

Où » indique un décalage à droite. Dans le logiciel de référence, $f = 2^{qbits}/3$ pour les blocs codés en Intra et pour ceux codés en Inter.

Le facteur de multiplication MF dépend de QP et de la position du coefficient (i,j) (Voir *Tableau A5*). Il est à noter que le facteur MF reste inchangé pour QP>5. Il peut être calculé par la formule suivante :

$$MF_{QP>5} = MF_{QP=QPmod6} \tag{A12}$$

TABLE A5 – Valeurs du facteur de multiplication MF selon le QP et la position (i,j). **OP** $(i, j) \cdot (0, 0) \cdot (2, 0) \cdot (0, 2) \cdot (2, 2) \cdot (j, j) \cdot (1, 1) \cdot (1, 3) \cdot (3, 1) \cdot (3, 3)$ Autres

ЧI	$(1, j) \cdot (0, 0), (2, 0), (0, 2), (2, 2)$	$(1, j) \cdot (1, 1), (1, 0), (0, 1), (0, 0)$	Autres
0	13107	5243	8066
1	11916	4660	7490
2	10082	4194	6554
3	9362	3647	5825
4	8192	3355	5243
5	7282	2893	4559

b) Transformation et quantification des coefficients DC de luminance (Mode Intra-16x16)

Si un macrobloc est codé en mode *Intra-16x16*, chaque bloc résiduel 4x4 est tout d'abord transformé en intégrant le cœur de la transformation décrite au-dessus $C_f X C_f^T$. Le coefficient DC de chaque bloc 4x4 est ensuite transformé par une transformation d'*Hadamard* 4x4:
Avec W_D est le bloc contenant les composantes continues et Y_D est le bloc obtenu après transformation. Les coefficients de sortie $Y_D(i, j)$ sont divisés par 2 (avec approximation), ils sont ensuite quantifiés selon la formule suivante :

$$\left| Z_{D(i,j)} \right| = \left(\left| Y_{D(i,j)} \right| \bullet MF_{0,0} + 2f \right) >> (qbits + 1)$$

signe $\left(Z_{D(i,j)} \right) = signe \left(Y_{D(i,j)} \right)$ (A14)

Avec $MF_{(0,0)}$ est le facteur de multiplication (Première colonne du *Tableau A5*), f et *qbits* sont tels qu'auparavant.

c) Transformation et quantification des coefficients DC de chrominance

Les coefficients DC de chaque bloc 4x4 de chrominance sont groupés dans un bloc 2x2 (W_D) qui sera transformé en appliquant une transformation d'Hadamard 2x2:

$$Y_D = \left(\begin{bmatrix} 1 & 1\\ 1 & -1 \end{bmatrix} [W_D] \begin{bmatrix} 1 & 1\\ 1 & -1 \end{bmatrix} \right)$$
(A15)

La quantification du bloc $2x^2$ de sortie (Y_D) est exécutée comme suit :

$$\left|Z_{D(i,j)}\right| = \left(\left|Y_{D(i,j)}\right| \bullet MF_{0,0} + 2f\right) >> (qbits + 1)$$

signe $\left(Z_{D(i,j)}\right) = signe\left(Y_{D(i,j)}\right)$ (A16)

Avec $MF_{(0,0)}$, f et qbits sont tels qu'auparavant.

Dans cette section, nous avons détaillé les deux processus de transformation et de quantification tels qu'ils sont appliqués dans le standard H.264/AVC. Une fois les coefficients sont transformés et quantifiés, ils vont être codés en s'appuyant sur deux types de codage : CAVLC et le codage de Golomb exponentiel pour réduire la quantité d'information transmise.

A1.8 Codage entropique

La norme H.264/AVC prend en charge deux méthodes de codage entropique. Le codage par défaut fait appel à un unique ensemble illimité de mots-codes défini pour tous les éléments syntaxiques, à l'exception des coefficients de transformée quantifiés. Ainsi, au lieu de concevoir une table de code à longueur variable (VLC) différente pour chaque élément syntaxique, seule la correspondance à la table de mots-codes unique, est personnalisée en fonction des statistiques de données. La table unique est un code Golomb exponentiel aux propriétés de décodage simples et régulières. Pour transmettre les coefficients de transformée quantifiés, on fait intervenir la méthode CAVLC (Context-Adaptive Variable Length Coding), plus sophistiquée. Dans ce cas, les tables VLC de plusieurs éléments syntaxiques sont remplacées en fonction de ceux déjà transmis. Comme ces tables sont spécialement conçues pour cadrer avec les statistiques correspondantes, les performances du codage entropique sont améliorées par rapport à celles obtenues par une seule table VLC [76].

a) Codage de Golomb exponentiel (Exp-Golomb)

Les codes du Golomb exponentiel sont à longueur variable et à construction régulière. Le $Tableau \ A6$ liste les neuf premiers mots-codes ; il est clair que ces mots-codes progressent en un ordre logique. Chaque mot-code est construit comme suit :

[M zeros][1][INFO]

Où *INFO* est un champ de *M* bits contenant l'information. Le premier mot-code ne comporte ni zéro ni information. La longueur de chaque mot-code est (2M+1) bits. Chaque mot-code peut être construit par le codeur en se basant sur son indice "code-num" comme suit :

 $M = \lfloor log_2 \lfloor code - num + 1 \rfloor \rfloor$ INFO=code-num+1-2^M

Table A6 – Mots	-codes du	codage de	e Golomb	o exponentiel.

Code-num	Mot-code
0	1
1	010
2	011
3	00100
4	00101
5	00110
6	00111
7	0001000
8	0001001
	•••

Un mot-code peut être décodé comme suit :

- Lecture des M zéros suivis par 1.
- Lecture des M-bit d'INFO.
- code-num= 2^M +INFO-1

Pour le premier mot-code, les champs M et INFO sont nuls.

Un paramètre v à coder est associé à code-num des trois manières suivantes (Tableau A?) :

Le *Tableau A8* liste une petite partie de la table du paramètre "coded-block-pattern". Pour les macroblocs codés en *Inter*, ce paramètre indique quels sont les blocs 8x8 au sein d'un macrobloc contenant des coefficients non nuls.

Chacune des associations décrites (ue, se et me) est désignée pour produire des mots-codes courts pour les valeurs les plus apparentes et des mots-codes longs pour les valeurs les moins apparentes. Par exemple, le type d'un macrobloc *Pred-L0-16x16* (Prédiction d'une partition 16x16 de luminance à partir d'une image précédente) est associé à la valeur *code-num=0* car il

TABLE A7 – Types de codage du paramètre v.				
Type	Description			
ue (v)	Association directe non signée, code-num=v			
	(Utilisée pour le type de macrobloc, Indice			
	de la trame de référence et autres).			
se (v)	Association signée, utilisée pour MVD, ΔQP			
	et autres : $code - num = 2 v $ si v<=0,			
	sinon, $code - num = 2 v - 1$ si v>0.			
me(v)	Associations spécifiques, v est associé			
	à code-num selon le $Tableau A8$.			

TABLE A8 – Partie de la table du paramètre "coded-block-pattern". coded-block-pattern (Prédiction Inter)code-num

I ()	
0 (pas de blocs non nuls)	0
16 (bloc de composante continue	1
de chrominance non nul)	
1 (bloc de luminance $8x8$	2
haut-gauche non nul)	
2 (bloc de luminance $8x8$	3
haut-droit non nul)	
4 (bloc de luminance 8x8	4
bas-gauche non nul)	
8 (bloc de luminance $8x8$	5
bas-droit)	
••••	

apparaît fréquemment alors que le type Pred-8x8 (Prédiction d'une partition 8x8 de luminance à partir d'une image précédente) est associé à la valeur code-num=3 car il est moins commun. En résumé, ce codage à longueur variable est défini pour tous les éléments syntaxiques, à l'exception des coefficients de transformée quantifiés qui seront codés en utilisant le CAVLC.

b) Context-based adaptive variable length coding (CAVLC)

Cette méthode est utilisée pour coder les données résiduelles des blocs 4x4 ou 2x2 parcourus en *zigzag (Figure A12*). CAVLC est conçu pour tirer avantage de plusieurs caractéristiques des blocs 4x4 quantifiés :

- Après prédiction, transformation et quantification, les blocs contiennent principalement des zéros. Le *CAVLC* les représente de manière compacte.
- Le plus grand nombre des coefficients après le parcours en zigzag sont souvent des séquences de ± 1 . Le *CAVLC* signale le nombre de ces coefficients (Trailing 1s ou T1s) de manière compacte.
- Le nombre de coefficients non nuls dans les blocs voisins est corrélé. Le nombre des coefficients est codé suivant une table de correspondance; le choix de cette table dépend du nombre des coefficients non nuls.
- L'amplitude des coefficients non nuls tend à être plus grande prés de la composante continue *DC* et faible vers les hautes fréquences. Le *CAVLC* tire avantage de cette caractéristique en adaptant le choix des tables de correspondance suivant les amplitudes codées récemment.



FIGURE A12 – Parcours en zigzag des blocs 4x4 de luminance.

Brièvement, le codage CAVLC d'un bloc de transformation procède comme suit :

- Coder le nombre des coefficients non nuls et le nombre de ± 1 : Le *coeff-token* code le nombre de coefficients non nuls (*TotalCoeffs*) et le nombre de (*T1s*).
- Coder le signe de chaque *T1s* : Le code *trailing-ones-sign-flag* signe la valeur Trailing One (un seul bit code le signe : 0=+, 1=-).
- Coder les niveaux des coefficients non nuls restants : Le code *level-prefix* est la première partie du code des coefficients non nuls (un par coefficient, sauf les ± 1). Le code *level-suffix* est la seconde partie du code des coefficients non nuls (pas toujours présent).
- Coder le nombre total des zéros avant le dernier coefficient : Le code *Totalzeros* code le nombre total de zéros apparaissant après le premier coefficient non nul dans l'ordre du parcours en zigzag (un par bloc).

• Coder le nombre de zéros précédant chaque coefficient non nul : Le code *run-before* code le nombre de zéros précédant chaque coefficient non nul, dans l'ordre inverse du parcours en zigzag [5].

L'efficacité du codage entropique a été encore améliorée par l'adoption d'un autre type de codage "CABAC" au sein des autres profils. Ce codage arithmétique a permis de réduire encore le débit binaire de 10 à 15% des signaux de télévision d'une même qualité en le comparant au CAVLC. De plus, pour améliorer la qualité visuelle des vidéos codées en H.264/AVC tout en réduisant le débit binaire, le standard applique un filtrage (filtre de déblocage de la Figure A1) dont nous allons décrire brièvement son utilité.

A1.9 Filtre de suppression des effets de blocs

Un défaut du codage axé sur le bloc est la visibilité de la structure en blocs. Les bords sont en général reconstitués avec moins de précision que les pixels intérieurs : la pixellisation est l'un des artefacts les plus visibles des méthodes de compression actuelles. Pour cette raison, la norme H.264/AVC définit un filtre de suppression des effets de blocs adaptatif en boucle où la puissance du filtrage est contrôlée par les valeurs de plusieurs éléments syntaxiques. La pixellisation est alors réduite sans affecter outre mesure la netteté du contenu et la qualité subjective est considérablement améliorée. En même temps, le filtre réduit le débit binaire - en général de 5 à 10% tout en produisant la même qualité objective que la vidéo non filtrée. La *Figure A13* illustre les performances du filtre anti-blocs [76].



FIGURE A13 – La performance du filtre anti-blocs des images très compressées, sans et avec filtre.

BIBLIOGRAPHIE

- [1] JM10.2, JM14.0, JM15.0, http://iphome.hhi.de/suehring/tml/ download/old_jm.
- [2] JVT-G012, JVT-O016, ftp3.itu.ch/av-arch/jvt-site.
- [3] ISO/IEC 11172-2 :1991, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbps", Part 2 : Visual. 1991.
- [4] ITU-T Recommendation H.261, "Video codec for audiovisual services at px64 kbit/s", version 1, 1990; version 2. 1993.
- [5] ISO/IEC 13818-2 :1994, "Information technology generic coding of moving pictures and associated audio, Part 2 : Visual". 1994.
- [6] ITU-T Recommendation H.263, "Video coding for low bitrate communication, version 1, 1995; version 2, 1998; version 3". 2000.
- [7] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264 ISO/IEC 14496-10 AVC)". 2003.
- [8] ISO/IEC 14496-2 :1999, "Information technology-coding of audio/visual objects", Part 2 : Visual, version 1, 1999; version 2, 2000; version 3. 2004.
- [9] www.pixeltools.com, "Rate Control and H.264". 2006.
- [10] ISO/IEC JTC/SC29/WG11, "MPEG test model 5". Apr. 1993.
- [11] ITU-R, Methodology for the Subjective Assessment of the Quality of Television Pictures, ITU-R Recommendation BT.500-11, Mar. 2000.

- [12] MPEG Video Group, "MPEG-4 video verification model v18.0", ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio N3908. Pisa, Italy, Jan. 2001.
- [13] ITU-T/SG15, "Video codec test model, TMN8". Portland, June 1997.
- [14] I. Amer, W. Badawy, and G. Jullien. "A proposed hardware reference model for spatial transformation and quantization in H.264". Journal of Visual Communication and Image Representation, 17(2):533–552, April 2006.
- [15] R. Barland and A. Saadane. "Automatic Video Summarization Driven by a Spatio-temporal attention Model". pages Proceedings of SPIE, the International Society for Optical Engineering, Human vision and electronic imaging XIII, pp. 68060Q.1–68060Q.9, 28-31 January 2008, San Jose, California, USA.
- [16] R. E. Bellman. "Dynamic Programming". Princeton, NJ: Princeton University Press, 1957.
- [17] T. Berger. "Rate Distortion Theory". Englewood Cliffs, NJ : Prentice Hall, 1971.
- [18] G. Bjontegaard. "Calculation of Average PSNR Differences between RD- Curves". In ITU-T Q.6/SG16 VCEG 13th Meeting, Document VCEG-M33, Austin, USA, Apr. 2001.
- [19] D. Chai and K. N. Ngan. "Foreground/background video coding scheme". In Proc. of IEEE International Conference on Circuits and Systems, ISCAS '97, Hong Kong, June 1997.
- [20] J. Chen, Y. Qu, and Y. He. "A Fast Mode Decision Algorithm In H.264". In Picture Coding Symposium (PCS2004), San Francisco, CA, USA, 2004.
- [21] Z. Chen and K. N. Ngan. "Recent advances in rate control for video coding". Signal Processing : Image Communication, 22 :2007, 19-38.
- [22] Z. Chen and K. N. Ngan. "Distortion variation minimization in real-time video coding". Signal Process. : Image Commun, 21 :273–279, Apr. 2006.
- [23] Z. Chen and K. N. Ngan. "Towards rate-distortion tradeoff in real-time color video coding". IEEE Trans. Circuits Syst. Video Technol, 17:158–167, Feb. 2007.
- [24] Z. Chen and K.N. Ngan. "Linear rate-distortion models for MPEG-4 shape coding". IEEE Trans. Circuits Systems Video Technol., 14 :869–873, 2004.
- [25] Z. Chen and K.N. Ngan. "Rate-distortion analysis for MPEG-4 binary shape coding". In *IEEE International Symposium on Intelligent Signal Processing and Communications Systems, Hong Kong*, 2005.
- [26] T. Chiang and Y.-Q. Zhang. "A new rate control scheme using quadratic rate-distortion modelling". *IEEE Trans. Circuits Syst. Video Technol.*, 7 :246–250, Feb. 1997.
- [27] T. Cormen, C. Leiserson, and R. Rivest. "Introduction to Algorithms". Cambridge, MA : MIT press, 1990.

- [28] T. M. Cover and J. A. Thomas. "Elements of Information Theory". New York : Wiley, 1991.
- [29] W. Ding and B. Liu. "Rate control of MPEG video coding and recording by ratequantization modelling". *IEEE Trans. Circuits Systems Video Technol.*, 6:12–20, 1996.
- [30] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli. "New full-reference quality metrics based on HVS". In *Proceedings of the Second International* Workshop on Video Processing and Quality Metrics, Scottsdale, USA, 2006.
- [31] H. Everett. "Generalized Lagrange multiplier method for solving problems of optimum allocation of resource". Oper. Res., 11:399–417, 1963.
- [32] Forney. "The Viterbi algorithm". In Proc. IEEE, volume 61, pages 268–278, Mar. 1973.
- [33] B. Fox. "Discrete optimization via marginal analysis". Manage. Sci., 13 :210-216, Nov. 1966.
- [34] H.-M. Hang and J. J. Chen. "Source model for transform video coder and its application-Part I : Fundamental theory". IRE Nat. Conv. Rec. New-York, 4 :287–298, Apr. 1997.
- [35] Z. He, Y. K. Kim, and S. K. Mitra. "Low-delay rate control for DCT video coding via ρdomain source modeling". *IEEE Trans on Circuits and Systems for Video Technol*, 11:928– 940, Aug. 2001.
- [36] Z. He and S. K. Mitra. "A unified rate-distortion analysis framework for transform coding". IEEE Trans on Circuits and Systems for Video Technol, 11 :1221–1236, Dec. 2001.
- [37] C. Huang and C. Lin. "A novel 4-D perceptual quantization modelling for H.264 bit-rate control". *IEEE Trans. On Multimedia*, 9(6) :1113–1124, 2007.
- [38] J. Y. Huang and P. M. Schultheiss. "Block quantization of correlated Gaussian random variables". *IEEE Trans. Commun. Syst.*, 11 :289–296, Sept. 1963.
- [39] I. Hurbain. "The MPEG4/AVC standard : description and basic tasks splitting ". 7 January 2004.
- [40] L. Itti and C. Koch. "Computational modelling of visual attention". Nature Reviews, Neuroscience, 2 :194–203, Mar, 2001.
- [41] L. Itti, C. Koch, and E. Niebur. "A model of saliency-based visual attention for rapid scene analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:1254–1259, Nov 1998.
- [42] M. Jiang and N. Ling. "On enhancing H.264/AVC video rate control by PSNR-based frame complexity estimation". *IEEE Trans Consumer Electron*, 51 :281–286, 2005.
- [43] M. Jiang and N. Ling. "Low-delay rate control for real-time H.264/AVC video coding". IEEE Transactions on Multimedia, 8(3):467–477, 2006.

- [44] M. Jiang, X. Yi, and N. Ling. "Improved frame-layer rate control for H.264 using MAD ratio". In *IEEE Int. Sym. on Circuits Syst, Vol 3*, pages 813–816, Vancouver, Canada, May 2004.
- [45] X. Jing and L. P. Chau. "A Novel Intra-Rate Estimation Method for H.264 Rate Control". In IEEE International Symposium on Circuits and Systems (ISCAC), 2006.
- [46] X. Jing, L. P. Chau, and W. C. Siu. "Frame Complexity-Based Rate-Quantization Model for H.264/AVC Intra-frame Rate Control". *IEEE Signal Processing Letters*, 15 :373–376, 2008.
- [47] N. Kamaci, Y. Altunbasak, and R.M. Mersereau. "Frame bit allocation for the H.264/AVC video coder via Cauchy-density based rate and distortion models". *IEEE Trans on Circuits* and Systems for Video Technol, 15(8) :994–1006, Aug. 2005.
- [48] L. S. Karlsson. "Spatio-temporal pre-processing methods for region-of-interest video coding". PhD thesis, PhD dissertation, Sundsvall, Sweden, 2007.
- [49] H.M. Kim. "Adaptive rate control using non linear regression". IEEE Trans. Circuits Systems Video Technol., 13:432–439, 2003.
- [50] J. Y. Kim, S. H. Kim, and Y. S. Ho. "A frame-layer rate control algorithm for H.264 using rate-dependent mode selection". *Lecture Notes in Computer Science*, PCM, 3768 :477–488, 2005.
- [51] S-K. Kwon, A. Tamhankar, and K.R. Rao. "Overview of H.264/MPEG-4 part 10". Journal of Visual Communication and Image Representation, 17(2):186–216, April 2006.
- [52] K.C. Lai, S.C. Wong, and K. Lun. "A rate control algorithm using human visual system for video conferencing systems". In *Proc. Int. Conf. Signal Processing*, volume 1, pages 656–659, Aug. 2002.
- [53] G. Laroche and B. Pesquet-Popescu. "Codage de vecteurs mouvement par compétition de prédicteurs spatio-temporels dans le standard H.264". In In CORESA, Caen, 2006.
- [54] J. Lee and B. W. Dickinson. "Rate-distortion optimized frame type selection for MPEG encoding". *IEEE Trans. Circuits Syst. Video Technol*, 7:501–510, June 1997.
- [55] J. Lee and B. Jeon. "Fast mode decision for H.264". In Proc. of IEEE International Conference on Multimedia and Expo, Taipei, Taiwan, June 2004.
- [56] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu, and Y. Lu. "Adaptive rate control for H.264". J. Visual Communication and Image Representation, 17(2):376–406, 2005.
- [57] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu, and Y. Lu. "Adaptive Rate Control with HRD Consideration". ISO-IEC/JTC1/SC29/WG11/JVT-H014. In 8 th JVT Meeting, Geneva, Switzerland, 23-27 May, 2003.

- [58] S. C. Lim, H. R. Na, and Y. L. Lee. "Rate control based on linear regression for H.264/MPEG-4 AVC". Signal Processing : Image Communication, 22 :39–58, 2007.
- [59] G. Lin and S. Zheng. "Perceptual importance analysis for H.264/AVC bit allocation". Journal of Zhejiang University SCIENCE A, 2 :225–231, 2008.
- [60] G. Lin, S. Zheng, and L. Zhu. "An improved frame-layer bit allocation scheme for H.264/AVC rate control". Journal Shanghai Jiaotong Univ, 14(2) :143–148.
- [61] L.-J. Lin and A. Ortega. "Bit-rate control using piecewise approximated rate-distortion characteristics". *IEEE Trans. Circuits Systems Video Technol.*, 8:446–459, 1998.
- [62] T. Liu, N. Zheng, W. Ding, and Z. Yuan. "Video attention : learning to detect a salient object sequence". In Proc. ICPR, pages 1–4, 2008.
- [63] S. W. Ma, W. Gao, F. Wu, and al. "Rate control for JVT video coding scheme with HRD considerations". In *Proceedings of IEEE ICIP (IEEE Signal Processing Society)*, pages 793–796, 2003, Barcelona, Spain.
- [64] H.S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky. "Low-complexity transform and quantization in H.264/AVC". *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):598–603, July 2003.
- [65] Z. Min, T. Ikenaga, and S. Goto. "A novel Rate Control Algorithm for H.264/AVC". In The 23rd International Conference On Circuits/Systems Computers and Communications (ITC-CSCC), pages 725–728, 6-9 July 2008, Japan.
- [66] K. Minoo and T.Q. Nguyen. "Perceptual Video Coding with H.264". In Conference on Signals, Systems and Computers (Conference Record of the Thirty-Ninth Asilomar), pages 741–745, November 2005.
- [67] A. G. Nguyen and J.-N. Hwang. "A novel hybrid HVPC/mathematical model rate control for low bit-rate streaming video". Signal Process. : Image Commun, 17:423–440, May 2002.
- [68] N.Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin. "On betweencoefficient contrast masking of DCT basis functions". In Proceedings of the Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VP-QM'07), Scottsdale, Arizona, USA, 25-26 January, 2007.
- [69] A. Ortega and K. Ramchandran. "Rate-distortion methods for image and video compression". IEEE Signal Process. Mag., 15(6):23–50, 1998.
- [70] F. Pan, Z. G. Li, K. P. Lim, X. Lin, S. Rahardja, D. J. Wu, and S. Wu. "Adaptive intraframe quantization for very low bit rate video coding". In Proc. Int. Symp. Circuits and Systems, ISCAS'04, pages 781–784, May 2004.

- [71] N. Ponomarenko, S. Krivenko, V. Lukin, K. Egiazarian, and J-T. Astola. "Lossy Compression of Noisy Images Based on Visual Quality : A Comprehensive Study". *EURASIP Journal on Advances in Signal Processing*, Volume 2010, Article ID 976436 :13 pages, 2010.
- [72] J. Ribas-Corbera and S. Lei. "Rate control in DCT video coding for low-delay communications". IEEE Trans. Circuits Syst. Video Technol., 9 :172–185, Feb. 1999.
- [73] I.E.G. Richardson. "H.264 and MPEG-4 Video Compression : Video Compression for Nextgeneration Multimedia". Edition John Wiley & Sons, 17 october 2003.
- [74] D. A. Robinson. "The mechanics of human smooth pursuit eye movement". J. Physiol, 180 :569-591, Oct. 1965.
- [75] G. Ru, W. Gao, and J. Fu. "An Improved rate control method based on modification at scene change for H.264". Journal WUJNS (Wuhan University Journal of Natural Sciences), 12(6):1047–1050, 2007.
- [76] R. Schäfer, T.Wiegand, and H. Schwarz. "H.264/AVC -La norme qui monte". UER-REVUE TECHNIQUE, SELECTION 2003 :1–10, 2003.
- [77] G. M. Schuster and A. K. Katsaggelos. "Rate-Distortion Based Video Compression". Norwell, MA : Kluwer Academic Publishers, 1997.
- [78] G. M. Schuster, G. Melnikov, and A. K. Katsaggelos. "A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers". *IEEE Trans. Multimedia*, 1 :3–17, Mar. 1999.
- [79] C. E. Shannon. "A mathematical theory of communications". *Bell System Tech*, 27, July and Oct. 1948.
- [80] C. E. Shannon. "Coding theorems for a discrete source with a fidelity criterion". IRE Nat. Conv. Rec, New-York, 4 :142–163, Mar. 1959.
- [81] I-H. Shin, Y.L. Lee, and H. Park. "Rate control using linear rate ρ- model for H.264". IEEE Trans on Circuits and Systems for Video Technol, 19 :441–452, Apr. 2004.
- [82] H. Song and C.-C. J. Kuo. "Rate control for low-bit-rate video via variable-encoding frame rates". *IEEE Trans. Circuits Syst. Video Technol*, 11:512–521, Apr. 2001.
- [83] A. Stoica, M-C. Larabi, and C. Fernandez-Maloigne. "Visual quality enhancement for color images in the framework of the JPEG2000 compression standard". *International Journal* of Robotics and Automation, 20(2):109–122, January 2005.
- [84] G. Sullivan, T.Wiegand, and K-P. Lim. "Joint Model Reference Encoding Methods and Decoding Concealment Methods", ISO IEC TC JTC 1/SC29 N5821, Draft ISO/IEC 14496-5 :2002/PDAM6. 2003.

- [85] G. Sullivan, T. Wiegand, and K.-P. Lim. "Joint model reference encoding methods and decoding concealment methods", JVT-I049. San Diego, Sept. 2003.
- [86] G. J. Sullivan and R. L. Baker. "Efficient quadtree coding of images and video". In Proc. IEEE Int. Conf. Acoustics, Speech and Singal Processing, Toronto, Canada, May 1991.
- [87] G. J. Sullivan and T. Wiegand. "Rate-distortion optimization for video compression". IEEE Signal Process. Mag., 15(6):74–90, 1998.
- [88] P. Sunna. "AVC/H.264-Un système de codage vidéo évolué pour la HD et SD ". UER-REVUE TECHNIQUE, SELECTION 2005 :54–58, 2005.
- [89] C. W. Tang, C. H. Chen, Y. H. Yu, and C. J. Tsai. "Visual sensitivity guided bit allocation for video coding". *IEEE Trans. Multimedia*, 8 :11–18, Feb. 2006.
- [90] L. Tong and K.R. Rao. "Region-of-interest based rate control for low-bit-rate video conferencing". Journal of Electronic Imaging, 15:3, July, 2006.
- [91] Y.-K. Tu, J.-F. Yang, and M.-T. Sun. "Rate-distortion estimation for H.264/AVC coders". In Proc. of IEEE International Conference on Multimedia and Expo, Amsterdam, Netherlands, July 2005.
- [92] A. Vetro, Y. Wang, and H. Sun. "A probabilistic approach for rate-distortion modeling of multiscale binary shape". *IEEE Trans. Image Process*, 12:356–364, 2003.
- [93] A. J. Viterbi and J. K. Omura. "Principles of Digital Communication and Coding". New York : McGraw-Hill, 1979.
- [94] S. Wan, Y. Chang, and F. Yang. "Frame-layer rate control for JVT video coding using improved quadratic rate distortion model". In VCIP, Beijing, China, Jul.2005.
- [95] S. Wan and E. Izquierdo. "A New Rate distortion Model for Rate Control of Video Coding". In 7th International Workshop On Image Analysis for Multimedia Interactive Services WIAMIS, Incheon, S. Korea, 2006.
- [96] C-C. Wang, T-S. Chen, and C-W. Tung. "Fast Intra-Mode Decision in H.264 using Interblock Correlation". In *IEEE International Conference on Image Processing (ICIP)*, Atlanta Marriott Marquis, GA, USA, October 2006.
- [97] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. "Image Quality Assessment : From Error Measurement to Structural Similarity". *IEEE Transactions On Image Processing*, 13(1), January 2004.
- [98] J. M. Wolfe. "Visual search in Attention". H. Pashler, Ed. London, UK : Psychology Press, 1998.
- [99] B. Xie and W. Zeng. "A sequence-based rate control framework for consistent quality real-time video". *IEEE Trans on Circuits and Systems for Video Technol*, 16:56–71, Jan. 2006.

- [100] C-L. Yang, L-M. Po, and W-H. Lam. "A fast H.264 intra prediction algorithm using macroblock properties". In *International Conference on Image Processing (ICIP apos)*, October 2004.
- [101] X. Yi and N. Ling. "Improved H.264 Rate Control by Enhanced MAD-Based Frame Complexity Prediction". Journal of Visual Communication Image Representation, 17:407–424, 2006.
- [102] M. Yin, Y. Xie, F. Guo, and S. Cai. "Improvements on MB-layer Rate Control Scheme for H.264 video using complexity estimation". In Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pages 280–284, July 30-August 01, 2007, China.
- [103] M. Yin and H. Y. Yuan. "A rate control scheme for H.264 video under low bandwidth channel". Journal of Zhejiang University - Science A, 7(6) :990–995, 2006.
- [104] P. Yin, A. M. Tourapis, and J. Boyce. "Fast mode decision and motion estimation for JVT/H.264". In Proc. of IEEE International Conference on Image Processing, Barcelona, Spain, Sept. 2003.
- [105] H. T. Yu, F. Pan, Z. P. Lin, and Y. Sun. "A Perceptual Bit Allocation Scheme for H.264". In IEEE International Conference Multimedia and Expo (ICME), July, 2005.
- [106] W. Yuan, S. Lin, Y. Zhang, W. Yuan, and H. Luo. "Optimum Bit Allocation and Rate Control for H.264/AVC". *IEEE Transactions on Circuits and Systems for Video Technology*, 16:705–715, 2006.
- [107] S.M. Zhou, J. T. Li, J. H. Fei, and Y. D. Zhang. "Improvement on rate-distortion performance of H.264 rate control in low bit-rate". *IEEE Trans. Circuits Sys. Video Technol*, 17(8) :996–1006, Aug. 2007.
- [108] Y. Zhou, Y. Sun, Z. Freng, and S. Sun. "New rate-distortion modeling and efficient rate control for H.264/AVC video coding". Signal Processing : Image Communication, 24 :345– 356, 2009.