



HAL
open science

Stratégies d'encodage pour codeur vidéo scalable

Yohann Pitrey

► **To cite this version:**

Yohann Pitrey. Stratégies d'encodage pour codeur vidéo scalable. Traitement du signal et de l'image [eess.SP]. INSA de Rennes, 2009. Français. NNT: . tel-00461631

HAL Id: tel-00461631

<https://theses.hal.science/tel-00461631>

Submitted on 5 Mar 2010

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

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

THÈSE

présentée devant

L'INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE RENNES

pour obtenir le grade de

DOCTEUR DE L'INSA DE RENNES

Spécialité : TRAITEMENT DU SIGNAL ET DE L'IMAGE

par

Yohann PITREY

Stratégies d'encodage pour codeur vidéo scalable

Équipe d'accueil : Équipe Image et Télédétection - IETR UMR 6064
École Doctorale : MATISSE
Composante Universitaire : INSA de Rennes

Soutenue le 16 septembre 2009

COMPOSITION DU JURY

Président

Atila Baskurt Professeur, INSA de Lyon, LIRIS

Rapporteurs

Béatrice Pesquet-Popescu Professeur, Télécom ParisTech, TSI

Patrick Le Callet Professeur, École Polytechnique de l'Université de Nantes, IRCCyN

Examineurs

Atila Baskurt Professeur, INSA de Lyon, LIRIS

Édouard François Ingénieur de Recherche et Développement, Thomson R&D, Rennes

Marie Babel Maître de Conférences - Co-encadrante de thèse, INSA de Rennes, IETR

Olivier Déforges Professeur - Directeur de thèse, INSA de Rennes, IETR

Table des matières

Table des matières	2
Introduction	11
0.1 Contexte et problématique	11
0.2 Objectifs et motivations	12
0.2.1 Compromis entre performances et complexité algorithmique . . .	12
0.2.2 Qualité d'image constante	13
0.2.3 Régulation de débit inter-couches	13
0.2.4 Travail envisagé	14
0.3 Organisation du document	14
I Étude Bibliographique	15
1 MPEG-4 AVC/H.264	17
1.1 Structure de l'encodeur MPEG-4 AVC/H.264	17
1.2 Macroblocs et codage prédictif	18
1.3 Transformée	19
1.4 Quantification	19
1.5 Codage entropique	22
1.6 Types d'images et groupes d'images hiérarchiques	22
1.7 Différents profils	24
1.8 Conclusion	25
2 MPEG-4 SVC	27
2.1 Scalabilité spatiale	28
2.2 Scalabilité en qualité	29
2.3 Scalabilité temporelle	30

2.4	Prédiction inter-couches	30
2.5	Scalabilité combinée	32
2.6	Conclusion	33
3	Régulation de débit	35
3.1	Problématique	35
3.2	Optimisation débit-distorsion	37
3.2.1	Optimisation Lagrangienne	38
3.2.2	Le dilemme de la poule et de l'œuf	39
3.3	Éléments d'un mécanisme de régulation de débit	40
3.3.1	Formulation des contraintes en fonction du contexte	41
3.3.2	Calcul de budget et stratégies de répartition	41
3.3.3	Différentes échelles de la régulation	43
3.3.4	Respect du budget	44
3.3.4.1	Recherche exhaustive	44
3.3.4.2	Relation entre le débit et le QP	44
3.3.5	Paramètres clés de la relation entre débit et QP	45
3.3.5.1	Entropie des résidus de prédiction	45
3.3.5.2	Macroblocs non codés	46
3.3.5.3	Données d'en-tête dans un macrobloc	46
3.3.6	Régulation de débit et qualité	46
3.3.6.1	Mesures objectives de qualité	47
3.3.6.2	Mesures perceptuelles de qualité	48
3.3.6.3	Minimisation des variations de la qualité	48
3.3.7	Gestion du buffer de transmission	48
3.4	Contextes applicatifs et contraintes supplémentaires	49
3.4.1	Optimisation complexité-performances	49
3.4.2	Scénarios à bas débit et temps réel	50
3.5	Le ρ -domaine, une approche intéressante	51
3.5.1	Relation entre ρ et le QP	51
3.5.2	Modèle de débit	52
3.5.3	Calcul du QP optimal	53
3.5.4	Au niveau image	54
3.5.5	Conclusion	54
3.6	Régulation de débit pour MPEG-4 SVC	55
3.7	Synthèse de l'étude bibliographique	58
3.7.1	Choix des contraintes appliquées	58

<i>Table des matières</i>	5
3.7.2 Régulation de débit niveau image	59
3.7.3 Validation d'un modèle de débit	59
3.7.4 Régulation de débit et qualité constante	60
3.7.5 Régulation à faible complexité algorithmique	60
3.7.6 Exploitation des informations inter-couches	60
II Régulation de débit pour MPEG-4 SVC	63
4 Validation du ρ-domaine pour MPEG-4 SVC	65
4.1 Relation entre ρ et le QP pour MPEG-4 SVC	66
4.2 Linéarité de la relation entre ρ et le débit pour MPEG-4 SVC	66
4.3 Conclusion	68
5 Mécanisme de contrôle de débit en deux passes pour MPEG-4 SVC	73
5.1 Problématique	73
5.2 Contraintes appliquées	74
5.3 Allocation de budget	75
5.3.1 Allocation de budget niveau GOP	75
5.3.2 Allocation de budget niveau image	76
5.4 Respect du budget	76
5.4.1 Passe d'initialisation	77
5.4.2 Calcul du QP optimal	77
5.4.3 Calcul de la valeur initiale du QP	78
5.4.4 Seconde passe d'encodage	79
5.5 Résultats expérimentaux	79
5.5.1 Précision de la régulation de débit	80
5.5.2 Interprétation des erreurs de contrôle	81
5.5.3 Influence de la taille des GOP	83
5.5.4 Qualité des flux reconstruits	83
5.5.5 Temps d'encodage	83
5.6 Conclusion	84
6 Régulation de débit et qualité constante	95
6.1 Qualité constante par GOP	96
6.2 Efficacité de codage des différents types et niveaux d'images	97
6.3 Mesure de l'efficacité de codage d'une image	98
6.4 Répartition du budget du GOP d'après l'efficacité de codage	101

6.5	Résultats expérimentaux	103
6.5.1	Comportement du débit au niveau des images	103
6.5.2	Atténuation des variations de la qualité	103
6.5.3	Évaluation visuelle du gain apporté sur la qualité	104
6.5.4	Influence de la prédiction inter-couches	105
6.5.5	Conclusion	105
7	Contrôle de débit en une passe pour MPEG-4 SVC	115
7.1	Initialisation du modèle de débit	116
7.2	Choix de l'image de référence pour le modèle de débit	118
7.3	Calcul du budget au niveau image	118
7.4	Résultats	119
7.4.1	En termes d'erreur de débit	119
7.4.2	En termes de qualité	120
7.4.3	En termes de temps d'encodage	120
7.5	Conclusion	121
8	Schémas avancés : travaux préliminaires	131
8.1	Poids d'images adaptatifs	131
8.1.1	Limites de la mesure d'efficacité de codage	132
8.1.2	Évolution du poids des types d'images	133
8.1.2.1	Choix de la fonction d'évolution des poids d'images	133
8.1.2.2	Adaptation de la vitesse d'évolution en fonction du niveau temporel	135
8.1.3	Conclusion	136
8.2	Exploitation des dépendances inter-couches	136
8.2.1	Comportements débit-distorsion similaires entre les couches	137
8.2.2	Problème d'échelle entre les couches	137
8.2.3	Choix du QP	138
8.2.4	Conclusion	139
	Conclusion	145
1.1	Bilan sur le travail effectué	145
1.1.1	Validation du ρ -domaine sur MPEG-4 SVC	145
1.1.2	Régulation de débit en deux passes pour MPEG-4 SVC	146
1.1.3	Régulation de débit et qualité constante	146
1.1.4	Régulation de débit en une passe pour MPEG-4 SVC	147

<i>Table des matières</i>	7
1.2 Travaux en cours et perspectives de recherche	148
1.2.1 Qualité constante et poids adaptatifs	148
1.2.2 Prise en compte du buffer de transmission	148
1.2.3 Régulation de débit inter-couches	149
1.2.4 Régulation globale du débit	149
Annexes	153
A Résultats complémentaires pour la régulation en deux passes	153
B Résultats complémentaires pour la régulation à qualité constante	163
C Résultats complémentaires pour la régulation en une passe	173
D Correspondance entre QP initial et contrainte en bits par pixel	183
Bibliographie	192
Table des Figures	193

Introduction

La compression d'une séquence vidéo consiste à réduire la quantité de données nécessaire pour la représenter, tout en cherchant à minimiser les phénomènes de perte de qualité. Les variations à l'intérieur d'une séquence, causées par exemple par des mouvements d'objets par rapport à la caméra ou des changements d'illumination, sont continues dans la majorité des cas. Les données contenues sont par conséquent imprégnées de fortes redondances, qu'il est possible de retirer pour représenter la scène en utilisant une quantité réduite d'informations. De plus, l'oeil humain ne perçoit qu'une partie des informations présentes dans les images, et il est ainsi possible de ne pas retirer les zones les moins pertinentes en limitant l'impact sur la qualité ressentie. L'utilisation conjointe de la suppression des redondances et des informations les moins visibles permet alors de condenser de manière significative les données à représenter.

Cette thèse s'inscrit dans le contexte de la compression de vidéos, dans le but d'apporter plus spécifiquement de nouvelles stratégies d'encodage adaptées à la diffusion de contenus hétérogènes sur les réseaux de communication actuels.

0.1 Contexte et problématique

Grâce à l'évolution des techniques de codage vidéo, accompagnée par les progrès technologiques dans le domaine des télécommunications, l'accès aux contenus vidéo est en plein essor depuis plus d'une dizaine d'années. La multiplication des moyens de transmission (ADSL, réseaux Wi-Fi, réseaux cellulaires, UMTS, *etc*) et la diversification des appareils capables de lire ces contenus (télévisions haute-définition, ordinateurs personnels, téléphones mobiles, *etc*) contraignent les diffuseurs à investir beaucoup de temps et de moyens pour être en mesure de fournir une qualité de vidéo optimale quel que soit le contexte. En effet, la compression vidéo conventionnelle (aussi appelée codage monocouche) est incapable d'adapter la résolution et le débit du flux après encodage en fonction des capacités du réseau et du dispositif de lecture. Chaque source vidéo est alors encodée en autant de versions qu'il existe de cibles à adresser. D'importantes pertes s'ensuivent, tant du point de vue de l'espace de stockage nécessaire pour entreposer tous ces flux, que du point de vue de la surcharge des réseaux occasionnée par la diffusion de ce volume important d'informations.

Le codage vidéo scalable a été développé en réponse à ce besoin d'adaptation des contenus aux différents contextes de diffusion. Un flux vidéo unique est encodé, contentant plusieurs couches de résolutions et de qualités différentes. Les couches sont codées les unes par rapport aux autres, de manière à exploiter les redondances pré-

sentes entre ces différentes versions du flux. Le codage scalable permet alors d'adapter les caractéristiques de la vidéo au moment même de la transmission de la vidéo, en choisissant quelles sont les parties du flux à envoyer en fonction des capacités de la cible.

Indépendamment des problèmes d'adaptation des flux encodés aux capacités des cibles de lecture, les variations du débit de données en sortie de l'encodage peuvent compliquer la transmission sur des canaux de communication. Le dépassement des capacités de transfert d'un canal cause en effet des interruptions dans la lecture de la vidéo, car l'intégralité des données ne peut être transmise à temps. À l'opposé, un débit trop faible ne permet pas de conserver une quantité d'informations suffisante et pénalise la qualité du flux décodé.

Afin de résoudre ces problèmes, des stratégies de régulation de débit sont mises en œuvre. Leur objectif est de moduler la quantité d'informations en sortie de l'encodeur pour obtenir un débit et une qualité optimaux, et ainsi exploiter au mieux les ressources disponibles.

0.2 Objectifs et motivations

Les travaux de cette thèse ont pour but de développer des stratégies de régulation de débit pour le codage vidéo scalable. Notre étude s'appuie sur le standard MPEG-4 Scalable Video Coding (SVC), qui représente actuellement l'implémentation de référence du concept de codage vidéo scalable. Plusieurs axes de recherche ont été suivis pour développer une approche pertinente et innovante de la régulation de débit sur ce standard. Nous présentons ici chacun de ces axes, en précisant leurs motivations et les points clés qui nous ont permis de fixer les objectifs de notre travail.

0.2.1 Compromis entre performances et complexité algorithmique

Il est nécessaire de trouver un compromis entre précision et complexité, de manière à construire des stratégies de régulation utilisables en pratique. Dans des applications où la contrainte de temps est forte, une approche peu complexe est préférable, car elle ne pénalise pas significativement la durée de l'encodage. Pour des applications davantage axées précision, une approche ayant une complexité plus lourde peut être envisagée pour obtenir une plus grande finesse.

La nécessité de faire un choix entre les deux approches amène les deux interrogations suivantes.

Quelle échelle pour la régulation ?

Selon l'échelle de la régulation, la précision et la complexité d'un mécanisme peuvent varier de manière importante. Une régulation effectuée à l'échelle microscopique (au niveau des macroblocs) bénéficie d'une grande précision, car le comportement du débit est modélisé avec une résolution fine, et les erreurs de contrôle peuvent être rapidement compensées. La complexité algorithmique est par contre élevée, car le processus de régulation est répété un grand nombre de fois au cours de l'encodage. En revanche, une régulation faite à l'échelle macroscopique (au niveau des images ou des groupes d'images) bénéficie d'une complexité algorithmique réduite, obtenue au prix d'une précision moins élevée.

Quel modèle de débit utiliser ?

Un des principaux enjeux de la régulation de débit est de trouver un modèle reliant les paramètres de l'encodage au débit des données en sortie de l'encodeur. Selon les outils mathématiques utilisés, les approximations nécessaires et le nombre de paramètres à évaluer, l'efficacité du modèle en termes de précision et de complexité est soumise à une grande variabilité.

0.2.2 Qualité d'image constante

La qualité d'une vidéo et le débit nécessaire pour la représenter sont deux notions antagonistes. Imposer une contrainte sur le débit d'une vidéo peut donc avoir des conséquences néfastes sur la qualité obtenue. Notamment, le respect de la contrainte de débit peut entraîner des variations de la qualité, qui provoquent une gêne sensible chez le spectateur. Il est alors primordial d'inclure des considérations liées à la qualité du signal encodé dans le mécanisme de régulation.

0.2.3 Régulation de débit inter-couches

Une des principales particularités du codage scalable est le fait que les différentes couches sont codées les unes par rapport aux autres, par le biais de la prédiction inter-couches. Ainsi, il existe de fortes relations entre les couches et dans leur comportement face à la régulation de débit. Ces dépendances permettent notamment d'anticiper les erreurs de régulation dans les couches d'amélioration en utilisant les informations obtenues dans la couche de base.

0.2.4 Travail envisagé

L'élaboration d'une méthode de régulation de débit adaptée à la vidéo scalable passe en premier lieu par l'étude des approches existantes dans le contexte du codage vidéo monocouche. La validation d'un modèle de débit performant, à la fois en termes de précision et de simplicité, constituera un des points de départ de notre travail.

Nous proposerons ensuite plusieurs approches de la régulation de débit, adaptées à des contextes applicatifs et à des scénarios d'encodage différents. Un mécanisme de régulation en deux passes, axé sur la précision du contrôle, sera construit dans un premier temps. Sa complexité relativement élevée nous amènera ensuite à proposer une seconde méthode de régulation en une passe, permettant de réduire considérablement la quantité de calculs nécessaires.

Par le biais d'une mesure de la complexité de codage des images, nous élaborons une méthode de répartition du débit permettant de réduire significativement les variations de la qualité. Les phénomènes de scintillement sont éliminés, ce qui améliore sensiblement l'impression visuelle.

0.3 Organisation du document

Ce manuscrit est composé de deux parties distinctes. La première partie est consacrée à l'étude bibliographique des stratégies de régulation de débit dans le codage vidéo. Les méthodes développées pour le codage monocouche seront abordées dans un premier temps, puis nous nous intéresserons aux contributions dédiées au codage vidéo scalable. Une synthèse de cette étude nous amènera ensuite à proposer une approche répondant aux différents besoins et problèmes posés par la régulation pour la vidéo scalable.

La seconde partie présente les différentes approches de régulation de débit mises en place. Le chapitre 4 établit tout d'abord les bases de notre travail en validant un modèle de débit pour le standard MPEG-4 SVC. Fort de cette validation, le chapitre 5 décrit une méthode de régulation en deux passes adaptée à la vidéo scalable. Dans le chapitre 6, des considérations liées à l'optimisation de l'impression visuelle du spectateur seront abordées. Le chapitre 7 présentera un mécanisme en une passe, permettant de réduire considérablement la complexité du processus de régulation. Pour finir, le dernier chapitre du manuscrit présente les travaux en cours et les perspectives envisagées. Premièrement, une amélioration de la méthode présentée au chapitre 6 sera proposée, offrant une meilleure impression visuelle dans le cas du mécanisme de régulation en une passe. La régulation de débit en prenant en considération les dépendances inter-couches sera enfin abordée.

Première partie

Étude Bibliographique

Chapitre 1

MPEG-4 AVC/H.264

Le standard MPEG-4/H.264 est un projet développé conjointement par deux organismes de normalisation : le *ITU-T Video Coding Expert Group* (VCEG) et le *ISO/IEC Motion Picture Expert Group* (MPEG). L'objectif de ce standard est de fournir une compression vidéo efficace et une structure adaptée à son transport sur des réseaux de communications. Il introduit des outils de codage outrepassant amplement les performances des standards précédents comme MPEG-2 [34] ou MPEG-4 Partie 2 [27], ainsi qu'un module d'adaptation au réseau rendant plus robuste sa transmission sur des canaux bruités.

Il existe un ensemble de documents officiels ainsi que plusieurs ouvrages et publications décrivant le standard de manière exhaustive [68, 81]. Dans ce chapitre, nous présentons les principes du codage MPEG-4 AVC/H.264 tout en nous attardant sur les aspects qui présentent un intérêt pour la régulation de débit, qui fait l'objet de cette thèse.

1.1 Structure de l'encodeur MPEG-4 AVC/H.264

L'encodeur MPEG-4 AVC/H.264 utilise un schéma de codage *hybride*, illustré par la Figure 1.1. Les modules de prédiction, de transformée, de quantification et de codage entropique sont décrits dans la suite du chapitre.

À la fin du processus de codage, le flux compressé est segmenté et converti en unités NAL (*Network Abstraction Layer*) pour faciliter sa transmission sous forme de paquets de données [68]. Chaque unité NAL est composée d'une séquence de bits correspondant aux données vidéo, encapsulées avec des renseignements relatifs au contenu de l'unité NAL et les paramètres nécessaires pour le décodage. Ces informations peuvent

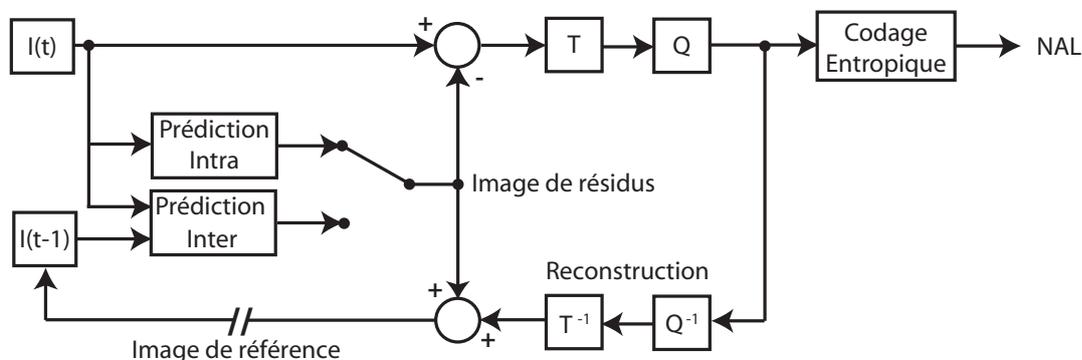


FIGURE 1.1 – Structure de l'encodeur MPEG-4 AVC/H.264.

être utilisées pour la synchronisation du flux avec des pistes audio ou des informations complémentaires telles que le sous-titrage. Dans le cadre de la vidéo scalable, elles sont notamment utilisées pour faciliter le changement entre les différentes couches.

1.2 Macroblocs et codage prédictif

La prédiction spatio-temporelle est un mécanisme largement utilisé, tant dans le codage des images fixes (comme le standard JPEG [29]) que dans le codage des vidéos (MPEG-2 à MPEG-4 *etc*). Elle permet d'enlever une grande partie de la redondance dans les données en ne codant que les différences entre des zones proches. Comme ses prédécesseurs, la norme MPEG-4 AVC/H.264 utilise le *macrobloc* comme unité de base. Les images sont divisées en blocs de taille variable, en fonction de la complexité du mouvement à représenter. Ces blocs sont ensuite prédits les uns par rapport aux autres, grâce à deux types de prédiction. La prédiction *spatiale* ou *intra* image permet de prédire un macrobloc par rapport aux macroblocs voisins déjà codés dans la même image. La prédiction *temporelle* ou *inter* image permet quant à elle de prédire le mouvement d'un macrobloc par rapport à une ou plusieurs images de référence. La prédiction est obtenue en soustrayant le bloc prédicteur au bloc prédit. Le signal résidu produit est ensuite analysé pour trouver la prédiction la plus efficace parmi plusieurs modes disponibles. La prédiction est effectuée sur des macroblocs de 16×16 pixels, et peut être affinée sur des blocs 8×8 ou même 4×4 si aucun autre mode satisfaisant n'est trouvé. Des blocs rectangulaires de 16×8 , 8×8 , 8×4 et 4×4 pixels peuvent également être utilisés dans la prédiction temporelle pour s'adapter à certaines zones de l'image (notamment les zones de forts contours ou de forte texture).

Le signal résidu de la prédiction subit ensuite un ensemble de traitements visant à en retirer autant de redondances que possible. L'étape de prédiction retire en effet une partie de la redondance statistique présente dans les données, mais pas la totalité. Pour enlever davantage de redondance, les résidus de prédiction sont passés dans le domaine fréquentiel par l'étape de transformée.

1.3 Transformée

MPEG-4 AVC/H.264 utilise plusieurs types de transformée en fonction du type de macrobloc à coder [68]. Ils sont tous basés sur une transformée entière en cosinus discret (*Integer Discrete Cosine Transform* : IDCT), qui permet d'obtenir une représentation décorrélée et compacte des données. Elle décompose en effet le signal en composantes fréquentielles indépendantes les unes des autres et dont les principales valeurs sont concentrées sur un petit nombre de coefficients. L'étape de transformée est suivie de l'étape de quantification, qui introduit une perte d'informations dans les données d'origine dans le but d'atteindre des taux de compression plus élevés. Ces deux étapes du codage sont relativement coûteuses en termes de calculs. Elles peuvent être couplées l'une à l'autre de manière à limiter le nombre d'opérations binaires nécessaires, comme le décrit la section suivante.

1.4 Quantification

La quantification constitue une étape très importante du processus d'encodage vis à vis de notre problématique, car elle permet de faire varier le débit nécessaire pour coder le signal en ajustant la quantité de données perdues. Nous décrivons donc cette étape de manière plus détaillée, afin de mettre en évidence le point d'ancrage de la régulation de débit dans le processus d'encodage.

Lors de l'étape de quantification, l'ensemble des valeurs possibles du signal d'entrée est découpé en intervalles d'une longueur qui dépend de la précision désirée. Une table représentant un index de ces intervalles est ensuite construite, et chaque valeur du signal est codée par l'indice de l'intervalle auquel elle appartient. En modulant la taille des intervalles, il est possible de contrôler la quantité d'informations perdues lors de la quantification.

Le standard MPEG-4 AVC/H.264 utilise un mécanisme de quantification scalaire [68]. Dans l'encodeur de référence, la quantification est directement couplée avec la transformée pour minimiser le nombre de multiplications à effectuer. Ainsi, la quasi

totalité de l'opération de transformation et de quantification peut être réalisée en n'utilisant que des additions et des décalages binaires, ce qui diminue considérablement la complexité des calculs. Ces opérations sont décrites dans la suite de cette section.

À partir d'un bloc de résidus de prédiction X , le bloc transformé Y est donné par

$$Y = AXA^T, \quad (1.1)$$

où A est la matrice de transformée et A^T est sa matrice transposée. Chaque coefficient Y_{ij} du bloc transformé est ensuite quantifié pour donner un coefficient Z_{ij} de la manière suivante :

$$Z_{ij} = \text{round} \left(\frac{Y_{ij}}{qstep_{ij}} \right), \quad (1.2)$$

où $qstep_{ij}$ est le *pas de quantification* utilisé. C'est en faisant varier ce pas de quantification que l'on peut moduler la quantité de données perdues.

Pour rendre plus rapide son enchaînement avec l'opération de quantification, la transformée est séparée en une étape de "transformée partielle", suivie d'une opération de décalage. Le bloc Y est décomposé en un bloc W de coefficients non décalés que l'on multiplie par une matrice constante de facteurs de décalage PF donnée dans [68] :

$$Y_{ij} = W_{ij} \times PF_{ij}. \quad (1.3)$$

La valeur d'un coefficient quantifié Z_{ij} est donc obtenue de la manière suivante :

$$Z_{ij} = \text{round} \left(W_{ij} \times \frac{PF_{ij}}{qstep_{ij}} \right). \quad (1.4)$$

Les valeurs possibles de $qstep$ sont disposées dans une table indexée par le *paramètre de quantification* (*Quantization Parameter* : QP), qui varie entre 0 et 51 (0 signifiant une perte minimale et 51 la quantification la plus forte). En pratique, c'est sur le QP que l'on agit pour faire varier la quantité d'informations perdues. Il est bon de souligner la différence entre le pas de quantification $qstep$ et le paramètre de quantification q . Si l'amalgame entre ces deux notions est courant, on s'intéresse généralement à la valeur du paramètre de quantification, qu'il est possible de contrôler directement.

Afin de remplacer l'opération de division par un ensemble de décalages binaires, on pose

$$\frac{PF_{ij}}{qstep_{ij}} = \frac{MF_{ij}}{2^{qbits}},$$

avec

$$qbits = 15 + \lfloor q/6 \rfloor, \quad (1.5)$$

où q est le QP correspondant à la valeur de $qstep_{ij}$. Les valeurs des facteurs de décalage PF et MF en fonction de la position (i, j) du coefficient sont données dans [68]. Ainsi, la valeur du coefficient quantifié Z_{ij} peut être calculée rapidement à l'aide d'une succession d'opérations binaires simples sous la forme

$$Z_{ij} = \text{round} \left(Y_{ij} \times \frac{MF}{PF \times 2^{qbits}} \right). \quad (1.6)$$

L'opération de quantification conduit à l'annulation des coefficients dont les valeurs ne sont pas significatives dans le signal transformé. Dans MPEG-4 AVC/H.264, un coefficient transformé Y_{ij} est codé comme un zéro si sa valeur se trouve en dessous d'un seuil qui caractérise la zone morte du mécanisme de quantification. Pour prendre en compte des aspects psycho-visuels liés à la fois à la prédiction et à la transformée, ce seuil dépend de la valeur du QP, de la position du coefficient dans le macrobloc et du mode de prédiction (intra ou inter) utilisés :

- plus la valeur du QP est grande, plus le seuil en dessous duquel un coefficient est annulé est élevé ;
- les coefficients situés près du coin supérieur gauche du bloc transformé représentent les basses fréquences, qui contiennent les structures principales du signal d'origine ; ils sont moins quantifiés que les coefficients proches du coin inférieur droit du bloc qui représentent les hautes fréquences, qui contiennent quant à elles des détails moins visibles ;
- les prédictions *intra* et *inter* n'ont pas la même efficacité, donc les coefficients des blocs intra sont quantifiés moins fortement que ceux des blocs inter.

Soit un coefficient transformé Y_{ij}^m situé à la position (i, j) dans un macrobloc m . On appelle $t(i, j, q)$ la valeur du seuil de la zone morte du mécanisme de quantification pour un QP égal à q . Les valeurs de $t(i, j, q)$ pour chaque position dans un macrobloc et chaque valeur de QP sont données dans [71]. Nous définissons ici $dz(Y_{ij}^m, i, j, q)$ comme étant la fonction caractéristique de la zone morte :

$$dz(Y_{ij}^m, i, j, q) = \begin{cases} 1 & \text{si } Y_{ij}^m < t(i, j, q) ; \\ 0 & \text{sinon.} \end{cases} \quad (1.7)$$

Finalement, l'opération de seuillage est appliquée à chaque coefficient quantifié, dont la valeur est donc obtenue par :

$$Z_{ij} = Z_{ij} \cdot dz(Y_{ij}^m, i, j, q). \quad (1.8)$$

La fonction $dz(Y_{ij}^m, i, j, q)$ sera notamment utilisée par la suite pour calculer le pourcentage de coefficients nuls dans un macrobloc après l'étape de quantification.

Nous verrons que cette quantité, appelée ρ , peut permettre de trouver une relation simple et efficace entre le QP et le débit nécessaire pour coder l'image, et qu'elle peut représenter un grand intérêt pour la régulation de débit.

L'étape de quantification revêt une importance supplémentaire car elle est impliquée dans le processus de prédiction spatio-temporelle décrit plus tôt. En effet, la prédiction se base sur la version reconstruite du macrobloc de référence, obtenue en effectuant les opérations inverses de quantification, de transformée et de prédiction. La présence de cette boucle dans le processus d'encodage, souvent appelée le *dilemme de la poule et de l'œuf*, aura une influence primordiale sur l'efficacité de la régulation de débit et sera abordée dans la suite de notre étude bibliographique.

1.5 Codage entropique

Après la quantification, l'ensemble des données du flux est encodé à l'aide de codes binaires à longueur variable. Ces codes prennent en compte la probabilité d'apparition d'un motif dans le flux pour adapter le nombre de bits nécessaire pour le coder. MPEG-4 AVC/H.264 fournit deux types de codage binaire : les *codes à longueur variable avec adaptation de contexte* (CAVLC) et le *codage binaire arithmétique avec adaptation de contexte* (CABAC) [53]. Ces deux techniques permettent de coder plus efficacement les signaux restant à la fin du processus d'encodage. Le CAVLC est basé sur des outils assez simples et jouit d'une relative rapidité de calculs. Le CABAC repose quant à lui sur des outils plus avancés qui permettent d'obtenir des performances 10 à 20% supérieures au CAVLC. Sa complexité algorithmique est cependant plus élevée et peut ralentir le processus d'encodage. Le flux binaire ainsi produit peut enfin être transmis sur un canal de communication pour un décodage immédiat ou entreposé sur un média de stockage pour une lecture différée.

1.6 Types d'images et groupes d'images hiérarchiques

Dans MPEG-4 AVC/H.264, trois types d'images sont différenciés en fonction des outils qu'ils utilisent. Les images Intra (ou images I) n'utilisent que la prédiction spatiale. Elles permettent notamment d'initialiser le codage d'une séquence car elles sont codées indépendamment des autres images. Les images Inter (ou images P) utilisent la prédiction intra et inter image, et exploitent les redondances temporelles dans les séquences d'images. Elles jouissent d'une meilleure efficacité de codage que les images I , mais le calcul de la prédiction inter-images est plus coûteux. Enfin, les images Bi-directionnelles

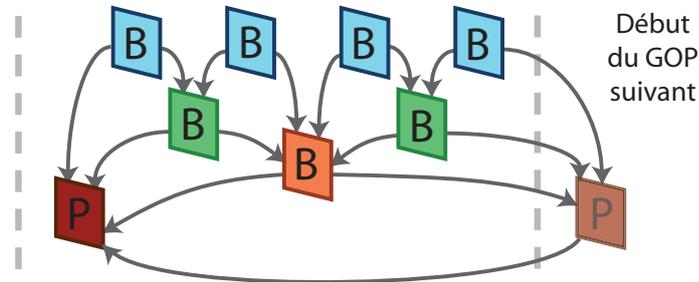


FIGURE 1.2 – Structure de GOP hiérarchique proposée dans MPEG-4 AVC/H.264.

(ou image B) utilisent un mécanisme de prédiction des macrobloccs par rapport à deux images de référence. La prédiction est effectuée à l'aide d'une interpolation entre les deux images de référence, ce qui produit un signal résidu généralement très faible. Les images B ont une efficacité de codage encore supérieure à celle des images P , mais leur construction requiert une quantité supérieure de calculs.

Les flux MPEG-4 AVC/H.264 sont découpés en Groupes d'Images (*Groups Of Pictures*, GOP). La structure des groupes d'images peut varier au long du flux pour s'adapter aux changements d'activité [9], ou rester constante pour limiter la quantité de calculs nécessaires. Une structure particulièrement intéressante est celle des GOP *hiérarchiques* [22, 21] (Figure 1.2). Des images B sont encodées en *niveaux temporels* successifs selon une structure pyramidale basée sur deux images P . Cette structure permet notamment d'adapter la fréquence temporelle de la séquence d'images, en ne décodant qu'une partie de la pyramide. Elle permet également d'obtenir un codage plus efficace des images B de haut niveau temporel, car les images de référence utilisées sont très proches, ce qui rend la prédiction temporelle plus précise.

À intervalles réguliers, l'image P au début du GOP est remplacée par une image I . Ce changement de structure permet de fournir des points d'accès dans le flux, étant donné que les images I peuvent être décodées indépendamment des autres images. Cet intervalle, appelé l'*Intra Period*, doit être un multiple de la taille du GOP.

Une conséquence importante de l'utilisation de la structure hiérarchique est que les images ne sont pas encodées dans leur ordre de lecture. En effet les dépendances entre les images B et les images qu'elles utilisent comme référence contraignent l'encodeur à traiter les images par niveau temporel croissant, comme le montre la Figure 1.3. L'image P qui entame le GOP suivant est donc encodée avant les images du GOP courant. Les niveaux d'images B sont ensuite encodés les uns après les autres. Cette structure de GOP permet de limiter l'intervalle de temps pendant lequel une image peut être utilisée comme référence à la taille d'un GOP. Ceci n'est pas nécessairement vrai

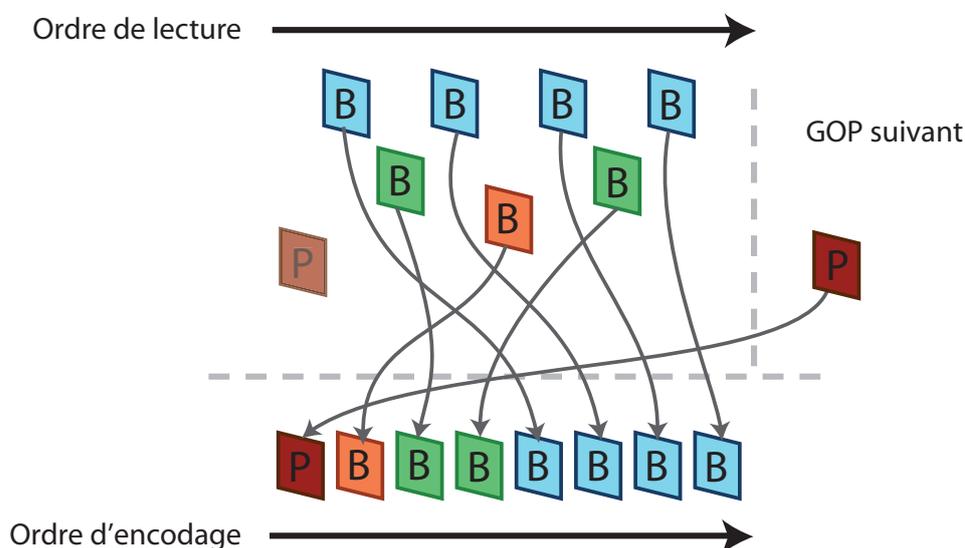


FIGURE 1.3 – Ordre d'encodage des images à l'intérieur d'un GOP hiérarchique.

pour des structures de GOP utilisant les images à références multiples, couramment utilisées pour la vidéo monocouche, mais permet de limiter l'espace mémoire nécessaire pour décoder un GOP entier.

1.7 Différents profils

Additionnellement, le standard définit différents *profils*, limitant les outils utilisés pour produire des flux s'adaptant à des contextes d'encodage ou de décodage avec des capacités de calculs restreintes. Il existe en effet trois profils, destinés à différents types de décodeurs. Le profil le plus simple (ou *Baseline Profile*) est destiné aux appareils ayant une faible puissance de calculs et proscrit par exemple l'utilisation des images B et du CABAC. Le profil intermédiaire (ou *Main Profile*) permet l'utilisation de la plupart des outils de codage disponibles dans MPEG-4 AVC/H.264, à l'exception de fonctionnalités avancées comme le découpage adaptatif des images en tranches (ou *slices*) indépendantes. Enfin, le profil étendu (ou *Extended Profile*) utilise tous les outils disponibles, tout en étant réservé à des applications hautes performances peu répandues.

1.8 Conclusion

Le standard MPEG-4 AVC/H.264 a apporté des avancées conséquentes dans l'efficacité du codage vidéo. Les outils de prédiction spatiale et temporelle, combinés avec les étapes de transformée/quantification et de codage entropique permettent d'atteindre des taux de compression surpassant largement ceux des standards précédents.

Pour notre problématique de régulation de débit, il est bon de souligner l'importance de l'étape de quantification dans le processus de contrôle. En effet, c'est en agissant sur la valeur du paramètre de quantification (QP) qu'il est possible de faire varier la quantité de données perdues et par conséquent de faire varier le débit et la qualité obtenus à la sortie de l'encodeur. Un problème majeur consistera donc à trouver un lien entre le nombre de bits nécessaire pour coder une image ou un macrobloc et la valeur du QP à utiliser. Le chapitre suivant est consacré à la description du standard MPEG-4 SVC, qui constitue le cadre de notre travail.

Chapitre 2

MPEG-4 SVC

Avec la diversification des cibles pour la lecture de vidéos, assurer une qualité optimale quel que soit le contexte de diffusion est devenu un problème de premier ordre. Plusieurs facteurs peuvent varier, comme la taille de l'écran (téléphones portables, téléviseurs standard, téléviseurs Haute-Définition, *etc*), la capacité de transfert du réseau utilisé (réseau cellulaire, ADSL, sans-fil, fibre optique, *etc*), ou la puissance de calculs de l'appareil qui lit la vidéo (DSP, processeur multicœur, *etc*). La disponibilité du réseau peut également varier, à cause de phénomènes de congestion ou d'éloignement de la source d'émission du signal par exemple, source de problèmes lors de la transmission du flux vidéo.

Sans le codage vidéo scalable, un même flux est encodé plusieurs fois séparément dans différentes résolutions adaptées à l'ensemble des cibles visées. Ce mécanisme, appelé *simulcast* (pour *simultaneous broadcast* : diffusion simultanée), représente une perte considérable en termes d'espace de stockage et de bande passante. En effet, les contenus encodés sont relativement proches, et il existe de fortes redondances statistiques entre les différentes versions de la même séquence vidéo. Bien entendu, le codage conjoint de ces différentes résolutions, comparé à un encodage en simulcast, doit mener à de meilleures performances en termes de débit et de qualité. Le codage vidéo scalable rend possible cette économie en encodant les différentes résolutions comme des couches d'un même flux vidéo et en autorisant la prédiction de mouvement et de texture entre les couches. Les redondances inter-couches sont alors exploitées, ce qui permet de gagner en efficacité de codage.

Initié en 2003, le travail de la *Joint Video Team* (JVT), jonction du comité MPEG de l'Organisation Internationale pour la Standardisation (ISO) [26] et de l'Union Internationale des Télécommunications (ITU) [28], a mené à l'élaboration du *Joint Scalable Video Model* (JSVM) et d'une norme de codage vidéo scalable appelée

MPEG-4 Scalable Video Coding (SVC). Cette norme a été finalisée en 2007 après des amendements visant à simplifier le codage, en reléguant au rang d'annexes les fonctionnalités les plus problématiques.

Pour schématiser, MPEG-4 SVC est une extension de la norme MPEG-4 AVC/H.264 assurant une adaptabilité du contenu vidéo à différents contextes de décodage. Un flux MPEG-4 SVC contient plusieurs couches vidéo de résolutions différentes. La couche de plus faible résolution, appelée *couche de base*, est compatible avec le codage mono-couche MPEG-4 AVC/H.264. Les couches supérieures, appelées *couches d'amélioration* peuvent ensuite utiliser les informations de la couche de base grâce à des mécanismes de *prédiction inter-couches*, permettant d'exploiter les redondances entre les différents niveaux du flux. Les outils de transformée, de quantification et de codage entropique restent identiques à ceux utilisés dans MPEG-4 AVC/H.264.

La norme autorise trois types de scalabilité : *spatiale*, *temporelle* et en *qualité*, pour pouvoir adapter la résolution du flux aux différents contextes de lecture. Nous présentons ici les principes du codage MPEG-4 SVC, sans prétendre en offrir une description exhaustive. Pour plus de précisions, plusieurs descriptions plus détaillées du standard existent [70, 82], et la spécification de la norme est disponible dans le document *JVT-X202* [39]. De plus, le logiciel de référence fourni par le JVT, utilisé pour mettre en pratique les travaux de cette thèse, est disponible à l'adresse indiquée en [24].

2.1 Scalabilité spatiale

La scalabilité spatiale est obtenue à l'aide d'une structure pyramidale de résolutions (Figure 2.1). Chaque résolution permet d'adapter la taille du flux à une taille d'écran différente. Il existe deux types de scalabilité spatiale : *dyadique* et *étendue*. La scalabilité spatiale dyadique permet de doubler les dimensions d'une couche d'amélioration par rapport à sa couche de base. Son principe de fonctionnement a l'avantage d'être simple mais elle ne permet pas une grande souplesse dans le choix des résolutions des couches. La scalabilité spatiale étendue (*Extended Spatial Scalability*, ESS) autorise quant à elle un rapport de taille libre entre une couche d'amélioration et sa couche de base. Elle permet également à une couche d'amélioration d'être une version élargie, et non agrandie (en termes de résolution), de sa couche de base (Figure 2.2). Elle offre ainsi une grande souplesse dans l'adaptation de la taille des images, qui se fait au prix d'une complexité de fonctionnement élevée.

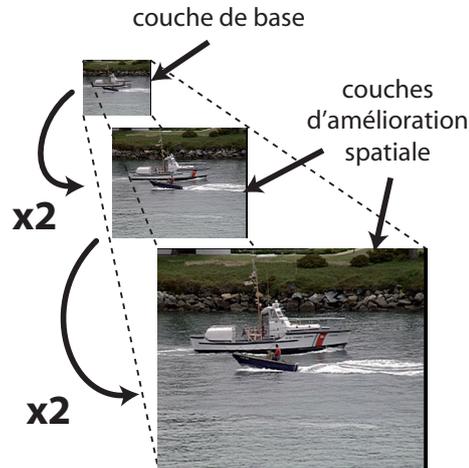


FIGURE 2.1 – Scalabilité spatiale dyadique

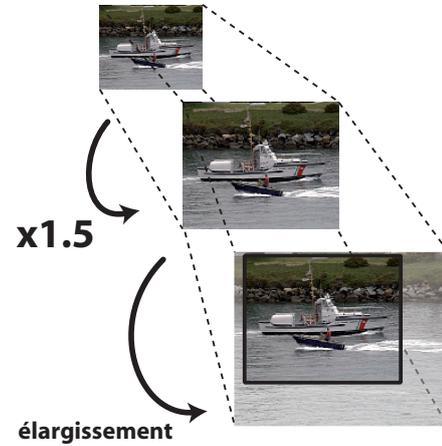


FIGURE 2.2 – Scalabilité spatiale étendue

2.2 Scalabilité en qualité

La scalabilité en qualité permet de fournir plusieurs niveaux de flux ayant une qualité croissante. Il est alors possible de choisir le niveau de fidélité souhaité en fonction des ressources disponibles (bande passante, capacités de décodage, *etc*). Il existe trois types de scalabilité en qualité. La scalabilité à *grain grossier* (*Coarse Grain Scalability* : CGS) raffine les résidus de prédiction par paquets successifs. Elle offre un nombre de niveaux de qualité fixé à l'avance au moment de l'encodage et bénéficie d'une complexité de calculs relativement peu élevée. La CGS se base sur les mêmes outils que la scalabilité spatiale, excepté le fait que les couches ont toutes les mêmes dimensions. La scalabilité à *grain fin* (*Fine Grain Scalability* : FGS) effectue un codage par raffinements progressifs des résidus de prédiction et permet d'adapter le niveau de qualité de manière arbitraire. La complexité des mécanismes mis en jeu et son manque d'adéquation avec les besoins des dispositifs de décodage actuels ont finalement mené la FGS à être retirée du standard et proposée en tant qu'annexe. Enfin, la scalabilité à *grain moyen* (*Medium Grain Scalability* : MGS) est un intermédiaire entre la scalabilité à grain grossier à la scalabilité à grain fin. Elle permet un ajustement du niveau de qualité plus précis que la CGS, tout en étant moins complexe que la FGS. Cependant ses applications restent limitées, et en pratique seule la CGS est implantée.

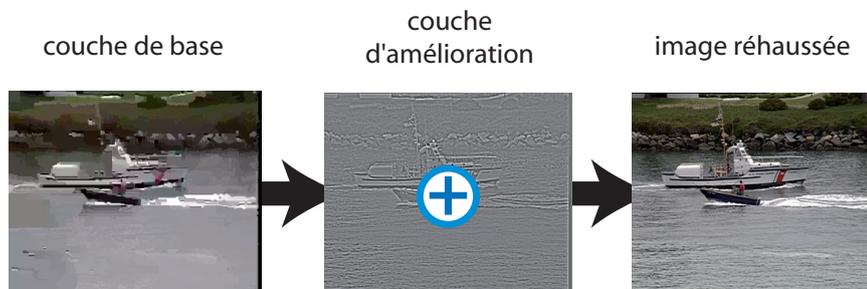


FIGURE 2.3 – Scalabilité en qualité.

2.3 Scalabilité temporelle

La scalabilité temporelle permet d'ajuster la fréquence temporelle du flux vidéo pour réduire l'intervalle entre deux images consécutives et améliorer la fluidité du mouvement (Figure 2.4). Elle est assurée par la structure de GOP hiérarchique présentée dans le chapitre précédent (Figure 1.2). La fréquence temporelle est modifiée en ne décodant qu'une partie des niveaux de la pyramide et l'ajout d'un niveau hiérarchique multiplie par deux le nombre d'images par seconde.

La scalabilité temporelle est déjà disponible dans MPEG-4 AVC/H.264 [22]. La différence dans MPEG-4 SVC réside dans le fait que les images portent un identifiant différent en fonction de leur couche d'appartenance, afin de séparer les niveaux de scalabilité. La structure de GOP hiérarchique est par conséquent distribuée parmi les différentes couches temporelles.

2.4 Prédiction inter-couches

Sans la prédiction inter-couches, encoder plusieurs couches dans un flux unique avec MPEG-4 SVC est équivalent en termes de débit à encoder plusieurs flux de résolutions différentes avec MPEG-4 AVC/H.264 (*cf.* simulcast). En effet, le codage des couches se fait alors indépendamment les unes des autres, et les redondances inter-couches ne sont pas exploitées. L'unique avantage que représente le codage scalable dans ce cas par rapport au simulcast est le fait de pouvoir changer la couche à décoder plus aisément qu'en manipulant plusieurs flux monocouches séparés. En effet, les unités NAL (Network Abstraction Layer, *cf.* chapitre 1) de l'ensemble des couches sont encodées par ordre de lecture. Il est donc plus aisé de changer de couche dans le même flux scalable que de changer entre plusieurs flux indépendants.

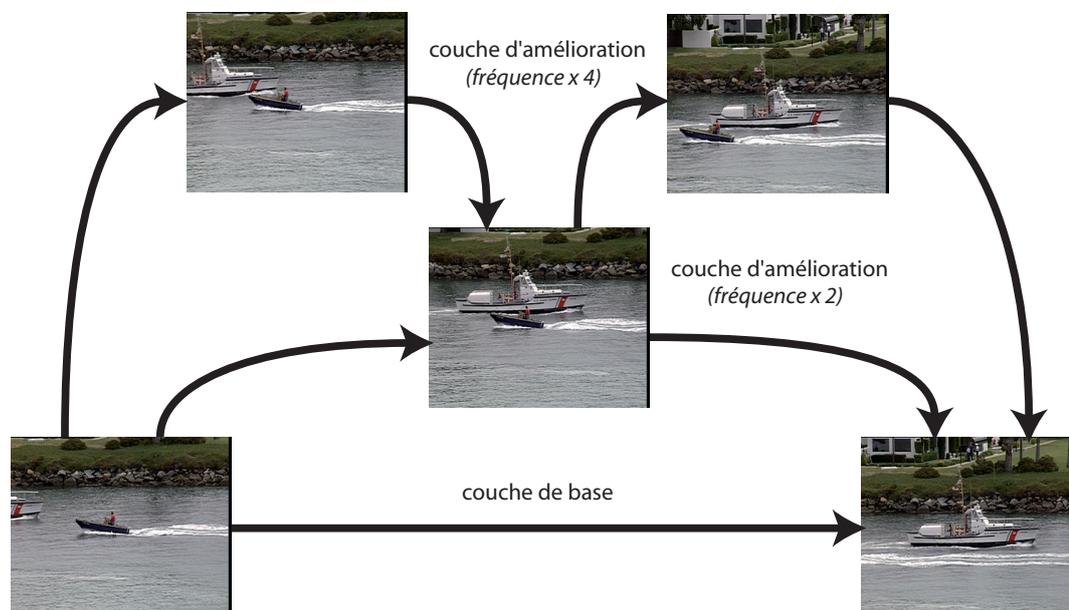


FIGURE 2.4 – Scalabilité temporelle.

La prédiction inter-couches permet de réutiliser les informations de prédiction d'une couche à l'autre pour gagner en efficacité de codage par rapport au simulcast. Il existe deux modes de prédiction inter-couches : *imposé* et *adaptatif*. Dans le mode imposé, les informations de la couche de base sont toujours utilisées pour prédire les macroblocs de la couche d'amélioration. Ce mode permet de réduire la quantité de calculs effectués en fixant le choix pour la prédiction mais peut amener à une mauvaise prédiction. Avec le mode adaptatif, la prédiction inter-couches n'est utilisée que si elle représente un gain par rapport à la prédiction du macrobloc depuis la couche elle-même. L'encodeur de référence permet en outre de désactiver la prédiction inter-couches. Cette possibilité existe avant tout pour des raisons de validation de l'implémentation, et ne présente aucun intérêt du point de vue des performances de codage.

La prédiction inter-couches réutilise deux types d'informations de prédiction :

- **Prédiction des vecteurs de mouvement** : Les vecteurs de mouvement de la couche de base sont utilisés pour prédire le mouvement dans les couches d'amélioration. Dans les couches d'amélioration, seule la différence avec les vecteurs de mouvement de la couche de base est codée. Dans le cas de la scalabilité spatiale, les vecteurs de mouvement sont sur-échantillonnés pour prendre en compte la différence de résolution entre les couches.

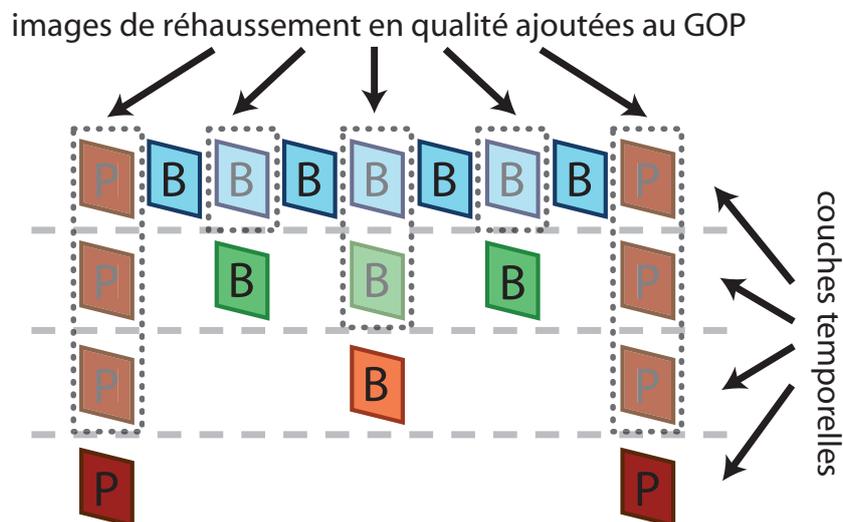


FIGURE 2.5 – Structure de GOP hiérarchique utilisée pour la scalabilité combinée temporelle/qualité dans MPEG-4 SVC.

- **Prédiction des résidus :** Les résidus de prédiction sont codés par rapport aux résidus de la couche de base. Il est également possible de prédire un macrobloc par rapport à la version reconstruite de son homologue dans la couche de base. Ce mode de prédiction implique cependant des restrictions sur la structure du flux puisque la couche de base doit avoir été décodée avant la couche d'amélioration.

2.5 Scalabilité combinée

Il est possible de combiner les trois types de scalabilité précédemment évoqués. Une couche peut être par exemple une amélioration à la fois en fréquence temporelle et en qualité de sa couche de base. En réalité, dans le cas de la scalabilité temporelle, l'encodeur de référence de MPEG-4 SVC ajoute systématiquement un rehaussement CGS. La structure des GOP hiérarchiques est par conséquent modifiée, comme le montre la Figure 2.5. Des images de rehaussement en qualité sont ajoutées en même temps que les images B hiérarchiques dans les couches d'amélioration. Cette structure permet notamment de décoder les couches d'amélioration sans avoir besoin des images de la couche de base lorsque la prédiction inter-couches est désactivée. Dans la suite, nous parlerons simplement de scalabilité temporelle lorsque la rigueur voudrait qu'elle soit appelée *scalabilité combinée temporelle/qualité*.

2.6 Conclusion

En réutilisant des concepts introduits par MPEG-4 AVC, le nouveau standard MPEG-4 SVC fournit trois types de scalabilité qui permettent d'adapter la résolution, la fréquence temporelle et la qualité d'un flux à des cibles hétérogènes. Cette capacité d'adaptation représente un gain considérable en termes de temps d'encodage et de ressources de stockage ou de transmission. Les flux scalables peuvent en outre être adaptés à leur cible au moment de la transmission ou du décodage, ce qui leur procure une grande flexibilité. Les tests effectués pour évaluer les performances de MPEG-4 SVC ont montré son intérêt face au simulcast [82]. Les outils de prédiction inter-couches compensent la perte en efficacité de codage introduite par la scalabilité pour obtenir des performances proches du codage monocouche.

Depuis les débuts du standard, les travaux sur MPEG-4 SVC se sont concentrés sur trois aspects de la chaîne d'encodage-transport-décodage du flux. Le premier aspect consiste à améliorer les performances du codage scalable en modifiant la structure de l'encodeur, par des méthodes de prédiction plus efficaces ou des algorithmes plus rapides pour l'estimation de mouvement [3, 83, 1]. Ces travaux convergent vers une version stable de l'encodeur MPEG-4 SVC, et les avancées majeures dans les techniques de codage se font moins nombreuses. Le deuxième aspect concerne la transmission du flux sur des canaux de communication bruités, et des techniques rendant le flux plus robuste aux erreurs de transmission ont notamment été proposées [59, 15, 2]. Au niveau du décodage enfin, l'optimisation des performances permet d'atteindre un décodage en temps réel pour des flux contenant trois couches spatiales dyadiques de résolution allant de $320 * 180$ à $1280 * 720$ pixels [25].

Du point de vue de leur structure, les signaux SVC et AVC sont relativement proches, car une grande partie du processus d'encodage est commune aux deux standards. Cependant, la prédiction inter-couches et les GOP hiérarchiques changent le comportement des signaux SVC en termes de débit et de qualité, et plus particulièrement dans les couches d'amélioration. Ces différences de nature sont à prendre en compte lors de l'élaboration de techniques de régulation de débit pour la vidéo scalable. Nous verrons dans le chapitre suivant que la plupart des contributions dans le domaine se basent sur des approches existantes pour MPEG-4 AVC et les étendent à la vidéo scalable. La validation des modèles utilisés dans un contexte scalable représentera par conséquent une étape essentielle de la construction d'une stratégie de régulation.

Chapitre 3

Régulation de débit

Ce chapitre est consacré à l'étude des stratégies de régulation de débit existantes pour le codage vidéo. Dans un premier temps, nous présenterons la problématique liée à la régulation de débit. Nous étudierons ensuite les outils mis en jeu dans un processus de régulation et les éléments qui le composent. Les approches adaptées à la vidéo monocouche seront d'abord présentées, en mettant notamment l'accent sur une approche particulièrement intéressante du point de vue de l'efficacité et de la complexité algorithmique, appelée le ρ -domaine. Nous passerons ensuite en revue les stratégies développées pour MPEG-4 SVC, ainsi que les spécificités du contrôle de débit pour la vidéo scalable. Pour finir, nous effectuerons une synthèse de cette étude bibliographique et décrirons les choix que nous avons faits pour répondre à la problématique de cette thèse.

3.1 Problématique

La régulation de débit a pour but de contrôler certains paramètres de l'encodage pour obtenir un flux qui respecte des contraintes imposées. Il s'agit généralement de contraintes de débit liées à la transmission ou au stockage du flux, ou de contraintes liées à la qualité désirée pour le flux vidéo décodé.

Dans les systèmes de codage d'images et de vidéos, le taux de compression est variable en fonction de la complexité des données à représenter, et de la quantité de données que l'on accepte de perdre. Les variations d'activité dans le contenu d'une séquence vidéo causent des variations de l'efficacité du codage, qui se traduisent par des variations du débit à la sortie de l'encodeur.

La Figure 3.1 illustre les variations du débit observées sur quelques séquences d'images classiques encodées avec MPEG-4 AVC/H.264, sans effectuer de régulation de débit. On y observe différents profils d'activité, traduisant les différents types de conte-

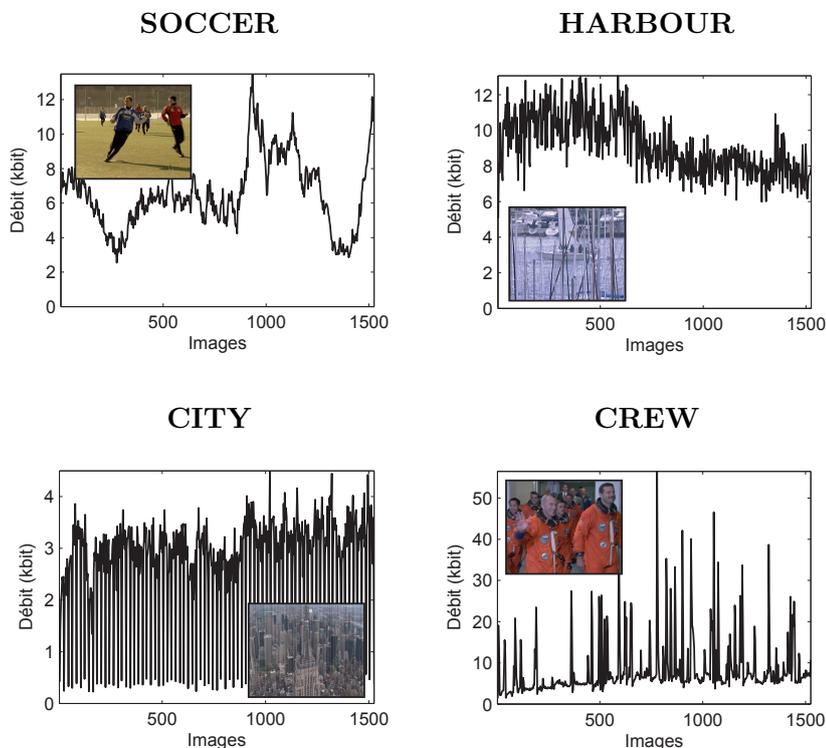


FIGURE 3.1 – Variations du débit observées sans mécanisme de régulation (tracés obtenus en encodant un flux MPEG-4 AVC/H.264 composé exclusivement d’images P).

nus des séquences d’images. Les séquences SOCCER et HOCKEY comportent beaucoup de mouvement, ce qui se traduit par des variations importantes du débit au cours de la séquence. La séquence CREW contient quant à elle de nombreux changements de luminosité, causés par des flashes lumineux très courts. Ils produisent des changements brutaux de l’activité très localisés dans le temps. Les séquences HARBOUR et CITY contiennent peu de mouvement, mais des détails très fins se déplacent sur de petites distances. On observe par conséquent des variations locales de l’activité, dues aux approximations faites au niveau de l’estimation de mouvement, qui ne permettent pas de représenter les mouvements trop faibles.

Ces variations de débit posent problème dans la majorité des applications du codage vidéo. Par exemple, pour le stockage hors-ligne, il est primordial que la taille des séquences à stocker n’excède pas l’espace disponible. De la même manière, pour la transmission sur des réseaux de communication, le débit des séquences encodées ne doit pas excéder les capacités de transfert du canal assurant le transport.

La régulation de débit est donc une étape primordiale du processus d'encodage. Elle permet de moduler la quantité de données perdues de façon à respecter les contraintes formulées d'après le contexte d'encodage et d'utilisation du contenu vidéo. Pourtant, elle ne figure pas dans la spécification des standards de codage, qui s'attachent exclusivement à la description du format que les données codées doivent respecter. Certaines stratégies font toutefois partie des recommandations liées à ces standards [91, 72, 67].

La section suivante décrit les mécanismes d'optimisation débit-distorsion, qui sont au cœur de la régulation de débit pour les standards MPEG-4 AVC et SVC.

3.2 Optimisation débit-distorsion

Dans un schéma de codage avec pertes, le débit et la qualité sont deux notions fortement liées. Un fort débit signifie généralement une bonne qualité, alors qu'un débit faible engendre une qualité plus mauvaise. Le choix des paramètres de codage pour un macro-bloc (taille, mode de prédiction, vecteur de mouvement, paramètre de quantification, etc) a donc un impact à la fois sur le débit et sur la qualité obtenue.

Dans MPEG-4 AVC/H.264, le concept d'optimisation débit-distorsion (*Rate-Distorsion Optimization* : RDO) est employé pour trouver les paramètres de codage optimaux en termes de débit et de qualité [75]. Il s'agit de minimiser la distorsion introduite dans le signal reconstruit après l'encodage, tout en respectant une contrainte de débit, selon :

$$\min\{D\}, \text{ avec } R \leq R_c, \quad (3.1)$$

où D mesure la distorsion, R est le débit produit et R_c est la contrainte de débit imposée [80]. Il est bon de remarquer que la distorsion D introduite durant le processus de codage est fréquemment utilisée à la place de la qualité du flux reconstruit. Ces deux notions complémentaires sont employées indifféremment pour désigner la perte de qualité occasionnée par le codage.

Pour trouver la solution à ce problème d'optimisation, la formulation Lagrangienne a été largement utilisée pour l'équilibre qu'elle représente entre simplicité et efficacité. Son principe de fonctionnement est expliqué dans la section suivante.

3.2.1 Optimisation Lagrangienne

L'optimisation Lagrangienne permet de passer d'un problème contraint comme celui présenté par l'équation (3.1) à un problème non contraint, en introduisant la notion de critère de débit-distorsion à minimiser [13, 75]. Ce critère, prenant en paramètres la distorsion D introduite par le codage et le débit R généré, est exprimé sous la forme :

$$J(D, R) = D + \lambda.R, \quad (3.2)$$

où λ est appelé le *Lagrangien*, ou multiplicateur de Lagrange [93]. Il est utilisé pour traduire l'équilibre souhaité entre le débit et la distorsion. Si la valeur de λ est élevée, l'accent est mis sur la minimisation du débit, tandis qu'une valeur faible pour λ privilégie la minimisation de la distorsion.

Dans MPEG-4 AVC/H.264, l'optimisation Lagrangienne est utilisée à la fois pour l'estimation de mouvement et pour le choix du mode de codage des résidus.

Pour l'estimation de mouvement, le critère de débit-distorsion est exprimé de la manière suivante :

$$J_{mvt}(v, q, \lambda_{mvt}) = D_{comp}(v, q) + \lambda_{mvt}R_{mvt}(v) \quad (3.3)$$

où q est le QP utilisé, D_{comp} est l'erreur résiduelle après compensation du mouvement et quantification et R_{mvt} est le coût de codage du vecteur de mouvement v .

Le vecteur \hat{v} choisi est celui qui minimise le critère J_{mvt} :

$$\hat{v} = \arg \min_v J_{mvt}(v, \lambda_{mvt}). \quad (3.4)$$

Suite au calcul du vecteur de mouvement, le mode de prédiction des résidus est choisi en utilisant le même mécanisme. Le critère de débit-distorsion est exprimé de la manière suivante :

$$J_{mode}(m, v, q, \lambda_{mode}) = D_{rec}(m, v, q) + \lambda_{mode}R_{rec}(m, v, q), \quad (3.5)$$

où D_{rec} est la distorsion observée sur le macrobloc reconstruit et R_{rec} est le nombre de bits nécessaire pour coder le vecteur de mouvement v , le mode de prédiction m et le signal résidu après quantification avec le paramètre q .

Le mode de prédiction choisi est celui qui minimise J_{mode} :

$$\hat{m} = \arg \min_m J_{mode}(m, v, q, \lambda_{mode}). \quad (3.6)$$

Dans le cas de la prédiction intra-image, seule l'équation (3.6) est naturellement utilisée, et le coût du vecteur v est nul.

Le choix des valeurs de λ_{mvt} et λ_{mode} est un enjeu majeur de l'optimisation Lagrangienne, et plusieurs approches ont été proposées [93, 50]. L'approche adoptée dans le codeur de référence de MPEG-4 AVC/H.264 est celle de [50], qui définit les deux paramètres comme suit :

$$\begin{aligned} \lambda_{mode} &= \alpha \times 2^{q/3} \\ \text{et } \lambda_{mvt} &= \sqrt{\lambda_{mode}}, \end{aligned} \tag{3.7}$$

où α est une constante déterminée de façon empirique [79] :

$$\alpha = \begin{cases} 0.85 & \text{pour les images I et P;} \\ 4 \times 0.85 & \text{pour les images B.} \end{cases} \tag{3.8}$$

Le choix des paramètres de Lagrange représente un problème à part entière, qui ne fait pas l'objet de cette thèse. En revanche il est bon de remarquer que la valeur du QP intervient dans le choix des paramètres de prédiction (*i. e.* vecteur de mouvement et mode de prédiction). Cependant, l'étape de quantification est supposée être effectuée après la prédiction. Cette interdépendance a des répercussions importantes dans le cadre de la régulation de débit dans MPEG-4 AVC/H.264, comme le montre la section suivante.

3.2.2 Le dilemme de la poule et de l'œuf

De façon à évaluer précisément le coût en termes de débit et de distorsion d'une prédiction donnée, l'encodeur compare le macrobloc d'origine au macrobloc reconstruit après prédiction, transformée et quantification (*cf.* Figure 1.1). Par conséquent, la valeur du pas de quantification doit être connue au moment où est effectué le choix des paramètres de prédiction.

Or, dans le cas de la régulation de débit, le choix du QP est assuré après les étapes de prédiction et de transformée, de manière à adapter le nombre de bits en sortie aux ressources disponibles. Ce phénomène complique considérablement le processus de contrôle de débit, et est appelé le dilemme de la poule et de l'œuf (*chicken & egg dilemma*) [91, 72].

Pour résoudre ce problème, l'approche optimale du point de vue débit-distorsion consiste à effectuer une recherche exhaustive sur l'ensemble des paramètres de prédiction

et des valeurs de QP possibles, en gardant ceux qui minimisent le critère de débit-distorsion. Toutefois, cette méthode nécessite une quantité prohibitive de calculs et n'est pas utilisée en pratique.

Des méthodes permettent de réduire l'espace de recherche lors de la prédiction en n'utilisant que les paramètres les plus courants ou les plus probables [37, 66]. Pour l'estimation de mouvement, des algorithmes rapides tels que la recherche en diamant [17], la recherche en hexagone [8] ou l'estimation hiérarchique [14], ont été proposés afin de réduire la complexité du choix des vecteurs de mouvement.

D'autres méthodes proposent de réutiliser les paramètres du macrobloc précédent pour initialiser la prédiction [91, 36]. D'autres approches effectuent une première passe sur chaque macrobloc en utilisant une valeur de QP initiale pour choisir les paramètres de prédiction. Le QP final est ensuite calculé par un mécanisme de régulation de débit, et les macroblocs sont encodés une seconde fois avec cette valeur de QP [11, 46].

Malgré l'utilisation de méthodes rapides telles que celles décrites ici, l'optimisation débit-distorsion représente la tâche la plus coûteuse en calculs dans le processus d'encodage. La prédiction spatio-temporelle peut effectivement occuper près de 80% du temps consacré à l'encodage d'une image [38]. Cependant, elle joue un rôle déterminant dans les performances du système de codage.

La régulation de débit interagit fortement avec ce mécanisme, en choisissant la valeur du QP en fonction des contraintes imposées. Le processus de régulation de débit utilise généralement les mêmes éléments, indépendamment du standard auquel il s'applique. Dans la section suivante, nous décrivons chacun de ces éléments, en les illustrant avec des approches classiques.

3.3 Éléments d'un mécanisme de régulation de débit

Un processus de régulation de débit est généralement composé de quatre éléments. En premier lieu, les contraintes à appliquer au flux sont déterminées en fonction du contexte envisagé. Dans un deuxième temps, les ressources disponibles sont ensuite calculées et réparties entre les différents niveaux du flux. La troisième étape consiste à choisir la valeur du paramètre de quantification en accord avec les ressources à respecter. Pour finir, les données sont encodées en utilisant le QP choisi et les ressources disponibles sont mises à jour.

3.3.1 Formulation des contraintes en fonction du contexte

Selon le contexte d'encodage, de stockage ou de transmission, puis de lecture du contenu vidéo, il existe plusieurs types de contraintes, que l'on peut associer à trois types d'applications :

- **Stockage sur un support** : pour le stockage d'un flux vidéo compressé sur un support de capacité donnée (*i. e.* : CD-ROM, DVD, carte mémoire, *etc*), la régulation de débit doit garantir le fait que l'espace nécessaire pour stocker le flux vidéo compressé soit aussi proche que possible de l'espace disponible sur le support. S'il y a un dépassement de cet espace, l'intégralité du flux ne peut pas être stockée, ce qui pose évidemment problème. Il n'est pas non plus souhaitable que l'espace nécessaire pour stocker le flux soit trop inférieur à l'espace disponible. L'espace inutilisé aurait effectivement pu être comblé en autorisant une perte de données moins importante, qui aurait donné une meilleure qualité.
- **Transmission sur un réseau** : la diffusion de contenu vidéo sur des réseaux d'information est contrainte par les capacités de transfert des canaux de communication. Dans le cas d'un encodage hors-ligne pour une transmission à la demande, les contraintes sont fixées d'après une estimation des capacités de transfert des canaux. Dans le cas d'un encodage et d'une transmission en temps réel, les contraintes peuvent s'adapter aux capacités du réseau en temps réel.
- **Téléchargement pour lecture hors-ligne** : si la contrainte de temps réel n'est pas présente, on impose généralement des contraintes de qualité pour fournir à l'utilisateur un flux vidéo aussi agréable que possible. Des contraintes plus lâches de taille de fichier ou de débit maximal peuvent aussi être appliquées pour garantir une lecture fluide sur un appareil donné.

Une fois les contraintes identifiées, il est nécessaire d'en extraire les objectifs en termes de débit que devront respecter les flux compressés. Un budget à respecter est alors calculé, puis distribué parmi les différentes parties du flux vidéo. Ces méthodes de calcul sont présentées dans la section suivante.

3.3.2 Calcul de budget et stratégies de répartition

Suite à l'identification des contraintes liées à l'application envisagée, on détermine un *budget* à respecter. Par exemple, si la contrainte principale est de garantir que les flux vidéo pourront être transmis sur un canal de communication donné, le budget est déterminé à partir d'une estimation des capacités de transfert du canal en question.

Dans la majeure partie des cas, la contrainte principale est exprimée en termes

de *débit par seconde*, pour adapter le flux aux capacités de transfert des canaux de transmission, évaluées en nombre de bits transmis par seconde (*e. g.* 384 kilobits par seconde pour l'UMTS [31]). L'approche la plus courante est alors de contrôler le débit de manière à ce que le nombre de bits par seconde soit le plus proche possible de la contrainte exprimée [91, 60].

Cette capacité de transfert peut également varier au cours du temps, à cause des phénomènes de congestion ou de l'éloignement du récepteur par rapport à la source du signal (pour les canaux de transmission sans fil). La bande passante est alors un paramètre à prendre en compte [88], soit en y adaptant la contrainte de débit en temps réel [55, 76, 92], soit en la fixant au préalable à une valeur garantissant la transmission du flux dans les cas extrêmes [44, 67, 12].

La contrainte de débit par seconde offre une relative liberté dans la répartition des ressources à une échelle plus petite. Il est alors possible d'optimiser un second critère tel que la qualité du flux reconstruit, tout en s'assurant que le débit reste globalement équilibré autour de la valeur souhaitée [7].

Une répartition s'adaptant aux variations du contenu des images à coder peut représenter un grand intérêt du point de vue de la qualité ressentie. L'approche développée dans [58] consiste à diviser la séquence en segments d'activité constante. Un budget égal est ensuite alloué à chaque segment, de manière à obtenir une distorsion équivalente sur chacun d'eux. Cette méthode offre une répartition intéressante des ressources disponibles entre les différents niveaux du flux, et réduit fortement la variance du PSNR sur l'ensemble de la séquence. On observe cependant des discontinuités dans la distorsion au niveau des jonctions entre les segments, ainsi que l'apparition de valeurs isolées de faible PSNR dans les scènes agitées.

La prise en compte des outils de codage utilisés, tels que la structure de GOP hiérarchique (*cf.* Figure 1.2), rend également possible l'optimisation la qualité du flux. Un certain nombre de méthodes proposent par exemple d'allouer un nombre de bits plus important aux images complexes qu'aux images simples [84, 69, 43]. Ce type d'approche compense par conséquent les différences d'efficacité de codage entre les images et permet d'atteindre une qualité relativement constante.

3.3.3 Différentes échelles de la régulation

Une fois l'objectif principal de la régulation déterminé, le budget correspondant doit être réparti à différents niveaux du flux. Nous avons vu dans le chapitre 1 que les flux vidéo sont divisés en éléments de taille plus ou moins fine. On distingue en effet les groupes d'images (GOP), les images et les macroblocs. En fonction des besoins et des contraintes, le niveau de la régulation peut changer.

La répartition de budget au niveau des GOP est souvent utilisée pour traduire la contrainte de plus haut niveau (*e. g.* débit par seconde) en une contrainte ayant un sens au niveau du flux [91, 89]. En effet, les GOP représentent le plus haut niveau de structuration des flux, et la "seconde" de vidéo n'a pas de signification directe. Généralement, le budget alloué à un GOP est calculé en fonction de la valeur de la contrainte de débit C , du nombre d'images par seconde F et de la taille S du GOP :

$$G = S \times \frac{C}{F}. \quad (3.9)$$

La répartition de budget au niveau des GOP est une étape intermédiaire, qui nécessite d'être affinée par une répartition au niveau des images, et éventuellement des macroblocs.

Il existe des mécanismes se contentant de contrôler le débit au niveau des images [41, 55, 56]. L'allocation de ressources est exécutée une fois par image et le même QP est utilisé pour encoder toute l'image. Ces approches offrent une régulation relativement grossière, car les erreurs de contrôle ne peuvent être corrigées qu'une fois par image, ce qui ne permet pas toujours de compenser des changements brusques d'activité dans les données à coder. De plus, le QP de l'ensemble des macroblocs est modifié simultanément. Le débit nécessaire pour coder l'image varie donc par sauts lorsque l'on change la valeur du QP. Néanmoins, ces approches ont une complexité faible. Elles sont donc adaptées à des applications où la contrainte de complexité algorithmique est forte, comme l'encodage en temps réel.

Certaines approches raffinent le contrôle au niveau des macroblocs [50, 47]. Elles sont axées sur les performances du contrôle en termes de précision et d'adaptation aux changements d'activité dans les données à coder. Les erreurs peuvent en effet être corrigées rapidement sur de petites portions du flux, ce qui donne une grande souplesse à la régulation effectuée. En revanche, leur complexité algorithmique est beaucoup plus élevée, car le mécanisme de contrôle est réitéré de nombreuses fois pour chaque image.

3.3.4 Respect du budget

Une fois un budget attribué à chaque niveau du flux, la seconde étape de la régulation de débit consiste à contrôler les paramètres d'encodage pour respecter ce budget. L'enjeu principal de cette étape est de choisir la valeur du paramètre de quantification (QP) qui génère un nombre de bits aussi proche que possible du budget à respecter.

3.3.4.1 Recherche exhaustive

La stratégie la plus simple consiste à encoder chaque image ou chaque macrobloc avec chaque valeur de QP possible. On retient ensuite la valeur pour laquelle la différence entre le débit obtenu et l'objectif fixé est la plus faible. Bien entendu, cette approche est extrêmement coûteuse en calculs. Il est possible d'alléger cette approche, en adoptant un algorithme de recherche par dichotomie ou par un algorithme de recherche logarithmique, comme cité dans [87].

Ce type d'approche exhaustive est essentiellement utilisé pour tester les performances d'un standard de codage, ou pour comparer l'efficacité d'un mécanisme de contrôle de débit plus simple avec le fonctionnement optimal d'un encodeur. En pratique, les approches applicables de la régulation se limite à une ou deux passes d'encodage. Elles se basent sur un modèle de la relation entre le débit et le QP, comme le décrit la section suivante.

3.3.4.2 Relation entre le débit et le QP

La plupart des approches réalistes de la régulation de débit tentent d'exprimer la relation entre la valeur du paramètre de quantification et le débit en sortie de l'encodeur. Le QP détermine en effet la quantité d'informations perdues lors de l'étape de quantification et a une influence directe sur le débit en sortie de l'encodeur (Figure 3.2(a)). Cette relation dépend du contenu des données à coder, déterminé par l'activité présente dans la séquence d'origine et le mode de prédiction utilisé.

Les approches basées sur des modèles sont naturellement moins efficaces, du point de vue de la précision, que les approches exhaustives décrites plus tôt. Elles se basent en effet sur une approximation de la relation entre le QP et le débit et se heurtent aux imprécisions de cette approximation. Cependant, elles permettent de contrôler le débit en limitant le nombre d'itérations nécessaires et leur efficacité moindre est amplement compensée par leur plus grande rapidité.

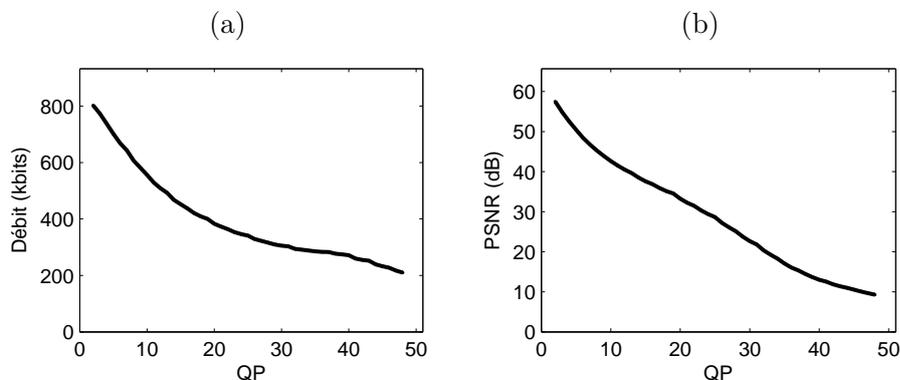


FIGURE 3.2 – (a) relation entre le débit et le paramètre de quantification. (b) relation entre la qualité et le paramètre de quantification (tracés obtenus à partir d'une image P taille $352 * 288$ codée avec MPEG-4 AVC/H.264).

3.3.5 Paramètres clés de la relation entre débit et QP

Différents modèles de la relation entre le QP et le débit ont été étudiés dans la littérature. Une synthèse des principales contributions est proposée dans [10]. Certaines approches se basent sur des modèles analytiques en tentant de prédire le comportement du débit d'après les données elles-mêmes [18, 40]. D'autres utilisent des modèles plus simples basés sur un apprentissage de cette relation [33, 67, 50]. Dans toutes ces approches, on retrouve un certain nombre de paramètres qui influent sur la relation débit-QP. Il sont passés en revue dans cette section.

3.3.5.1 Entropie des résidus de prédiction

L'entropie d'un élément, qui caractérise son activité, est un bon indicateur de son coût de codage. Elle mesure la complexité du contenu, et par conséquent, plus elle est élevée, plus le nombre de bits nécessaires pour coder l'élément est grand [50, 54].

Dans le contexte du codage prédictif, le coût de codage est essentiellement lié à l'entropie du signal de résidus de prédiction. À valeur de QP égale, le codage d'un macrobloc ayant une entropie forte demandera en effet plus de bits que le codage d'un macrobloc ayant une entropie faible. Pour évaluer le coût de codage, il est donc intéressant de mesurer l'entropie du bloc de résidus.

Il existe plusieurs manières d'évaluer le coût des résidus de prédiction, suivant l'échelle à laquelle on se place. Au niveau d'une séquence ou d'un groupe d'images, on le calcule généralement avec la somme des différences entre les pixels des images consécutives [58]. Au niveau d'une image, on la calcule à l'aide d'un critère statistique

équivalent à la moyenne des distances absolues (MAD) entre deux images consécutives ou entre l'image courante et une valeur moyenne [55, 69]. La variance est également régulièrement utilisée pour symboliser l'activité, tant au niveau des macroblocs qu'au niveau des images.

3.3.5.2 Macroblocs non codés

Après la prédiction spatio-temporelle, les macroblocs sont transformés puis quantifiés. Lorsque la prédiction est très efficace, il arrive que le bloc de résidus ne contienne que des valeurs très faibles. On peut alors décider de ne représenter le macrobloc que par le vecteur de mouvement ou le mode de prédiction trouvés.

Dans [12], une approche consistant à prédire si un bloc sera codé ou ignoré est introduite. Une fois les macroblocs non codés identifiés, le modèle de débit peut se concentrer sur les blocs restants, ce qui offre une régulation plus juste, en évitant notamment d'allouer une partie des ressources aux blocs qui seront ignorés.

3.3.5.3 Données d'en-tête dans un macrobloc

Un macrobloc encodé contient toutes les informations nécessaires à sa reconstruction au niveau du décodeur. Ces informations sont séparées en deux catégories : les coefficients quantifiés, appelés données de *texture*, et les informations de codage (vecteur de mouvement, mode de prédiction, paramètre de quantification), appelés données d'*en-tête*.

Les données d'en-tête sont codées sans perte, et la quantité de bits nécessaire pour les coder n'obéit pas aux mêmes règles que les données de texture. Des modèles de débit dédiés à ces données d'en-tête ont été proposés pour évaluer plus précisément la quantité de bits nécessaire pour coder un macrobloc [44, 57]. Ils permettent d'obtenir une modélisation du débit plus juste, notamment à bas débit, lorsque la quantité de bits d'en-tête peut représenter près de 20% du débit utilisé pour coder un macrobloc.

3.3.6 Régulation de débit et qualité

La qualité du flux encodé peut être affectée de manière significative par la régulation de débit. Le débit et la qualité sont en effet deux notions fortement liées, un débit faible allant généralement de pair avec une mauvaise qualité, tandis qu'un débit élevé permet de conserver une qualité plus satisfaisante (Figure 3.3). Par suite, imposer des contraintes sur le débit a des conséquences sur la qualité produite.

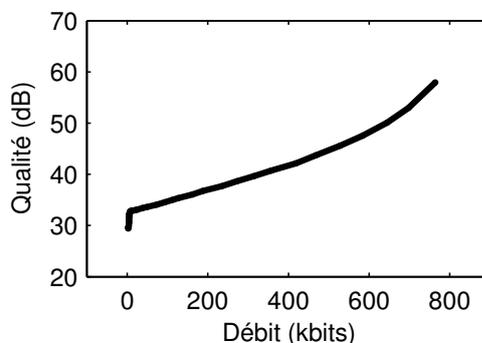


FIGURE 3.3 – Relation entre le débit et la qualité pour une image P de taille $352 * 288$ encodée avec MPEG-4 AVC/H.264.

Pour évaluer la qualité d'un encodage, il existe plusieurs types de mesures. Certaines sont des mesures purement statistiques et sont dites *objectives*, d'autres sont basées sur l'impression visuelle ressentie par un utilisateur et sont dites *perceptuelles*.

3.3.6.1 Mesures objectives de qualité

Les mesures objectives de qualité ont l'avantage de reposer sur des critères statistiques relativement simples et peu coûteux à calculer. Une des mesures les plus utilisées est le rapport signal sur bruit (*Peak Signal-to-Noise Ratio* : PSNR) [42, 54]. Étant donnée une image f_0 de dimensions $m \times n$ et l'image reconstruite f_r après encodage, le PSNR est calculé de la manière suivante :

$$\text{PSNR}(f_0, f_r) = 10 \cdot \log_{10} \frac{(2^b - 1)^2}{\frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \|f_0(i, j) - f_r(i, j)\|^2}, \quad (3.10)$$

où b est le nombre de bits sur lequel sont codées les valeurs des pixels de l'image (typiquement, $b = 8$).

D'autres critères statistiques équivalents sont aussi couramment utilisés, comme la moyenne des différences absolues (*Mean Absolute Difference* : MAD), la somme des distances absolues (*Sum of Absolute Distances* : SAD), ou la somme des erreurs quadratiques (*Sum of Squared Errors* : SSE) [33, 67, 55].

Le PSNR est une mesure de qualité objective, qui n'est certes pas parfaitement corrélée avec l'impression visuelle de qualité, mais dont le calcul ne fait pas intervenir de modèle complexe. De plus, c'est une mesure très largement utilisée pour évaluer les performances d'un standard de codage.

L'inconvénient majeur de ces mesures objectives de qualité est qu'elles ne sont pas bien adaptées au Système Visuel Humain, et ne reflètent pas toujours l'impression visuelle ressentie par un utilisateur standard. Pour remédier à ce problème, il existe des mesures perceptuelles, qui tentent de s'approcher de l'impression visuelle de qualité.

3.3.6.2 Mesures perceptuelles de qualité

Les mesures perceptuelles cherchent d'évaluer de façon plus fidèle la qualité au sens de l'impression visuelle. Elles tentent de prédire cette impression en se basant sur des caractéristiques du Système Visuel Humain. Des exemples de mesures de qualité impliquant des critères subjectifs sont traités dans [77, 61, 16].

Ces modèles offrent la possibilité d'optimiser la qualité en accord avec la perception d'un utilisateur. Cependant, elles mettent en jeu des outils complexes et leur mise en place dans un codeur vidéo n'est pas simple.

3.3.6.3 Minimisation des variations de la qualité

Les variations de qualité au cours d'une séquence ont une influence sur l'impression visuelle ressentie par un utilisateur [58, 41]. Il est donc intéressant de minimiser ces variations, en cherchant à obtenir une qualité constante. On peut également relâcher cette contrainte en cherchant des variations continues de la qualité [30, 48]. Cette deuxième approche permet en outre de conserver les variations de qualité à l'échelle de la séquence, qui font partie intégrante des données à coder.

3.3.7 Gestion du buffer de transmission

Dans les applications qui requièrent le transport du flux encodé, les images sont généralement placées dans une mémoire tampon à la suite du processus d'encodage. Les données accumulées sont ensuite envoyées sur le canal avec un débit constant, dépendant de la bande passante du canal. Le contrôle du taux d'occupation du buffer est un aspect important de la régulation de débit. En effet, s'il arrive à saturation, les données en sortie de l'encodeur ne peuvent pas être stockées et l'encodage doit s'arrêter jusqu'à ce que le transfert des données sur le canal libère un espace suffisant [67]. À l'inverse, si le buffer se vide, il se peut que le débit du flux sur le canal soit interrompu, ce qui représente un gaspillage de bande passante. De plus, dans le cas d'un décodage instantané, le manque de données à décoder donne lieu à des interruptions qui ne sont pas souhaitables.

Avant que les données d'une image puissent être envoyées vers le canal de communication, les données des images précédentes encore présentes dans le buffer doivent

d'abord être transmises. Si le nombre de bits utilisés pour coder les images est plus élevé que la capacité de transfert du canal, des phénomènes d'accumulation peuvent apparaître. Afin d'éviter que cette accumulation provoque la saturation du buffer, la stratégie adoptée dans [67] détermine le nombre de bits alloués à une image en fonction de l'occupation du buffer. Ce mécanisme permet de compenser l'accumulation en allouant moins de bits à une image lorsque le buffer est près de la saturation. De plus, il permet d'éviter que le buffer ne se vide, en allouant plus de bits à une image lorsque l'occupation est faible.

Du côté du décodeur, un buffer identique reçoit les données envoyées par l'encodeur, dans l'attente de leur décodage. Le fonctionnement du buffer de décodage est sensiblement différent de celui du buffer d'encodage. En effet, les données y arrivent à débit constant depuis le canal de communication, et quittent le buffer par à-coups.

Une partie de la norme MPEG-4 AVC/H.264 appelée le HRD (*Hypothetical Reference Decoder*, [78, 49]), modélise le comportement du buffer de décodage et fournit des directives à respecter pour qu'il ne sature pas². Plusieurs stratégies de régulation de débit ont été élaborées en rapport avec le HRD. Dans [51, 73], le budget alloué à une image est contraint par une borne inférieure et une borne supérieure. Le calcul de ces bornes repose sur un modèle qui prend en compte les caractéristiques du buffer (débit d'arrivée des données et taille du buffer) pour calculer son taux d'occupation à l'instant d'encodage de l'image. Le processus de régulation de débit permet ainsi d'éviter la saturation du buffer et de garantir un décodage sans interruption.

3.4 Contextes applicatifs et contraintes supplémentaires

Certains contextes applicatifs imposent des contraintes supplémentaires à la régulation de débit. Dans les scénarios à bas débit ou les applications temps réel, les ressources sont par exemple très limitées, ce qui implique une gestion particulière. Nous décrivons dans cette section ces contextes et les contraintes qui leur sont associées.

3.4.1 Optimisation complexité-performances

Globalement, les performances de l'encodeur et la complexité de calcul sont deux notions entre lesquelles il faut trouver un équilibre. L'optimisation complexité/performances fournit un cadre théorique à cette problématique. Pour permettre

2. Par abus de langage, nous utilisons le terme *saturation* pour exprimer à la fois le cas où le buffer est plein, et le cas où le buffer est vide.

d'ajuster le processus d'encodage aux capacités de calculs disponibles, l'optimisation complexité-performances consiste à voir la qualité de codage comme un critère à maximiser, tout en respectant des contraintes sur la quantité de calculs effectués.

Dans [66], l'étape de prédiction spatio-temporelle est répétée jusqu'à atteindre un certain seuil de qualité. Une prédiction sous-optimale est donc généralement trouvée, mais la quantité de calculs a de grandes chances d'être réduite. Dans [32], la quantité de calculs est de plus limitée, de manière à arrêter la recherche d'une prédiction si la quantité de calculs devient trop élevée.

L'adaptation de la complexité de l'encodeur trouve notamment des applications dans la transmission vidéo à partir d'appareils fonctionnant sur batterie (*e. g.* PDA, téléphones mobiles, *etc*), pour lesquels la consommation énergétique est liée à la charge de calculs demandée au processeur. Dans [19] un lien est établi entre le nombre de macroblocs non codés dans une image, la complexité de l'encodeur et la consommation énergétique. Le nombre de macroblocs qu'il est acceptable de coder d'après les ressources disponibles est alors calculé, pour ne traiter que ceux qui ont les SAD les plus élevées. Le reste des macroblocs n'est tout simplement pas codé, pour économiser les ressources. Cette méthode permet par conséquent d'économiser les ressources en énergie tout en minimisant l'impact sur l'image reconstruite.

3.4.2 Scénarios à bas débit et temps réel

Dans les applications à bas débit, la quantité des bits d'en-tête peut représenter la plus grande partie des données codées [88]. Il est donc important de construire des modèles de débit dédiés aux bits d'en-tête. Dans [84], une approche simple se base sur le nombre de bits d'en-tête du macrobloc précédent pour prédire celui du macrobloc courant.

Cependant, la quantité de bits d'en-tête peut varier de façon importante au cours de la séquence. L'approche présentée dans [12] modélise donc le nombre de bits d'en-tête séparément pour la prédiction intra et la prédiction inter image. Pour les blocs prédits en intra, la taille de l'en-tête est relativement constante, ce qui permet de l'estimer à l'aide de la moyenne du nombre de bits d'en-têtes des macroblocs précédents. Pour les blocs prédits en inter, on effectue le choix des vecteurs de mouvement, puis la taille de l'en-tête est estimée en comptant le nombre de composantes non nulles dans les vecteurs choisis.

Le codage temps réel est un des contextes les plus restrictifs. En effet, la rapidité de l'encodage et du décodage sont primordiales, ce qui limite fortement leur complexité. De plus, la taille du buffer de transmission est fortement réduite, de manière à limiter l'accumulation de données à transmettre et le délai entre l'encodage et le décodage [47]. Dans [57], un calcul plus précis de l'occupation du buffer avant l'encodage de chaque image permet de limiter ce délai, en arrêtant d'encoder les images tant que l'occupation du buffer est trop élevée.

En outre, la fréquence temporelle est généralement faible pour limiter la quantité de données à coder. Il en résulte une corrélation relativement faible entre deux images consécutives, qui peut rendre les images P, et à plus forte raison les images B, moins efficaces que les images I [88]. De plus, elles ajoutent une complexité au processus d'encodage qui achève de les rendre inadaptées. Il peut donc parfois être préférable de ne pas utiliser la prédiction temporelle dans le cadre du codage temps réel.

3.5 Le ρ -domaine, une approche intéressante

Les approches classiques tentent d'exprimer les relations entre le débit, la distorsion et le QP. La Figure 3.2 montre un exemple des relations liant le débit la qualité et le QP. Il est difficile d'obtenir des approximations précises de ces relations. Dans [20], Z. He *et al* introduisent ρ comme étant le pourcentage de coefficients nuls dans un macrobloc après quantification. Après avoir introduit une relation simple permettant d'obtenir la valeur de ρ à partir de la valeur du pas de quantification, ils montrent que ce paramètre a une influence directe sur le débit nécessaire pour coder un macrobloc. Dans les schémas de codage classiques tels que MPEG-2 ou MPEG-4 AVC/H.264, la relation qui lie ce paramètre ρ avec le débit R est linéaire pour des variations limitées de ρ [20]. Il est donc possible d'estimer la valeur du débit en fonction de ρ de façon très simple. Partant de ces deux observations élémentaires, on utilise alors ρ comme un intermédiaire pour exprimer la relation plus complexe qu'il existe entre le débit et le pas de quantification. Cette approche a pour particularité d'être très efficace pour modéliser la relation qui existe entre le débit à la sortie de l'encodeur et la valeur du pas de quantification, tout en étant basée sur un principe très simple.

3.5.1 Relation entre ρ et le QP

L'approche du ρ -domaine introduit ρ comme un paramètre intermédiaire pour caractériser la relation entre le débit et le QP. On cherche dans un premier temps à exprimer la relation entre ρ et le QP. Pour cela, il suffit d'analyser le processus de quantification

utilisé pour coder les coefficients d'un macrobloc après l'étape de transformée. Nous avons décrit le mécanisme de quantification associé dans le chapitre 1.4. D'après l'équation (1.7), la valeur de ρ pour un macrobloc m en fonction de la valeur du QP utilisée, notée q , est donnée par

$$\rho(q) = \frac{1}{M} \sum_{(i,j) \in m} dz(Y_{ij}^m, i, j, q), \quad (3.11)$$

où M est le nombre de coefficients dans le macrobloc m .

Il est donc aisé de calculer ρ en fonction de la valeur du QP. Le paragraphe suivant montre comment la relation entre ρ et le débit nécessaire pour coder le macrobloc peut être estimée par un modèle linéaire.

3.5.2 Modèle de débit

Intuitivement, il est aisé de constater que le débit décroît strictement quand le nombre de coefficients nuls dans un macrobloc augmente. La Figure 3.4 montre la relation ρ et le débit obtenue expérimentalement en encodant un macrobloc avec toutes les valeurs de pas de quantification comprises entre 0 et 51. Cette relation offre un aspect très linéaire, et peut être approchée par la formule :

$$R(\rho) = \theta(1 - \rho), \quad (3.12)$$

où θ est la pente de la droite à calculer. Il est directement visible que cette droite passe par le point $(1; 0)$. En effet, quand ρ vaut 1, tous les coefficients du macrobloc sont nuls, et le débit R est égal à 0. Pour caractériser cette relation, il est donc suffisant de déterminer θ , la pente de la droite. Dans [20], la valeur de θ est calculée après l'encodage de chaque image et utilisée pour prédire celle de l'image suivante.

Dans [71], une autre formulation de la relation entre ρ et le débit est utilisée :

$$\rho(R) = \frac{R_0 - R \times (1 - \rho_0)}{R_0}, \quad (3.13)$$

où R_0 et ρ_0 sont des paramètres du modèle à déterminer. Par une simple reformulation, on observe que cette équation est équivalente à

$$R(\rho) = \frac{R_0}{1 - \rho_0}(1 - \rho). \quad (3.14)$$

Cette relation est donc bien linéaire et passe également par le point $(1; 0)$. Les paramètres R_0 et ρ_0 jouent le même rôle que la pente de la droite θ dans l'équation (3.12). Ils correspondent aux coordonnées d'un deuxième point sur la droite, qui peut être obtenu à partir d'une phase d'initialisation ou d'une image précédente.

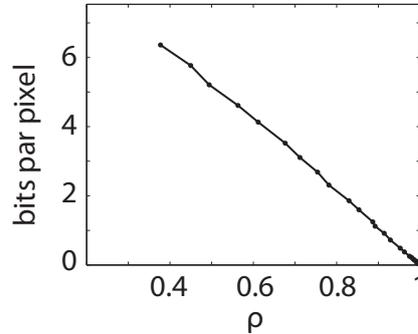


FIGURE 3.4 – Relation entre ρ et le débit sur une image P encodée avec MPEG-4 AVC/H.264.

Les deux équations (3.12) et (3.13) sont simplement deux formulations équivalentes de la même relation. Dans la suite de nos travaux, nous nous baserons sur l'équation (3.13), car elle est tirée d'un article validant l'usage du ρ -domaine pour MPEG-4 AVC/H.264 [71]. Il est important de noter que le raisonnement que nous avons suivi est parfaitement reproductible en utilisant la formule (3.12) pour caractériser la relation entre ρ et le débit.

Finalement, en combinant les équations (3.11) et (3.14), on obtient une relation entre le débit R et le QP q :

$$R(q) = R(\rho(q)). \quad (3.15)$$

3.5.3 Calcul du QP optimal

L'équation (3.15) permet de trouver le nombre de bits nécessaire pour coder un macrobloc en fonction de la valeur du pas de quantification utilisée. Elle est utilisée pour le calcul du QP optimal à partir du nombre de bits souhaité pour le macrobloc. Soit R_T le budget alloué au macrobloc. En utilisant l'équation (3.15), la valeur optimale du QP, notée q_T , est celle qui vérifie :

$$q_T = \arg \min_{q \in \{0;51\}} |R_T - R(q)|. \quad (3.16)$$

Le pas de quantification étant une valeur entière, la relation le liant le débit n'est pas continue. C'est pourquoi on cherche la valeur du QP qui minimise la différence entre le nombre de bits atteint et le nombre de bits voulu.

3.5.4 Au niveau image

Pour économiser une quantité non négligeable de calculs, il est parfois préférable d'effectuer le contrôle de débit au niveau des images et d'encoder chaque macrobloc avec la même valeur de QP. Au niveau d'une image f , la valeur de ρ est obtenue en sommant les valeurs de ρ sur l'ensemble des macroblocs m :

$$\rho(q) = \sum_{m \in f} \left(\frac{1}{M} \sum_{(i,j) \in m} dz(c_{ij}^m, i, j, q) \right). \quad (3.17)$$

Les équations (3.14) restent valides, dans la mesure où les valeurs de R_0 et ρ_0 sont calculées elles aussi au niveau de l'image. Dans la suite, nous conserverons les mêmes notations pour le contrôle de débit au niveau image.

3.5.5 Conclusion

Le ρ -domaine est une approche intéressante pour la régulation de débit tant du point de vue de la précision de sa représentation que de sa faible complexité en termes algorithmiques.

La plupart des approches de la régulation de débit tentent de modéliser directement la relation entre le débit et le QP. Cette relation est complexe et les approximations proposées utilisent classiquement des modèles Laplaciens ou Gaussiens en fonction de l'activité dans les données à coder. Ces modèles souffrent d'imprécisions qui se repercutent sur l'allocation de ressources et le choix du QP. En revanche, la représentation de la relation entre le débit et ρ par un modèle linéaire est très précise. De plus, la relation entre ρ et le QP résulte d'une analyse du mécanisme de quantification et est par conséquent exacte. Le ρ -domaine est donc une approche qui permet de modéliser la relation entre le débit et le QP de manière très précise.

En outre, des approches basées sur une représentation directe de la relation entre le débit et le QP sont généralement plus coûteuses et moins précises. Elles recourent à des approximations quadratiques et doivent estimer la valeur de plusieurs paramètres à l'aide de mécanismes coûteux tels que la régression linéaire. Au contraire, le ρ -domaine utilise un modèle linéaire de débit dont les paramètres peuvent être obtenus à partir des données elles mêmes, sans utiliser de mécanisme d'estimation. Sa complexité algorithmique est donc très faible.

3.6 Régulation de débit pour MPEG-4 SVC

Les contributions dans le domaine du contrôle de débit pour MPEG-4 SVC sont relativement peu nombreuses. Dans [6], les auteurs proposent d'adapter le mécanisme de référence de MPEG-4 AVC/H.264 [91] au logiciel de référence de MPEG-4 SVC. Leur contribution n'affecte cependant que la couche de base des flux SVC et son intérêt est avant tout de permettre une implémentation plus aisée de futures stratégies plus élaborées.

Dans [86], un mécanisme de contrôle de débit est construit pour MPEG-4 SVC à partir de deux méthodes existantes pour MPEG-4 AVC/H.264 [72, 94]. Le mécanisme reprend les considérations liées au HRD et au contrôle du buffer vidéo introduites par ces deux méthodes (voir section 3.3.7). En revanche le contrôle n'est pas appliqué sur les couches d'amélioration en scalabilité temporelle, ce qui restreint le champ d'action de l'algorithme. De plus, l'état d'avancement du processus de normalisation au moment de la publication de cet article ne permettait pas aux auteurs d'évaluer leur méthode sur une version stable de l'encodeur de référence. Les résultats obtenus sont donc difficilement extrapolables.

Dans [48], une analyse de la répartition des modes de prédiction utilisés par les différents types d'images met en évidence des comportements différents en fonction de la taille des GOP hiérarchiques. Lorsqu'elle augmente, le pourcentage de blocs Intra utilisés dans les images P augmente à son tour pour compenser l'écart entre les images de référence. Pour des GOP de taille 8, les images P utilisent en moyenne 60% de blocs Intra, et jusqu'à 80% pour des GOP de 16 images. De même, les performances des images B changent en fonction de leur niveau temporel. Les images de niveau temporel bas sont prédites par rapport à des images relativement éloignées dans le temps et ont par conséquent une efficacité de codage relativement proche des images P. Les images B de niveau temporel élevé sont quant à elles prédites par rapport à des images de plus en plus proches et leur efficacité de codage est plus élevée.

Cette étude met en évidence la nécessité de construire des méthodes de contrôle de débit dédiées à la structure de GOP hiérarchique utilisée dans MPEG-4 SVC. Les différences d'efficacité de codage entre les types d'images doivent être compensées par une répartition particulière des ressources disponibles pour le groupe d'images. La qualité du flux reconstruit en est naturellement fortement affectée.

Un autre mécanisme intéressant est présenté dans [87], avec des considérations davantage liées à la vidéo scalable. Pour prédire les caractéristiques d'une image dans une couche d'amélioration, les informations sont extraites à la fois de l'image correspondante dans la couche de base et de l'image précédente dans la même couche. Une mesure d'activité dans les images de résidus de prédiction permet de décider quelle image utiliser comme référence pour le contrôle de débit. Si l'activité est élevée, les corrélations sont plus faibles avec l'image précédente qu'avec l'image de la couche de base. Les informations de la couche de base sont par conséquent privilégiées.

Ce mécanisme permet donc de s'adapter aux changements brusques d'activité dans les couches d'amélioration en utilisant la couche de base. Le fait d'utiliser des images déjà encodées comme base pour le modèle de débit permet à cette méthode de n'encoder chaque image qu'une seule fois. Néanmoins elle opère à la fois au niveau image et au niveau macrobloc, ce qui augmente sa complexité algorithmique. En outre, les performances au niveau des images ne sont pas présentées, et seul le débit moyen obtenu par seconde est comparé avec la contrainte formulée. La qualité n'est de plus évaluée que par la valeur moyenne du PSNR calculée sur la séquence, sans montrer les variations au niveau des images .

Cette contribution représente toutefois à l'heure actuelle la version plus aboutie de régulation de débit pour MPEG-4 SVC. L'exploitation des dépendances inter-couches en fait un mécanisme réellement adapté à la vidéo scalable et il figure parmi les seuls qui ne requièrent pas de phase de pré-encodage.

Dans [45], une méthode d'optimisation de l'encodage SVC sur l'ensemble des couches est proposée. Un facteur d'importance est affecté à chaque couche en fonction du nombre d'utilisateurs à qui elle s'adresse et des contraintes qui lui sont imposées en termes de qualité. De plus, l'impact de la régulation du débit sur la couche de base dans les couches d'amélioration est considéré comme une contrainte supplémentaire.

Le choix d'une couche de base différente pour chaque image est également abordé dans le cas de la scalabilité combinée. Si deux couches représentent des candidats intéressants pour une couche d'amélioration, le choix de la couche de base peut poser problème. Par exemple, pour une couche d'amélioration combinée en qualité et en spatial, une couche de base ayant une résolution inférieure mais une qualité équivalente permet d'exploiter les redondances spatiales de manière efficace. En revanche, une couche de base ayant une résolution identique mais une qualité inférieure permet d'exploiter les redondances en termes de qualité.

Malgré le manque de résultats expérimentaux complets, la répartition globale du

débit dans le flux reste une idée intéressante, optimisant les performances du codage pour l'ensemble du flux, et non pas pour chaque couche indépendamment.

Dans [5], une stratégie de régulation de débit pour SVC basée exclusivement sur le contrôle du buffer de transmission est présentée. Il suffit en effet à garantir que le débit en sortie de l'encodeur est compris entre des bornes inférieures et supérieures. Du point de vue du décodeur, le buffer se remplit avec un débit constant au fur et à mesure que les données transmises arrivent, et se vide par à-coups lorsqu'une image est retirée car elle n'est plus utilisée comme référence.

L'approche proposée consiste donc à contrôler que l'occupation du buffer reste comprise entre un seuil minimal et un seuil maximal. Trois niveaux de contrôle sont implémentés. Le premier niveau contrôle l'occupation du buffer à chaque image I (*i. e.* : à chaque *Intra Period*, cf. chapitre 2), pour éviter les dérives lentes de l'occupation en maintenant un niveau constant avant le décodage de chaque image I. Le deuxième niveau contrôle le buffer pour chaque image, et agit sur le QP si le niveau d'occupation s'approche des seuils à ne pas dépasser. Enfin, le troisième niveau contrôle le buffer au niveau de chaque ligne de macrobloc, en vérifiant que les ressources disponibles pour l'image sont dépensées de manière progressive. Un buffer pour chaque couche est géré de manière indépendante, et sa taille dépend du débit souhaité pour la couche.

Le contrôle de buffer est un aspect intéressant dans la transmission vidéo, car il permet de garantir que le flux peut être envoyé sur un canal de communication à débit fixé et décodé par un appareil ayant des capacités de stockage limité. Cette méthode de régulation de débit ne nécessite aucune phase d'initialisation coûteuse et permet de respecter les contraintes de débit à la sortie de l'encodeur de manière précise. Toutefois, le calcul du QP proposé dans l'article est quelque peu simpliste et consiste simplement à en diminuer la valeur si le buffer se vide, et à l'augmenter si le buffer se remplit trop vite. De plus, le cas où le flux de données arrivant dans le buffer est rompu n'est pas géré, ce qui peut mener à des problèmes d'épuisement du buffer.

Pour finir, on note quelques contributions concernant la régulation de débit et son influence sur la transmission, la robustesse et l'adaptabilité des flux scalables. Dans [90], une stratégie d'ordonnancement de la transmission des flux scalables est présentée. Une technique utilisant des buffers multiples est notamment introduite pour rendre la transmission plus robuste aux pertes aléatoires de paquets de données. Une priorité est affectée à chaque buffer, en fonction de l'impact des images qu'il contient sur la qualité de la vidéo. Ainsi, lorsque les capacités de transfert du réseau diminuent, les images des

buffers prioritaires sont d'abord transmises, suivies par les images de ayant une priorité moins élevée. Les pertes de qualité sont donc réduites et le système s'adapte mieux aux phénomènes de congestion du réseau ou de baisse des capacités de transfert.

L'impact de la répartition du débit entre les couches sur la qualité du flux reconstruit est également étudié dans [4]. Il en ressort que le débit de la couche de base doit se situer entre 25% et 30% du débit de la couche supérieure pour constituer une base efficace pour la prédiction inter-couches. Dans [2, 52], des stratégies de distribution des unités NAL dans les différents niveaux de raffinement en qualité sont présentées en fonction de leur influence importante sur la qualité du flux reconstruit. Ainsi, l'envoi des unités les plus importantes est privilégié pour transmettre en priorité les niveaux de base et assurer une qualité de service minimale dans tous les cas.

Au regard de l'étude bibliographique que nous avons présentée dans ce chapitre, il est possible de dégager les points clés de la régulation de débit dans un contexte scalable. La section suivante synthétise les notions vues dans ce chapitre et présente les enjeux liés à notre problématique.

3.7 Synthèse de l'étude bibliographique

Notre étude bibliographique a mis en évidence les points clés de la régulation de débit pour le codage vidéo. Nous allons maintenant présenter les axes de recherche développés pour répondre à la problématique de cette thèse, à savoir la régulation de débit associée au codage vidéo scalable MPEG-4 SVC.

3.7.1 Choix des contraintes appliquées

Comme nous l'avons vu, le processus de régulation commence par le choix des contraintes à imposer aux flux. Dans la plupart des applications envisagées pour MPEG-4 SVC, chaque couche est destinée à une cible particulière, caractérisée par ses capacités de décodage et la vitesse de transfert disponible pour transmettre la vidéo.

Nous imposerons donc une **contrainte de débit constant par seconde à chaque couche**, qui pourra être déterminée directement à partir des capacités du canal de communication de la cible. Chaque couche sera dans un premier temps gérée indépendamment des autres couches, avant de considérer les dépendances entre couches.

Cette contrainte de débit constant par seconde est en outre assez générale pour pouvoir évoluer, notamment vers des scénarios dynamiques prenant en compte les variations de bande passante.

3.7.2 Régulation de débit niveau image

Le niveau de résolution de la régulation constitue un facteur clé dans la complexité et la souplesse d'un mécanisme. Une régulation au niveau des macroblocs permet de compenser rapidement les erreurs de débit et d'appliquer un traitement spécifique aux différentes zones de l'image, notamment pour prendre en compte les aspects liés à la qualité visuelle. En revanche elle pénalise fortement le mécanisme en lui donnant une complexité en calculs élevée.

Une régulation au niveau des images permet au contraire de construire un mécanisme peu complexe, au prix d'un contrôle relativement moins précis. L'approche retenue effectue donc une **régulation au niveau des images**, de manière à limiter la complexité algorithmique du mécanisme.

3.7.3 Validation d'un modèle de débit

Nous avons vu que le choix du modèle de débit est une des pièces maîtresses de l'efficacité d'un mécanisme de régulation. C'est ce modèle qui permet de relier le paramètre de quantification au nombre de bits générés par l'encodage d'une image ou d'un macrobloc, et par conséquent d'adapter le choix du QP aux ressources et aux besoins. Plusieurs types de modèles ont été décrits dans ce chapitre, certains étant basés sur des approximations de la relation QP-débit par des fonctions mathématiques plus ou moins complexes, d'autres s'appuyant sur des modèles empiriques.

Le modèle de débit basé sur le ρ -domaine a particulièrement retenu notre attention de par sa grande précision, sa simplicité et sa faible complexité en termes de calculs. En effet la relation qui relie ρ au débit est fortement linéaire, ce qui lui confère à la fois une bonne précision et un faible coût de calculs. Au regard de ces deux qualités essentielles, nous proposons d'utiliser le ρ -domaine comme **modèle de débit pour notre stratégie de régulation** sur la vidéo scalable.

Un des premiers enjeux qui découlent de ce choix consiste en la validation de ce modèle dans le contexte de MPEG-4 SVC. Nous étudierons la validité du modèle sur des configurations représentatives, de manière à montrer qu'il est utilisable dans un contexte scalable.

Cette validation mènera ensuite à l'incorporation du modèle de débit dans un mécanisme de régulation en deux passes pour MPEG-4 SVC, démontrant l'efficacité du ρ -domaine.

3.7.4 Régulation de débit et qualité constante

Il est primordial de prendre en compte la qualité dans le processus de régulation de débit. Notamment, les variations de qualité sont perçues de manière négative par le spectateur, et leur minimisation représente un enjeu majeur.

La structure de GOP hiérarchique utilisée dans MPEG-4 SVC implique une répartition particulière des ressources disponibles pour compenser les différences de comportement entre les types d'images et leurs niveaux temporels.

Nous proposons par conséquent une **méthode de répartition du budget au niveau des images**, dont l'objectif est de **minimiser les variations de la qualité au sein d'un groupe d'images**. Une mesure de complexité de codage reflétant les particularités de la structure de GOP hiérarchique sera utilisée pour ajuster la quantité de ressources allouée en fonction des besoins de chaque image.

3.7.5 Régulation à faible complexité algorithmique

Un des principaux problèmes du codage vidéo scalable réside dans sa complexité algorithmique. L'encodage simultané de plusieurs couches et l'utilisation de la prédiction inter-couches engendrent en effet des temps de calculs élevés. Il est par conséquent essentiel d'optimiser la complexité du mécanisme de régulation de débit proposé, de manière à limiter la surcharge en calculs.

L'approche en deux passes proposée permet d'initialiser efficacement le modèle de débit et offre un contrôle de débit très précis. En revanche, le fait d'encoder chaque élément deux fois augmente dramatiquement sa complexité algorithmique. Nous proposons par conséquent d'étudier **une approche en une passe** pour MPEG-4 SVC, basée sur la **réutilisation des informations des images précédentes**. Cette méthode devrait constituer un compromis entre efficacité et complexité, plus applicable dans des contextes où les ressources en temps sont limitées.

3.7.6 Exploitation des informations inter-couches

La prédiction inter-couches permet à MPEG-4 SVC d'atteindre des performances intéressantes par rapport à des scénarios équivalents en multicast avec MPEG-4 AVC/H.264. Les dépendances statistiques entre les couches, à l'origine de son efficacité, constituent également un atout majeur pour la régulation de débit. En effet, les différents niveaux de scalabilité ont un comportement proche en termes de débit et de qualité, ce qui nous laisse envisager la possibilité d'exploiter le comportement d'une

couche pour anticiper celui des couches supérieures, dans le cas par exemple de variations brusques de l'activité.

Nous présenterons donc des travaux préliminaires permettant d'établir les bases d'une approche utilisant les dépendances inter-couches pour anticiper les erreurs de contrôle dans les couches d'amélioration.

Deuxième partie

Régulation de débit pour MPEG-4
SVC

Chapitre 4

Validation du ρ -domaine pour MPEG-4 SVC

Nous avons vu dans la section 3.5 que le ρ -domaine tire son efficacité de la linéarité de la relation qui lie ρ et le débit (équation (3.14)). Le ρ -domaine a originellement été introduit pour le contrôle de débit dans MPEG-2 [20], puis a été utilisé pour MPEG-4 Partie 2 et MPEG-4 AVC/H.264 [18, 71]. Pour ces trois standards, la linéarité de la relation a été vérifiée. En revanche, ce n'est pas le cas pour MPEG-4 SVC. Certes, les signaux AVC et SVC sont très proches car les deux standards utilisent une grande partie d'outils communs. Il semble donc naturel de penser que la relation entre ρ et le débit dans MPEG-4 SVC est elle aussi linéaire. Cependant, l'influence des mécanismes de prédiction inter-couches sur cette relation n'a pas encore été étudiée. La relation entre le débit et le QP est elle-même modifiée par la réutilisation des informations de la couche de base dans les couches d'amélioration. Il convient donc de vérifier si la relation entre ρ et le débit conserve sa linéarité dans MPEG-4 SVC, lorsque la prédiction inter-couches est utilisée. Le ρ -domaine s'appuie par ailleurs sur la relation qui existe entre ρ et le QP, exprimée dans l'équation (3.17), pour trouver la valeur optimale du QP. Il est par conséquent nécessaire de vérifier que cette relation est valide dans MPEG-4 SVC.

L'objectif de ce chapitre se résume donc à déterminer si les deux hypothèses suivantes sont vérifiées dans le cadre de MPEG-4 SVC :

Hypothèse 1 : La relation qui relie ρ à la valeur du paramètre de quantification (QP) présentée dans l'équation (3.17) est valide pour MPEG-4 SVC.

Hypothèse 2 : Si l'Hypothèse 1 est vérifiée, la relation entre ρ et le débit reste linéaire dans MPEG-4 SVC, tout en utilisant la prédiction inter-couches.



FIGURE 4.1 – Séquences d’images utilisées pour la validation du ρ -domaine dans MPEG-4 SVC.

4.1 Relation entre ρ et le QP pour MPEG-4 SVC

L’Hypothèse 1 part du principe que la relation entre ρ et la QP obéit toujours à l’équation (3.17) dans MPEG-4 SVC. Il est aisé de vérifier cette assertion en analysant le mécanisme de quantification utilisé par MPEG-4 SVC. Au niveau image et macrobloc, MPEG-4 SVC utilise le même processus de quantification que son prédécesseur MPEG-4 AVC/H.264 [39]. Les coefficients transformés sont annulés si leur valeur est inférieure à la valeur seuil de la zone morte du quantificateur et les valeurs de seuil sont les mêmes que pour MPEG-4 AVC/H.264. Par conséquent l’Hypothèse 1 est vérifiée, et les équations (3.11) et (3.17) restent valides pour MPEG-4 SVC.

4.2 Linéarité de la relation entre ρ et le débit pour MPEG-4 SVC

Pour représenter la relation qui existe entre ρ et le débit, le ρ -domaine utilise un modèle linéaire. Cette hypothèse n’a jamais été vérifiée pour MPEG-4 SVC. Pour la valider, nous observons expérimentalement la relation entre ρ et le débit pour un ensemble représentatif de configurations pour un encodage MPEG-4 SVC, appliqué à des séquences de référence.

Ainsi, nous testons la linéarité de la relation entre ρ et le débit pour les trois types de scalabilité, avec des configurations reflétant des applications classiques pour MPEG-4 SVC, résumées dans le Tableau 4.1. Elles sont basées sur le "Plan de Vérification" de MPEG-4 SVC élaboré par le JVT [35]. Les flux encodés contiennent trois couches, et les couches d’amélioration utilisent la couche immédiatement inférieure comme couche de base. La prédiction inter-couches est utilisée en mode adaptatif (*cf.* chapitre 2). Les groupes d’images (GOP) contiennent ici 8 images (soit trois niveaux d’images B hiérarchiques).

TABLE 4.1 – Configurations testées pour vérifier la linéarité de la relation entre ρ et le débit dans MPEG-4 SVC.

Scénario	Couche	Scalabilité utilisée	Couche de base	Taille d'image	Images par seconde	Taille de GOP
SPATIAL	0	-	-	176 * 144	30	8
	1	Dyadique	0	352 * 288	30	8
	2	Dyadique	1	704 * 576	30	8
QUALITÉ	0	-	-	352 * 288	30	8
	1	CGS	0	352 * 288	30	8
	2	CGS	1	352 * 288	30	8
TEMPOREL	0	-	-	352 * 288	15	2
	1	+ 2 im. B	0	352 * 288	30	4
	2	+ 4 im. B	1	352 * 288	60	8

Pour les trois scénarios, plusieurs images de chaque type et dans chaque couche sont encodées en utilisant l'ensemble des valeurs de QP comprises entre 0 et 51. Pour chaque valeur de QP, nous mesurons le débit généré par l'image à la sortie de l'encodeur et calculons la valeur de ρ correspondante d'après l'équation (3.17). Chaque niveau d'images B hiérarchiques est traité comme un type d'image à part entière, que l'on note B1 à B3. Les tests sont répétés pour les cinq séquences d'images de référence présentées par la Figure 4.1 [74]. Ces séquences d'images couvrent une grande variété de contenus, avec différentes intensités de mouvement, de texture et de variations d'activité.

Nous traçons ensuite les valeurs du débit en fonction des valeurs de ρ pour chaque valeur de QP et calculons les coefficients de corrélation entre ρ et le débit. Les Figures 4.2 à 4.4 présentent un échantillon représentatif des résultats obtenus pour les trois scénarios en traçant les valeurs du débit en fonction des valeurs de ρ pour chaque valeur de QP. Les coefficients de corrélation entre ρ et le débit sont présentés dans les Tableaux 4.2 à 4.4.

Sur l'ensemble des configurations testées, on observe une relation très linéaire entre ρ et le débit. De plus, la corrélation très élevée entre les deux confirme les observations graphiques. Le même comportement est observé sur l'ensemble des séquences utilisées pour les tests. L'utilisation de la prédiction inter-couches ne semble pas affecter la linéarité de la relation. On constate également que la relation reste linéaire pour tous

les niveaux d'images B. La structure de GOP hiérarchique n'a donc pas d'incidence sur la linéarité de la relation entre ρ et le débit.

Il est cependant intéressant de remarquer les différences entre les coefficients directeurs des droites en fonction du scénario, de la couche de scalabilité et du niveau temporel des images. Toutes les droites passent par le point $(1;0)$, qui correspond au cas où tous les coefficients transformés sont annulés, ce qui produit un débit nul. En revanche, la pente varie en fonction des configurations. Nous verrons par la suite que la détermination de cette pente représente une étape majeure du respect de la contrainte imposée pour la régulation de débit. Elle permet en effet de déterminer l'équation du modèle de débit, utilisé pour choisir la valeur de QP la plus appropriée. Plusieurs approches seront proposées pour estimer cette pente en se basant sur les caractéristiques des images de coefficients.

Les résultats présentés dans cette section montrent que la relation entre ρ et le débit est linéaire pour l'ensemble des configurations testées. L'Hypothèse 2 est donc vérifiée sur un ensemble représentatif des configurations pouvant apparaître dans MPEG-4 SVC, confirmant la validité du ρ -domaine pour MPEG-4 SVC.

4.3 Conclusion

Les résultats présentés dans ce chapitre montrent que les deux hypothèses à la base du fonctionnement du ρ -domaine sont vérifiées pour MPEG-4 SVC. Ceci signifie que les relations présentées par les équations (3.14) et (3.17) demeurent valides et qu'il est possible de les utiliser pour construire une stratégie de régulation de débit adaptés à MPEG-4 SVC. La validation du ρ -domaine pour MPEG-4 SVC a donné lieu à la publication d'un article dans la conférence internationale *IEEE International Symposium on Multimedia* à Berkeley (USA), en Décembre 2008 [65]. Les résultats établis représentent la base ayant permis l'ensemble des travaux qui suivent, en validant le modèle de débit qui constitue l'élément clé lors de la régulation de débit.

Fort des conclusions tirées de cette étude préliminaire, le chapitre suivant est consacré à l'intégration du modèle basé sur le ρ -domaine dans un mécanisme complet de régulation de débit pour MPEG-4 SVC. Le modèle présenté dans la section 3.5 y est adapté à MPEG-4 SVC et utilisé pour respecter une contrainte de débit pour chaque image. Une méthode de régulation efficace et peu coûteuse en termes de calculs est ainsi construite à partir du ρ -domaine.

FIGURE 4.2 – Débit en bits par pixel en fonction de ρ pour le scénario **spatial**.

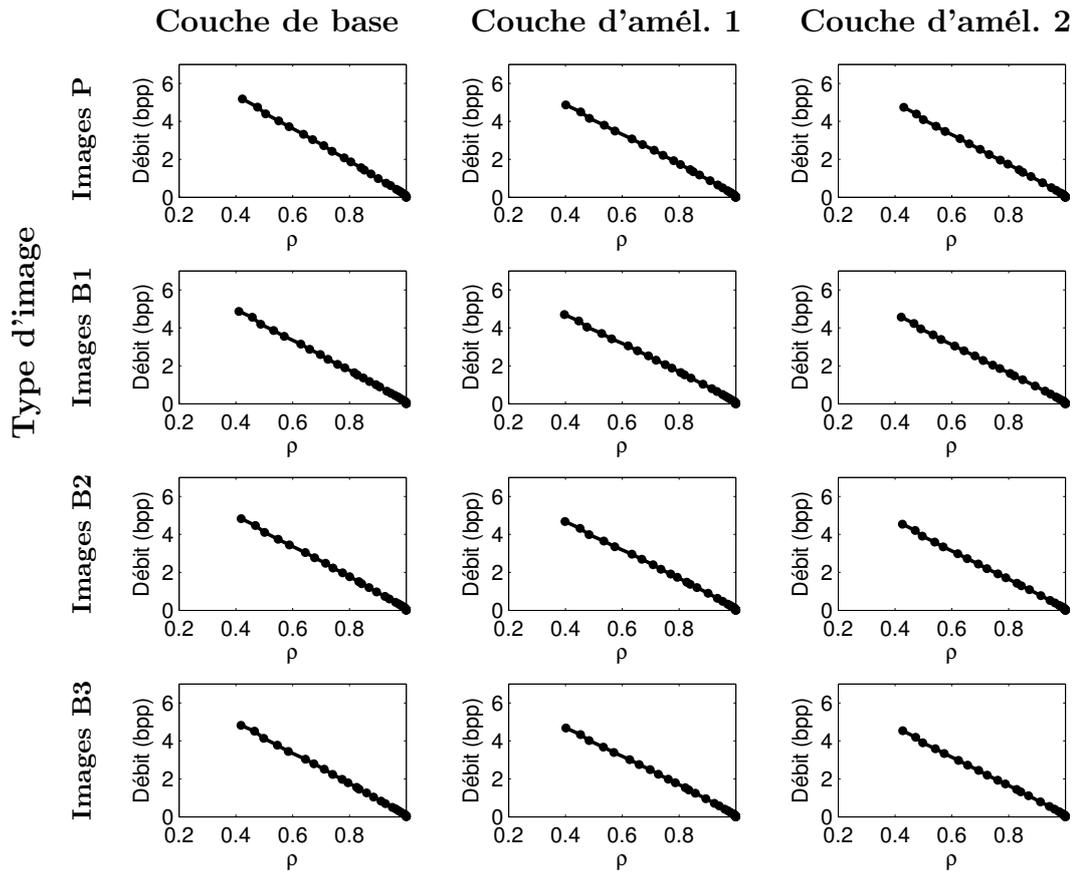
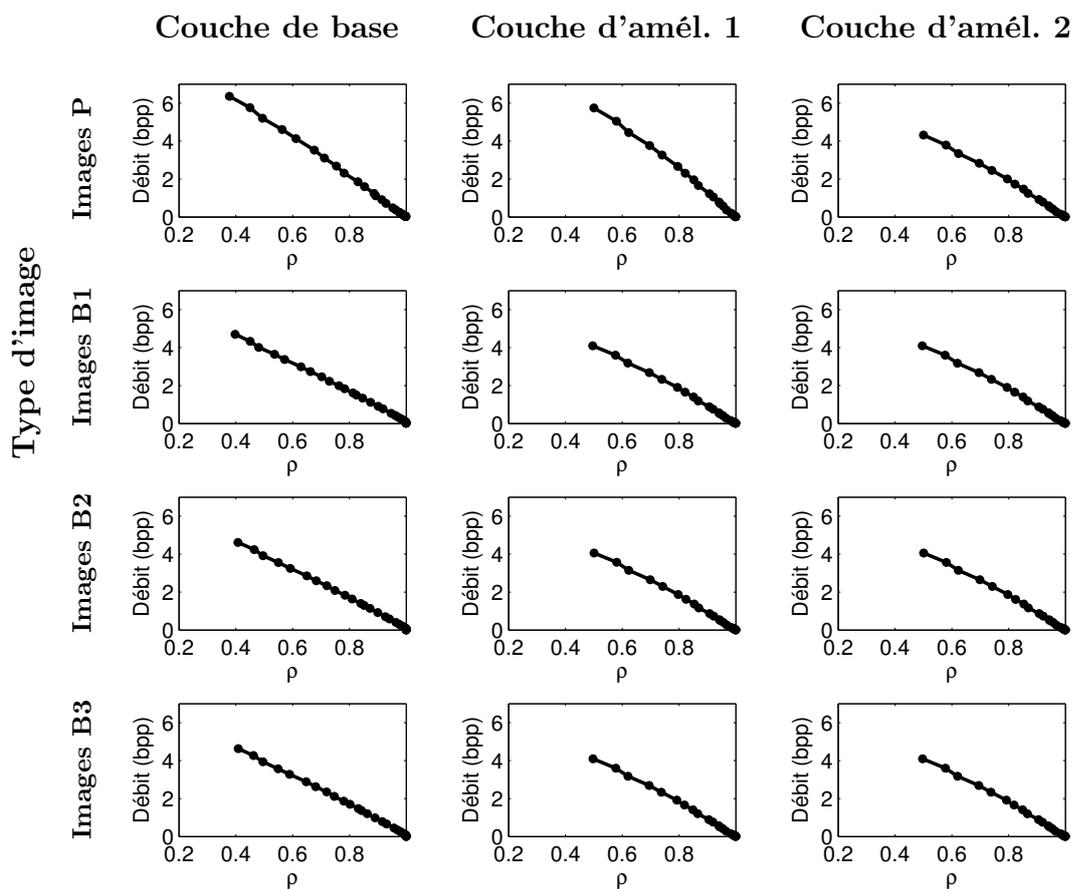


TABLE 4.2 – Coefficients de corrélation entre ρ et le débit pour le scénario **spatial**.

Type d'image	Couche de base	Couche d'amél. 1	Couche d'amél. 2
P	-0.99948	-0.99929	-0.99966
B1	-0.99954	-0.99934	-0.99964
B2	-0.99945	-0.99936	-0.99959
B3	-0.99958	-0.99940	-0.99970

FIGURE 4.3 – Débit en bits par pixel en fonction de ρ pour le scénario en **qualité**.TABLE 4.3 – Coefficients de corrélation entre ρ et le débit pour le scénario en **qualité**.

Type d'image	Couche de base	Couche d'amél. 1	Couche d'amél. 2
P	-0.99969	-0.99885	-0.99885
B1	-0.99926	-0.99877	-0.99878
B2	-0.99917	-0.99880	-0.99880
B3	-0.99926	-0.99883	-0.99884

FIGURE 4.4 – Débit en bits par pixel en fonction de ρ pour le scénario **temporel** (le symbole \emptyset signifie qu'il n'y a pas d'image de ce type dans la couche en question).

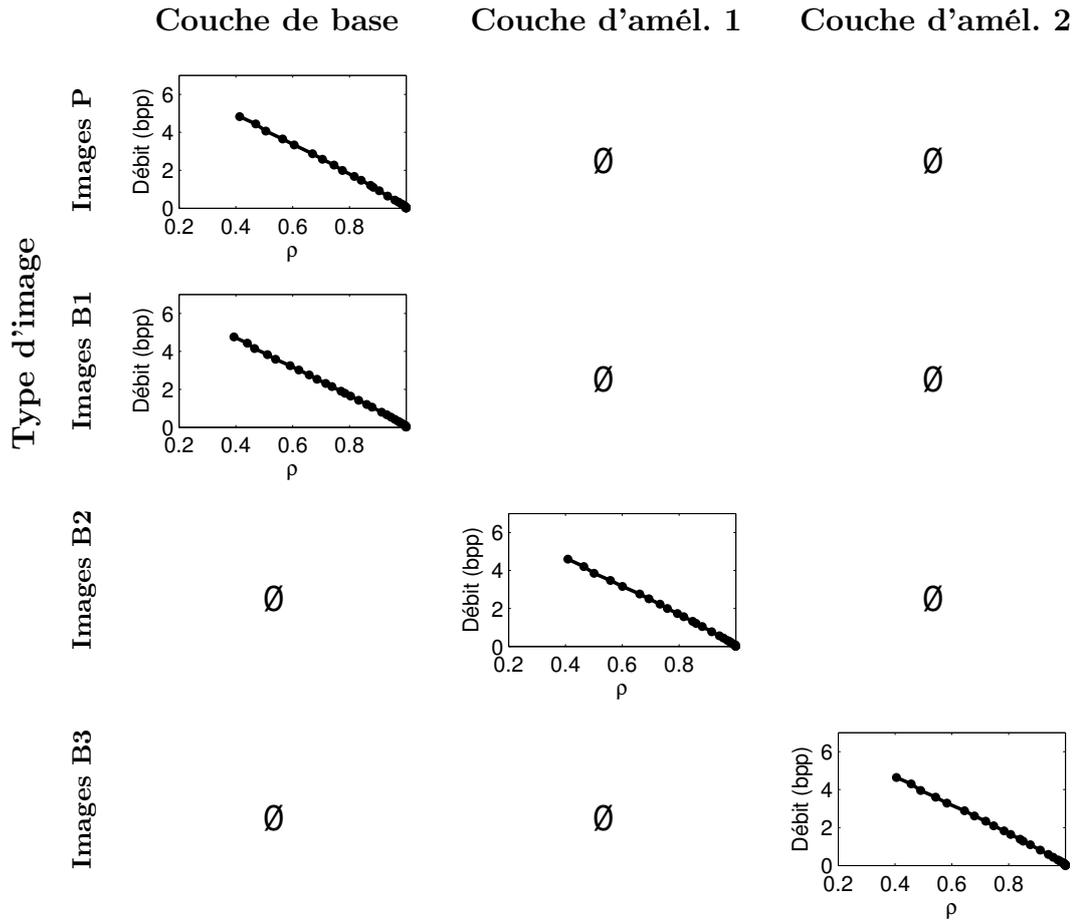


TABLE 4.4 – Coefficients de corrélation entre ρ et le débit pour le scénario **temporel**.

Type d'image	Couche de base	Couche d'amél. 1	Couche d'amél. 2
P	-0.99921	\emptyset	\emptyset
B1	-0.99960	\emptyset	\emptyset
B2	\emptyset	-0.99933	\emptyset
B3	\emptyset	\emptyset	-0.99925

Chapitre 5

Mécanisme de contrôle de débit en deux passes pour MPEG-4 SVC

Nous avons montré la validité du ρ -domaine pour MPEG-4 SVC. Il est donc possible d'utiliser cette approche dans un mécanisme de régulation de débit adapté à la vidéo scalable. L'objectif de ce chapitre est la construction d'un mécanisme élémentaire pour vérifier de façon pratique la viabilité du ρ -domaine pour MPEG-4 SVC. Une simple contrainte de débit par seconde est spécifiée, puis répercutée au niveau des GOP et des images. Le modèle de débit basé sur le ρ -domaine présenté dans la section 3.5 est ensuite utilisé pour respecter le budget alloué à chaque image. Chaque image est encodée deux fois : la première passe est utilisée pour initialiser le modèle de débit, alors que la seconde passe encode définitivement l'image en utilisant le QP choisi par le mécanisme de régulation.

Après avoir précisé la problématique du chapitre et les contraintes appliquées au flux, les méthodes de répartition de budget au niveaux des GOP et des images seront explicitées. La technique permettant de respecter le budget alloué sera ensuite présentée, et nous analyserons les performances du mécanismes en fournissant des résultats sur un ensemble représentatif de configurations.

5.1 Problématique

L'objectif des travaux présentés dans ce chapitre est de construire un mécanisme de contrôle de débit ayant une faible complexité algorithmique. En effet, le processus d'encodage de MPEG-4 SVC est d'ores et déjà complexe car plusieurs couches sont encodées en même temps, en utilisant la prédiction inter-couches pour gagner en efficacité de codage. Pour ne pas ralentir davantage le processus d'encodage, il est donc important

de minimiser la complexité algorithmique du mécanisme de régulation de débit.

Le niveau de résolution d'un mécanisme de régulation de débit a une grande influence sur sa complexité algorithmique (*cf.* chapitre 3). Si l'unité de base du contrôle est le macrobloc [87], l'allocation de ressources et le calcul du QP sont répétés un grand nombre de fois pour chaque image. Le cumul de ces opérations impute au mécanisme une complexité algorithmique élevée qui ralentit le processus d'encodage. À l'inverse, une régulation au niveau des groupes d'images possède une complexité très faible mais ne permet pas d'obtenir un contrôle suffisamment précis.

Par conséquent, si l'unité de base du contrôle est l'image, le mécanisme réalise un équilibre entre performances et complexité. La régulation est certes moins précise car les erreurs de contrôle ne peuvent pas être compensées aussi finement qu'au niveau macrobloc. En revanche, l'allocation de ressources et le calcul du QP sont exécutés un nombre de fois beaucoup plus modéré et la complexité du mécanisme est significativement plus faible.

Étant donné notre objectif de faible complexité, l'approche préconisée est donc un contrôle de débit au niveau image. La répartition de budget et le calcul du QP sont exécutés une fois par image, et aucun raffinement n'est effectué au niveau des macroblocs. L'image entière est encodée en utilisant la valeur de QP choisie.

La faible complexité du mécanisme est également assurée par l'utilisation du modèle de débit simple présenté dans la section 3.5, tirant parti des spécificités du ρ -domaine. La relation entre le débit et la QP est ainsi modélisée par une relation simple, qui permet d'obtenir une régulation de débit ayant un impact faible sur la complexité du processus d'encodage de MPEG-4 SVC.

5.2 Contraintes appliquées

Le codage vidéo MPEG-4 SVC est étroitement lié à la transmission même de la vidéo. Dans les applications ayant motivé l'élaboration du standard, on retrouve le besoin d'adéquation d'une couche de scalabilité avec une cible caractérisée par un canal de communication et un dispositif de décodage. Le canal de communication a une capacité de transfert a priori connue, généralement exprimée en termes de bits par seconde. Pour transmettre un flux vidéo sur ce canal, il est donc judicieux de contrôler le débit à la sortie de l'encodeur de façon à ne pas dépasser cette capacité.

Pour chaque couche l , nous spécifions une contrainte, notée C_l , exprimée en bits par seconde à notre mécanisme. L'objectif de la régulation est que la valeur moyenne du débit obtenu par seconde corresponde à cette contrainte. Dans ce chapitre, la régulation

de débit ne prend pas en compte la qualité du flux reconstruit. L'optimisation conjointe de la qualité et du débit fera l'objet du chapitre 6.

5.3 Allocation de budget

Lors de l'étape d'allocation du budget, le nombre de bits par seconde spécifié par la contrainte doit être réparti entre les différentes échelles du flux. Les deux échelles prises en compte par notre mécanisme sont le GOP et l'image. Il est donc nécessaire dans un premier temps de traduire la contrainte de débit par seconde en termes de débit par GOP, puis en termes de débit par image.

5.3.1 Allocation de budget niveau GOP

Au sein d'une couche l , on divise équitablement le nombre de bits par seconde voulu entre les GOP [91]. Le budget alloué à un GOP, noté G_l , est calculé de la manière suivante :

$$G_l = S_l \times \frac{C_l}{F_l}, \quad (5.1)$$

où C_l est la contrainte de débit par seconde, S_l est le nombre d'images contenues dans un GOP de la couche l et F_l est le nombre d'images par seconde dans la couche l . Pour corriger les éventuelles erreurs de contrôle accumulées au cours de l'encodage, nous ajoutons un terme de contrôle E à l'équation (5.1), qui devient :

$$G_l = S_l \times \frac{C_l}{F_l} + E. \quad (5.2)$$

Ce terme de correction donne plus ou moins de souplesse à la régulation de débit au niveau des GOP, en permettant de compenser des erreurs plus ou moins importantes d'un GOP à l'autre. Pour une valeur de E élevée, le mécanisme s'adapte aux erreurs commises sur un GOP en les compensant sur le GOP suivant, ce qui permet d'avoir un débit par seconde plus proche de la contrainte C_l . En revanche, la correction de cette erreur cause des variations significatives du budget du GOP suivant, ce qui peut produire des variations de qualité indésirables. Au contraire, une valeur trop faible ne permet pas de compenser les erreurs de régulation, qui peuvent s'accumuler et faire perdre en précision à la régulation par seconde. Pour nos expérimentations, la valeur de E est fixée à 30% du budget d'un GOP, qui s'est avérée être un compromis satisfaisant.

Une fois le budget du GOP alloué, il doit être réparti entre les différentes images qui le composent. La méthode de répartition du budget au niveau des images est décrite dans le paragraphe suivant.

5.3.2 Allocation de budget niveau image

Dans ce chapitre, la stratégie d'allocation de budget au niveau des images est élémentaire. Le budget du GOP est réparti de façon uniforme entre les images, tout en propageant les erreurs de débit au fur et à mesure. Pour la i -ème image d'un GOP de la couche l , le budget R_t est calculé de la manière suivante :

$$R_t = \frac{G_l}{S_l - i}, \quad (5.3)$$

où S_l est le nombre d'images contenues dans un GOP de la couche l . Dans l'équation (5.3), G_l est en fait le budget restant pour le GOP courant, si bien que les éventuelles erreurs de débit faites sur les images précédentes sont prises en compte lors du calcul du budget de l'image courante.

Comme nous l'avons vu dans le chapitre 3, des stratégies de répartition de budget plus élaborées permettent de limiter les variations de la qualité, d'assurer une bonne transmission du flux en évitant la saturation des buffers de communication ou d'optimiser l'efficacité de codage d'un ensemble de couches scalables. Ce type de stratégies sera abordé dans le chapitre 6.

L'objectif de ce chapitre est bien de valider l'utilisation du ρ -domaine en tant que modèle de débit pour MPEG-4 SVC. Nous ne cherchons donc pas ici à construire une stratégie d'allocation optimale mais à vérifier qu'en appliquant une répartition simple, la régulation de débit se déroule convenablement.

Une fois le budget de l'image calculé, le mécanisme de régulation de débit doit choisir la valeur de QP pour laquelle le débit se rapproche le plus du budget.

5.4 Respect du budget

Étant donné un budget R_T à respecter, le mécanisme de contrôle de débit s'appuie sur le modèle de débit basé sur le ρ -domaine pour calculer la valeur optimale du QP. Dans un premier temps, la valeur de ρ correspondant au budget R_T est calculée selon l'équation (3.13) :

$$\rho_t = \rho(R_t) = \frac{R_0 - R_t \times (1 - \rho_0)}{R_0}.$$

Pour fonctionner, le modèle de débit requiert une phase d'initialisation. En effet, l'équation (3.13) utilise deux valeurs R_0 et ρ_0 qu'il faut déterminer. Ils permettent de déterminer la pente de la droite représentant la relation entre ρ et le débit pour l'image à encoder. Dans ce chapitre, ces valeurs sont obtenues en encodant l'image lors d'une première passe.

5.4.1 Passe d'initialisation

L'image est encodée une première fois avec une valeur initiale de QP notée q_0 . La valeur de q_0 est déterminée de la manière suivante :

$$q_0 = \begin{cases} QP_{INIT} & \text{pour la première image encodée;} \\ q_{t-1} & \text{pour les autres images,} \end{cases} \quad (5.4)$$

où q_{t-1} est la valeur du QP utilisée pour encoder l'image précédente de même type et dans la même couche. Au début de l'encodage, la valeur de QP_{INIT} peut être évaluée à partir de la contrainte de débit. Une méthode simple de calcul de QP_{INIT} sera décrite dans la section 5.4.3.

Le nombre de bits R_0 et la valeur $\rho_0 = \rho(q_0)$ sont ensuite calculés. Une fois les valeurs de R_0 et ρ_0 déterminées, la valeur de ρ_t est calculée selon l'équation (3.13). La valeur du QP qui minimise l'écart avec le budget de l'image peut alors être calculée, comme décrit dans la section suivante.

5.4.2 Calcul du QP optimal

La valeur q_t de QP qui minimise la différence avec le budget R_t est déterminée par :

$$q_t = \arg \min_{q \in [0;51]} |\rho_t - \rho(q)|, \quad (5.5)$$

où $\rho(q)$ est donné par l'équation (3.17). Dans la pratique, le QP varie modérément entre la première et la seconde passe si la valeur de q_0 est choisie en respectant l'équation (5.4). Aussi, afin de réduire le nombre d'opérations effectuées, nous limitons la variation du QP entre la première et la deuxième passe d'encodage à une valeur maximale Δ_q . L'équation (5.5) devient alors :

$$q_t = \arg \min_{q_0 - \Delta_q \leq q \leq q_0 + \Delta_q} |\rho_t - \rho(q)|. \quad (5.6)$$

Nous avons pu observer qu'en fixant la valeur de Δ_q à 6, la valeur de q_t trouvée est systématiquement la même en utilisant les équations (5.5) et (5.6). Le nombre d'itérations nécessaires pour trouver la valeur de q_T est réduit de 52 à 13, sans perte de précision.

Du fait de l'échelle choisie pour notre mécanisme de régulation, il n'est généralement pas possible d'atteindre le budget exact alloué à une image. En effet le même QP est utilisé pour encoder tous les macroblocs de l'image, et un incrément de 1 dans sa valeur produit naturellement des variations importantes dans le débit nécessaire pour coder l'image. La valeur du budget à atteindre se trouve généralement comprise entre deux valeurs de débit qu'il est possible d'atteindre :

$$\begin{cases} R(q_t - 1) \geq R_t \geq R(q_t) & \text{si } R_t \geq R(q_t), \\ R(q_t) \geq R_t \geq R(q_t + 1) & \text{si } R_t \leq R(q_t). \end{cases} \quad (5.7)$$

En conséquence, le débit \hat{R}_t le plus proche du budget qu'il est possible d'atteindre, appelé le *budget réel*, est défini tel que :

$$\hat{R}_t = R(q_t), \quad (5.8)$$

où $R(q_t)$ est donné par l'équation (3.15).

Une fois la valeur du QP correspondant le mieux au budget trouvée, l'image est encodée une seconde fois, comme décrit dans la section suivante.

5.4.3 Calcul de la valeur initiale du QP

Au début de la séquence, on observe généralement une phase pendant laquelle le processus de prédiction spatio-temporelle doit se stabiliser. Une initialisation du mécanisme de régulation avec une valeur de QP adaptée à la contrainte de débit peut donc favoriser une convergence plus rapide vers un fonctionnement nominal.

Pour déterminer la valeur du QP pour les premières images du flux, nous avons analysé la relation entre le QP choisi après convergence de la régulation et la valeur de la contrainte de débit exprimée en bits par pixel selon la formule :

$$bpp(C_l) = \frac{C_l}{(W_l \times H_l) F_l}, \quad (5.9)$$

où C_l est la contrainte appliquée à la couche l , $(W_l \times H_l)$ représente la dimension des images dans la couche l et F_l sa fréquence temporelle.

Pour chaque type de scalabilité, une table de correspondance entre le QP et cette contrainte a été construite d'après un échantillon représentatif de séquences et de débits. Pour plus de détails, le lecteur est invité à se reporter à cette table, disponible en annexe D.

5.4.4 Seconde passe d'encodage

La valeur trouvée pour q_t est ainsi finalement utilisée pour encoder l'image lors d'une seconde passe. Le nombre de bits R_f généré et la valeur $\rho_f = \rho(R_f)$ sont ensuite mesurés pour évaluer les performances du contrôle de débit. Enfin, le budget du GOP G_l est mis à jour de la manière suivante :

$$G_l = \max\{0; G_l - R_f\}. \quad (5.10)$$

Dans la section suivante, nous analysons les performances du mécanisme décrit en termes de précision du contrôle de débit sur un ensemble de configurations représentatives des utilisations prévues pour MPEG-4 SVC. Les trois types de scalabilité sont testés, en combinaison avec la prédiction inter-couches.

5.5 Résultats expérimentaux

Dans cette section nous présentons les performances du mécanisme de contrôle de débit en deux passes en termes d'adéquation entre le budget alloué et le débit effectif. Le Tableau 5.1 présente les configurations testées. Elles sont similaires à celles utilisées dans le chapitre 4, avec des GOP contenant 16 images. Une contrainte de débit par seconde est spécifiée pour chaque couche, d'après le plan de vérification du standard MPEG-4 SVC élaboré par le JVT [35]. Nous utilisons les mêmes séquences d'images que dans le chapitre 4 (*cf.* Tableau 4.1).

Pour évaluer les performances du contrôle de débit, nous utilisons deux métriques. Tout d'abord, l'efficacité du mécanisme au niveau des images est mesurée par l'erreur δ_t entre le débit R_f obtenu pour une image et le débit alloué R_t . Cette erreur est calculée de la manière suivante :

$$\delta_t = \frac{R_f - R_t}{R_t} * 100. \quad (5.11)$$

TABLE 5.1 – Configurations testées pour évaluer les performances du contrôle de débit en deux passes.

Scénario	Couche	Couche de base	Taille d'image	Images par seconde	Taille de GOP	Débit par seconde (kbits/sec.)
SPATIAL	0	-	176 * 144	30	16	100
	1	0	352 * 288	30	16	400
	2	1	704 * 576	30	16	1200
QUALITÉ	0	-	352 * 288	30	16	400
	1	0	352 * 288	30	16	1200
	2	1	352 * 288	30	16	3600
TEMPOREL	0	-	352 * 288	15	4	200
	1	0	352 * 288	30	8	400
	2	1	352 * 288	60	16	800

De manière similaire, la précision du modèle de débit indépendamment des erreurs dues à l'utilisation d'une valeur de QP unique par image est mesurée par l'erreur $\hat{\delta}_t$ avec le budget réel \hat{R}_t de l'image :

$$\hat{\delta}_t = \frac{R_f - \hat{R}_t}{\hat{R}_t} * 100. \quad (5.12)$$

Pour ces deux erreurs, la valeur moyenne et l'écart type sont mesurés (*respectivement* $\mu(\delta_t)$, $\sigma(\delta_t)$, $\mu(\hat{\delta}_t)$ et $\sigma(\hat{\delta}_t)$). Afin d'illustrer le comportement du contrôle de débit, nous traçons également l'évolution du débit par rapport à la contrainte fixée pour chaque image, chaque GOP et chaque seconde.

5.5.1 Précision de la régulation de débit

Les Figures 5.2 à 5.7 constituent un ensemble représentatif des performances du mécanisme pour les trois scénarios d'encodage. Des résultats complémentaires sur l'ensemble des séquences d'images utilisées pour nos tests sont disponibles en annexe A. L'erreur $\hat{\delta}_t$ moyenne ne dépasse pas 7% du débit alloué pour l'ensemble des configurations testées, ce qui confirme la précision de la régulation. On observe une efficacité particulièrement élevée pour le scénario en qualité. En effet les images des trois couches ont les mêmes dimensions, ce qui rend la prédiction inter-couches très stable et permet

une régulation très précise. Dans le scénario temporel, les débits des différentes couches se superposent car le débit alloué aux images des trois couches est identique. L'erreur reste cependant comparable aux scénarios spatial et en qualité.

L'écart type de l'erreur varie selon les séquences, traduisant les variations dans l'efficacité du contrôle de débit en fonction du contenu à encoder. Pour les séquences SOCCER et HOCKEY, l'écart type est globalement très bas (inférieur à 15). Il s'agit de deux séquences d'images contenant une grande quantité de mouvement mais peu de texture. Les changements d'activité sont donc modérés, ce qui rend la régulation plus aisée. En revanche, sur les séquences HARBOUR, CREW et CITY, l'écart type de l'erreur est élevé (supérieur à 15). Dans CREW, les variations de luminosité dues aux flashes des photographes provoquent des variations brusques d'activité qu'il est très difficile de prédire. Dans CITY et HARBOUR, les images sont très texturées, avec des détails très fins et un mouvement de caméra par rapport à une scène comportant plusieurs plans d'objets. Les données à coder changent alors de manière importante lorsque la valeur de QP varie et il est difficile de prédire le comportement des images à partir du modèle de débit.

Au niveau des GOP et du débit par seconde, le contrôle de débit est néanmoins très efficace. L'écart entre la consigne et le débit obtenu est très faible sur l'ensemble des configurations testées. L'objectif principal de débit par seconde est par conséquent atteint avec une grande précision.

5.5.2 Interprétation des erreurs de contrôle

De façon prévisible, l'erreur δ_t entre le budget alloué et le débit atteint est globalement plus élevée que l'erreur $\hat{\delta}_t$ avec le débit qu'il est réellement possible d'atteindre. En effet, le seuil provoqué par l'utilisation d'une seule valeur de QP par image implique que le débit réel varie autour de la contrainte appliquée. De la même manière, l'écart type est moins élevé entre le budget réel et le débit atteint.

La raison majeure pour laquelle on observe des imprécisions dans le contrôle de débit est l'inter-dépendance entre le QP et la prédiction spatio-temporelle dans MPEG-4 SVC (*cf. chicken & egg dilemma*, chapitre 3). La relation entre ρ et le QP s'appuie sur les coefficients transformés de l'image pour déterminer la valeur optimale du QP. Or, cette relation suppose que l'image transformée ne change pas entre la première et la seconde passe. Cependant la valeur du QP intervient au moment de la prédiction spatio-temporelle des macroblocs, car elle conditionne le choix des modes de prédiction et des vecteurs de mouvement. Les changements de QP entre les deux passes affectent donc le contenu de l'image transformée, ce qui rend impossible de

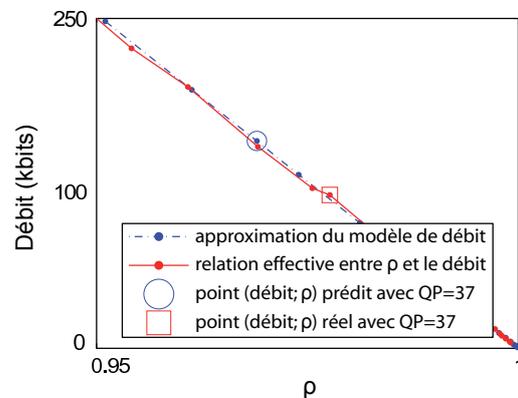


FIGURE 5.1 – Décalage entre le modèle de débit et le débit effectif.

connaître son contenu exact après l'avoir encodée avec le QP choisi pour la seconde passe. La Figure 5.1 illustre ce phénomène en montrant le décalage entre la valeur du débit prédite à partir des données de la première passe et du modèle linéaire, et la valeur réelle obtenue après la deuxième passe. Sur cet exemple, la première passe a utilisé un QP de 41 pour encoder l'image et initialiser le modèle de débit. À partir des coefficients transformés de l'image, le mécanisme de contrôle de débit prédit que le débit produit sera égal à 160 kbits en utilisant un QP de 37. Malheureusement, l'encodage de l'image avec un QP de 37 produit en réalité un débit de 120 kbits, soit environ 25% de moins. En résultat, le modèle de débit ne parvient pas à déterminer la bonne valeur de QP, malgré le fait qu'il suit parfaitement la relation entre ρ et le débit.

Les imprécisions augmentent avec l'écart entre le QP utilisé pour la première passe et le QP choisi pour la seconde passe, car les modifications des coefficients transformés sont de plus en plus nombreuses et rendent le modèle moins précis. Les quantités de mouvement et de texture interviennent également, car elles provoquent davantage de variations dans l'image transformée. Par ailleurs, il est visible que dans la plupart des configurations, l'erreur de débit mesurée sur les images B est supérieure à l'erreur mesurée sur les images P. La raison de cet écart est que le nombre de possibilités pour prédire un macrobloc est plus élevé dans les images B que dans les images P. Ainsi, les changements de QP ont une probabilité plus élevée de faire changer la prédiction et les erreurs du modèle de débit sont plus importantes.

Ces erreurs pourraient être réduites par un raffinement du choix du QP au niveau des macroblocs. Cependant, comme nous l'avons déjà évoqué, ce type d'approche augmente fortement la quantité de calculs effectuée et n'est pas en adéquation avec notre objectif

de faible complexité algorithmique. Qui plus est, les performances du mécanisme au niveau des GOP et du débit par seconde restent très satisfaisantes et la contrainte principale est respectée avec une grande précision.

5.5.3 Influence de la taille des GOP

La taille des GOP a une influence sur le comportement du débit. En effet, le calcul du budget alloué à une image par rapport au budget restant pour le GOP (et non pas le budget total) permet de compenser les erreurs de débit au cours de l'encodage du GOP. Par conséquent, si la taille du GOP est grande, les erreurs peuvent être davantage compensées avant le début du GOP suivant. La Figure 5.8 compare les débits obtenus pour des GOP de 8, 16 et 32 images. Pour des GOP de 8 images, le débit varie fortement autour de la contrainte. Pour des GOP de 16 images, ces variations sont plus modérées du fait du plus grand nombre de compensations possibles. Pour les GOP de 32 images, les variations sont encore atténuées et le mécanisme est moins sensible aux erreurs de contrôle localisées sur des périodes courtes.

5.5.4 Qualité des flux reconstruits

Pour évaluer la qualité des flux reconstruits, nous mesurons le PSNR sur chaque image (*cf.* équation (3.10)). La Figure 5.9 présente quelques exemples du PSNR obtenu par le mécanisme de contrôle de débit en deux passes présenté dans ce chapitre. On constate des oscillations brutales, qui ont un effet néfaste sur la qualité ressentie. Ce phénomène trouve son origine dans le fait que chaque image se voit allouer un débit égal au sein d'un GOP. Les types d'images et leurs différences d'efficacité ne sont pas prises en compte, ce qui amène une qualité moindre sur les images P que sur les images B. Une solution à ce problème est proposée dans le chapitre suivant, avec l'élaboration d'une méthode de répartition du budget prenant en compte les différents types d'images.

5.5.5 Temps d'encodage

Pour évaluer l'impact de notre mécanisme sur la rapidité de l'encodage, nous comparons les temps obtenus avec et sans régulation⁵ dans le Tableau 5.2. L'impact de la régulation de débit en deux passes est clairement visible, avec une augmentation du temps moyen d'encodage de près de 100%.

Le mécanisme de régulation en lui-même a une complexité faible grâce à l'utilisation du ρ -domaine et au choix du paramètre de quantification au niveau des images. Cependant, chaque image est encodée deux fois, ce qui augmente la complexité du processus

TABLE 5.2 – Temps d’encodage moyen par image avec et sans régulation de débit.

	SPATIAL	QUALITÉ	TEMPOREL
Sans régulation	1.19 s.	0.56 s.	0.37 s.
Régulation 2 passes	2.46 s.	1.13 s.	0.78 s.
<i>Augmentation</i>	<i>105%</i>	<i>102%</i>	<i>108%</i>

d’encodage dans son ensemble. Ce problème se traduit par une durée d’encodage multipliée par deux par rapport à un fonctionnement de l’encodeur sans régulation de débit. Notre méthode reste donc difficilement applicable dans un contexte où les ressources en temps sont limitées.

La réduction du temps d’encodage avec régulation représente un enjeu majeur. Une méthode permettant de supprimer la première passe d’encodage sera donc abordée dans la suite de notre travail.

5.6 Conclusion

Dans ce chapitre nous avons présenté un mécanisme de régulation de débit qui utilise le ρ -domaine pour prédire le comportement du débit en fonction du paramètre de quantification. Un budget est alloué à chaque image, en répartissant de manière uniforme le budget du GOP. Pour initialiser le modèle de débit, une première passe est effectuée sur chaque image. La valeur optimale du QP est ensuite choisie grâce au modèle de débit, et l’image est encodée une seconde fois avec le QP choisi.

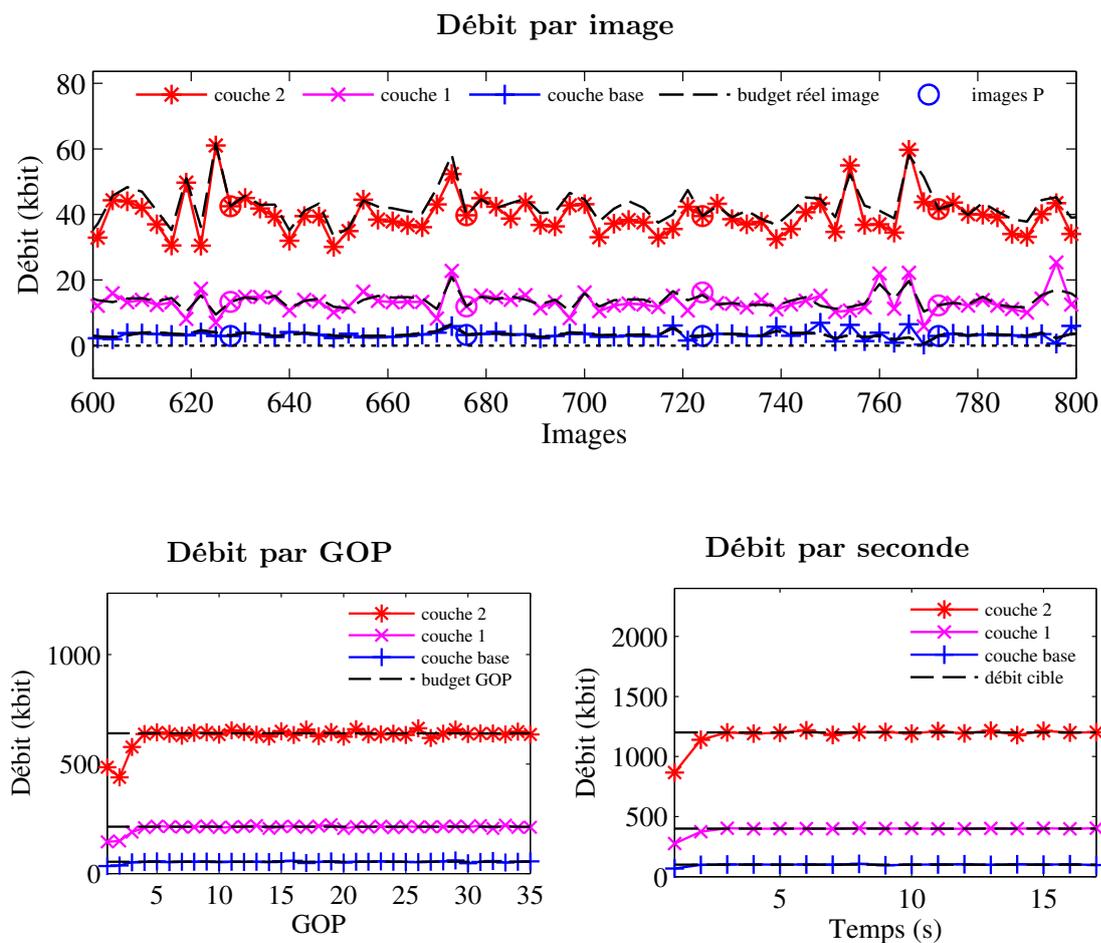
Les résultats montrent une grande efficacité de la régulation, en affichant une erreur de contrôle inférieure à 7% du débit alloué à chaque image. La contrainte de débit par GOP et par seconde est atteinte avec une grande précision. Les trois types de scalabilités proposées par MPEG-4 SVC sont prises en compte, ainsi que la prédiction inter-couches et les images B hiérarchiques. L’objectif du chapitre est donc atteint, avec la validation du ρ -domaine comme approche efficace de la régulation de débit sur MPEG-4 SVC.

Le mécanisme de régulation présenté dans ce chapitre a été publié également dans la conférence internationale *IEEE International Symposium on Multimedia* à Berkeley (USA) en Décembre 2008 [65].

5. Temps indicatifs mesurés sur une machine équipée d’un processeur Intel Xeon E5430 cadencé à 2.66 GHz et de 4 Go de mémoire RAM.

En marge des bonnes performances du mécanisme, on constate quelques imprécisions pour les séquences contenant beaucoup d'activité. Toutefois les séquences en question sont reconnues pour être difficiles à coder, du fait de la complexité de leur contenu. En outre, malgré les imprécisions constatées, les performances de l'algorithme sur ces séquences restent néanmoins très satisfaisantes.

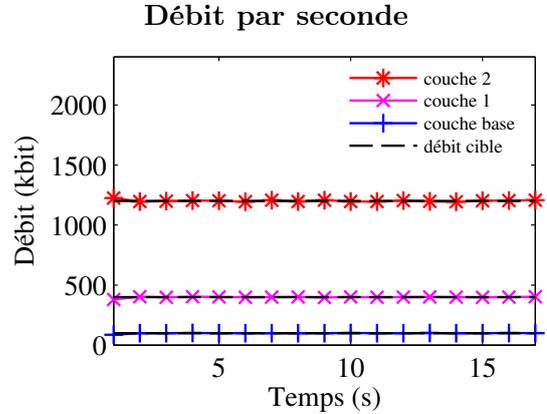
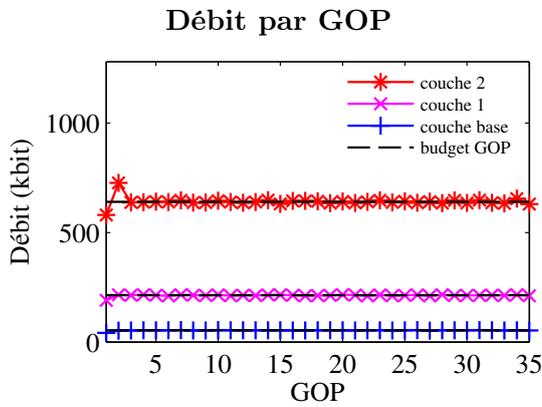
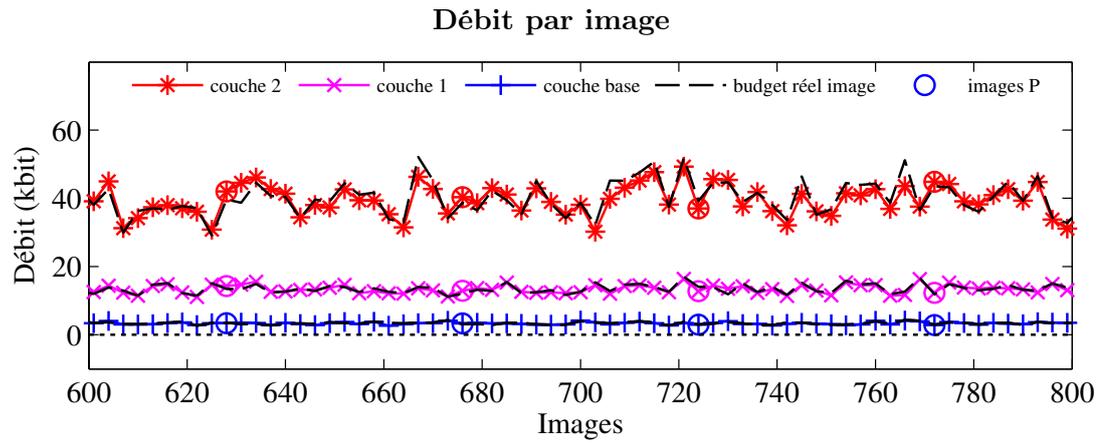
Du point de vue de la qualité, le fait d'allouer un débit équivalent à toutes les images d'un GOP produit des oscillations du PSNR dans les flux reconstruits. L'impression visuelle en est considérablement dégradée, particulièrement pour des GOP de grande taille. Pour obtenir une qualité plus constante, une répartition plus élaborée du budget des GOP entre les images qui les constituent est envisagée dans le chapitre suivant. Une méthode de répartition prenant en compte les différences d'efficacité de codage des types d'images est élaborée, pour réduire les variations de la qualité au sein d'un GOP dans le flux reconstruit.

FIGURE 5.2 – Performances du contrôle de débit pour le scénario **spatial** avec des GOP de 16 images sur la séquence **HARBOUR**.

Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	+1.00 %	37.76	+1.77 %	36.57
couche 1	-3.64 %	19.22	-3.39 %	14.36
couche 2	-6.02 %	12.65	-6.45 %	6.24

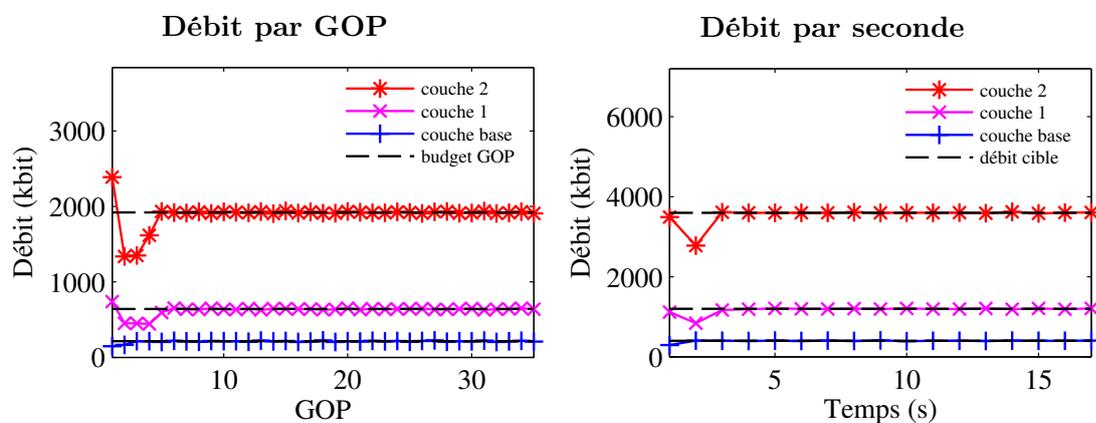
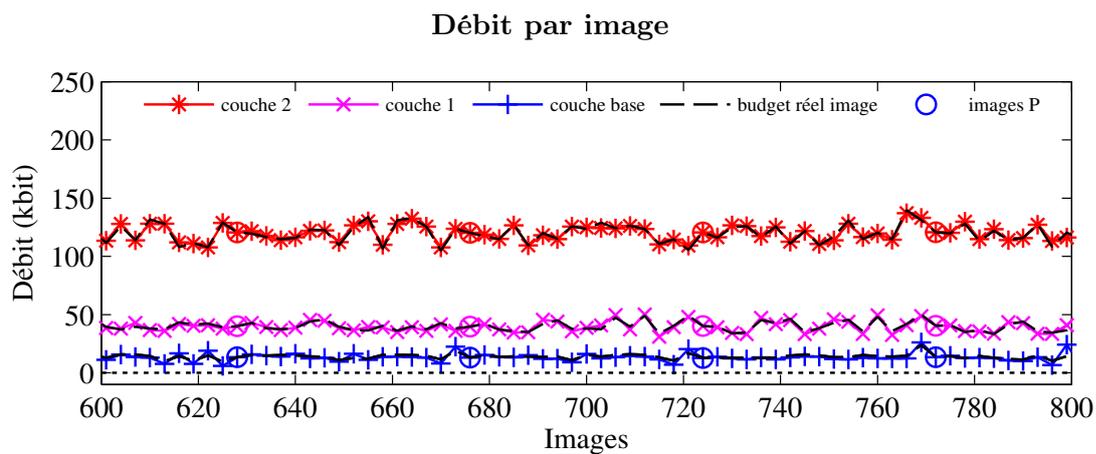
FIGURE 5.3 – Performances du contrôle de débit pour le scénario **spatial** avec des GOP de 16 images sur la séquence **HOCKEY**.



Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-2.22 %	8.71	-1.51 %	5.64
couche 1	-2.34 %	8.86	-1.74 %	4.51
couche 2	-4.16 %	9.25	-3.15 %	5.35

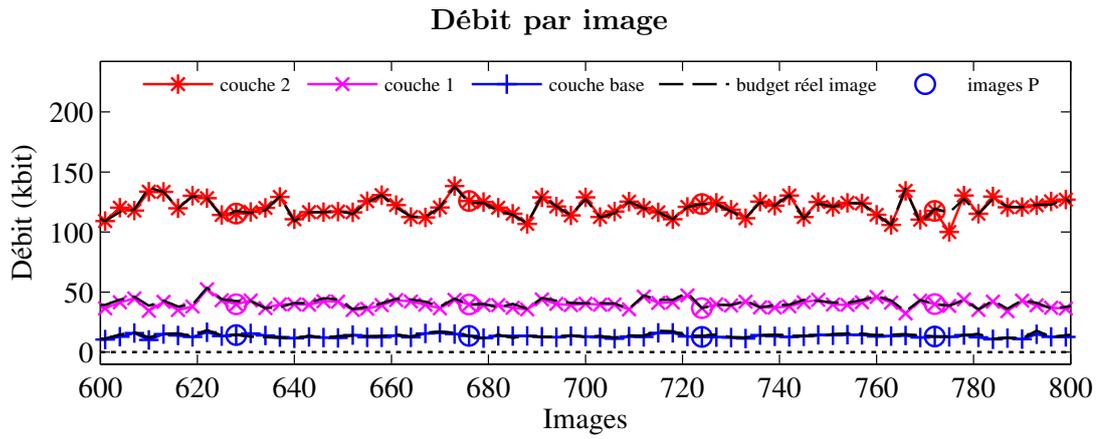
FIGURE 5.4 – Performances du contrôle de débit pour le scénario en **qualité** avec des GOP de 16 images sur la séquence **HARBOUR**.



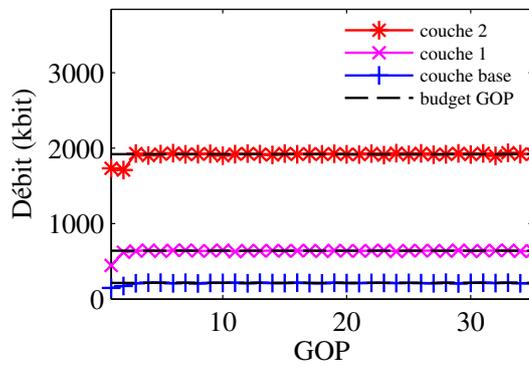
Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-5.66 %	17.75	-5.37 %	13.10
couche 1	-1.68 %	14.19	-0.95 %	5.10
couche 2	+0.47 %	7.88	+0.29 %	1.48

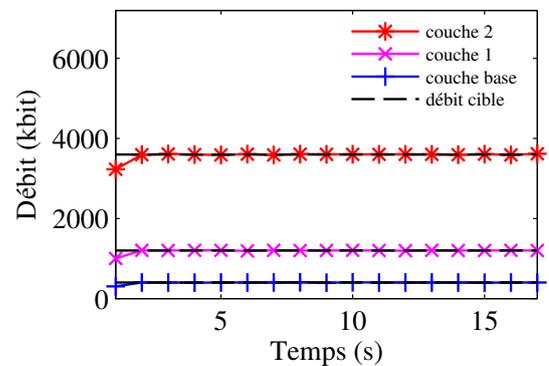
FIGURE 5.5 – Performances du contrôle de débit pour le scénario en **qualité** avec des GOP de 16 images sur la séquence **SOCCER**.



Débit par GOP

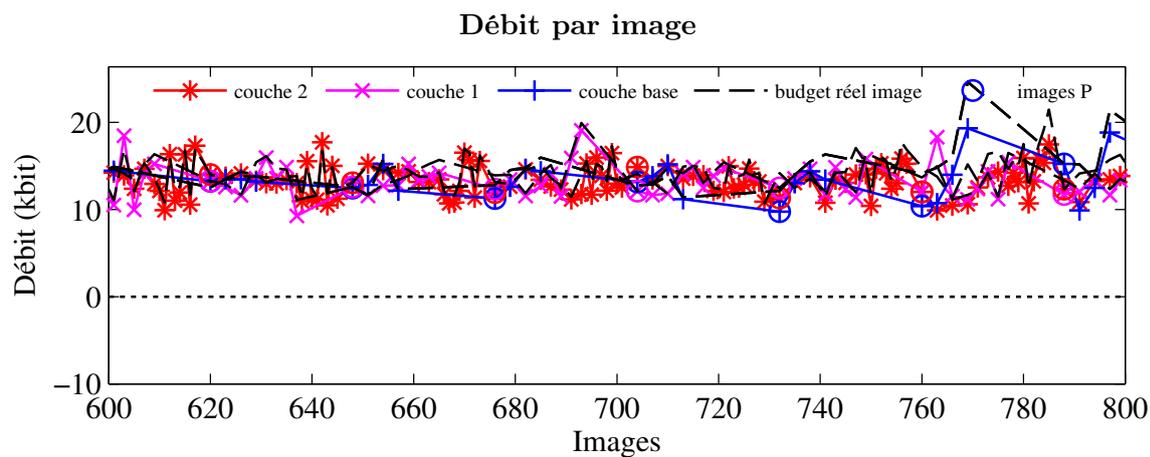


Débit par seconde



Erreur de débit par image pour chaque couche.

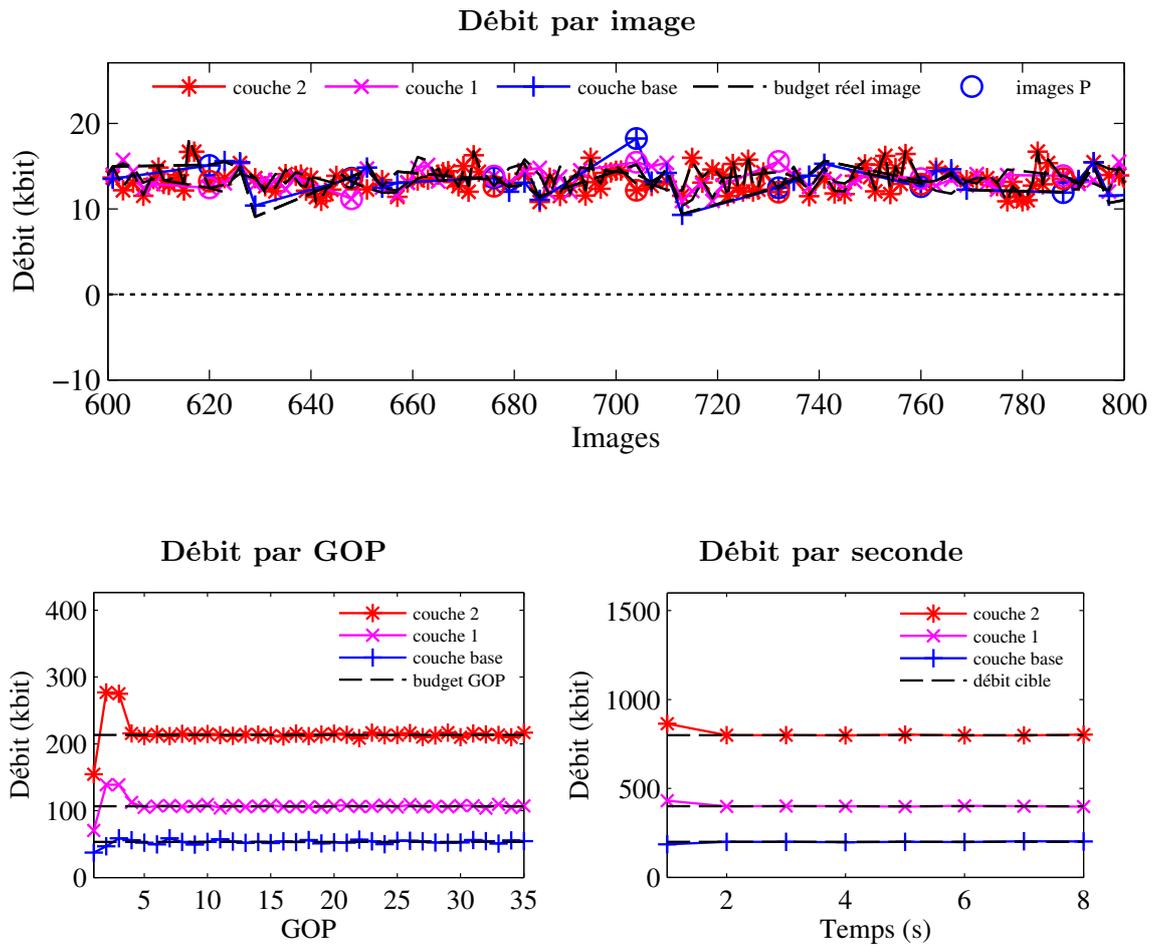
	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-6.35 %	12.91	-5.92 %	8.55
couche 1	-3.61 %	9.82	-3.04 %	3.81
couche 2	+0.27 %	5.35	+0.51 %	2.09

FIGURE 5.6 – Performances du contrôle de débit pour le scénario **temporel** avec des GOP de 16 images sur la séquence **SOCCER**.

Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-2.07 %	12.82	-1.67 %	11.51
couche 1	-4.36 %	11.92	-3.07 %	6.89
couche 2	-4.44 %	11.77	-4.30 %	7.14

FIGURE 5.7 – Performances du contrôle de débit pour le scénario **temporel** avec des GOP de 16 images sur la séquence **HOCKEY**.



Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-0.80 %	11.72	-0.47 %	10.29
couche 1	-0.68 %	7.92	-0.72 %	5.48
couche 2	-2.50 %	8.71	-1.73 %	5.15

FIGURE 5.8 – Comportement du mécanisme de contrôle de débit sur la séquence CREW avec différentes tailles de GOP.

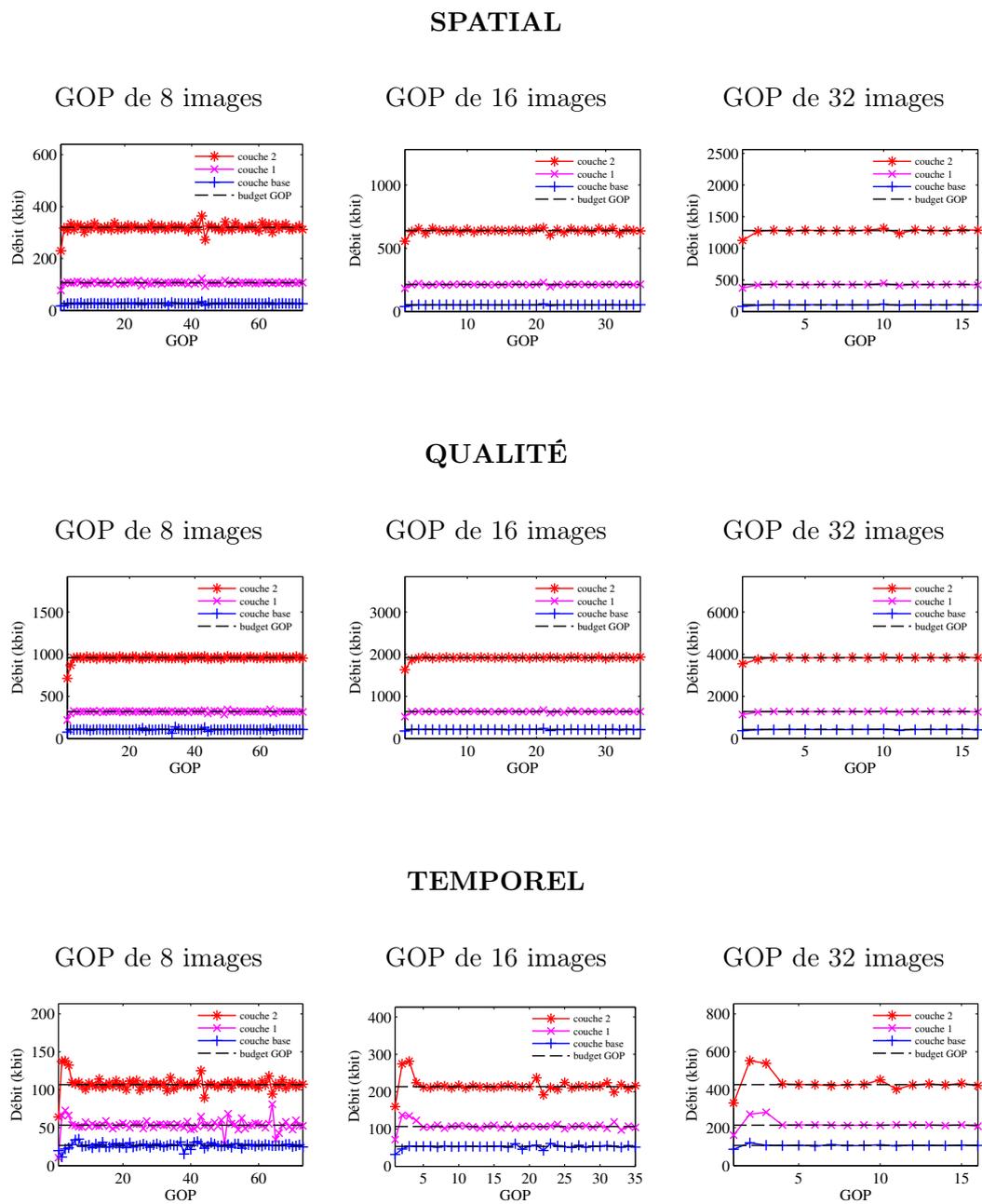
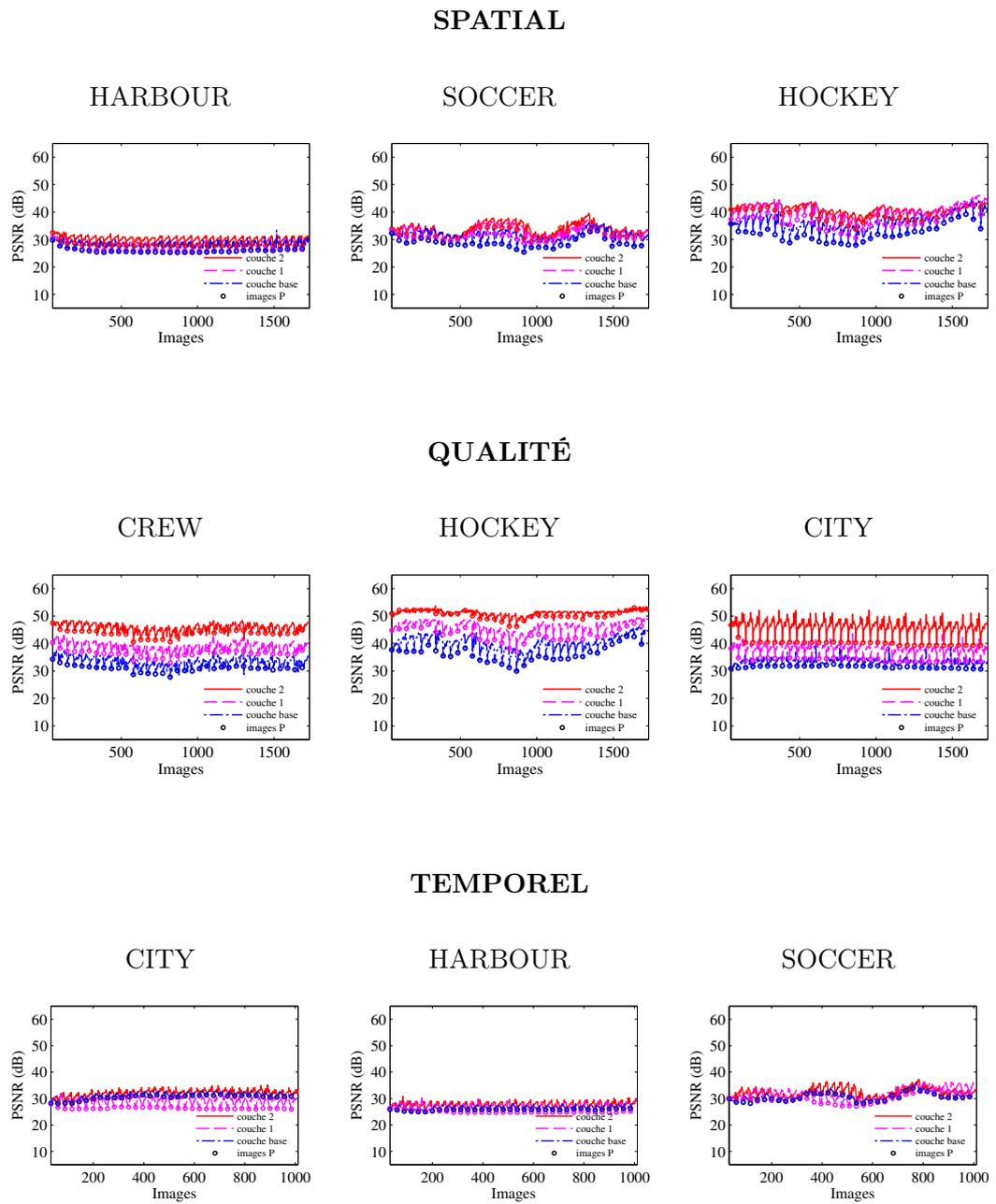


FIGURE 5.9 – Variations de la qualité des images dans les flux reconstruits avec le mécanisme de régulation de débit en deux passes.



Chapitre 6

Régulation de débit et qualité constante

Le mécanisme de régulation de débit présenté dans le chapitre précédent alloue un budget constant à chaque image. On observe des variations sensibles de la qualité sur de courts intervalles, qui provoquent des effets de scintillement clairement déplaisants visuellement. Ces variations sont dues aux différences d'efficacité de codage selon le type et le niveau temporel d'une image dans le GOP. La Figure 6.1 montre les variations du PSNR observées sur les trois couches de la séquence HARBOUR en utilisant le mécanisme du chapitre précédent avec la scalabilité en qualité et des GOP de 16 images. Les images P ont une qualité moindre que les images B, tandis que la qualité des images B hiérarchiques croît avec leur niveau temporel. Le Tableau 6.1 confirme cette observation en montrant l'écart entre le PSNR moyen des différents types d'images sur cette séquence. Selon la couche considérée, le PSNR des images P varie de 2.85 dB à 6.32 dB par rapport à celui des images B de plus haut niveau temporel. Ces différences de PSNR traduisent des oscillations rapides de la qualité qui provoquent les phénomènes de scintillement observés visuellement.

De manière à supprimer ces effets de scintillement, l'objectif de ce chapitre est d'obtenir une qualité constante sur toutes les images d'un GOP. Nous proposons une méthode de répartition de budget basée sur l'efficacité de codage au niveau image permettant de compenser les différences entre chaque type et chaque niveau temporel.

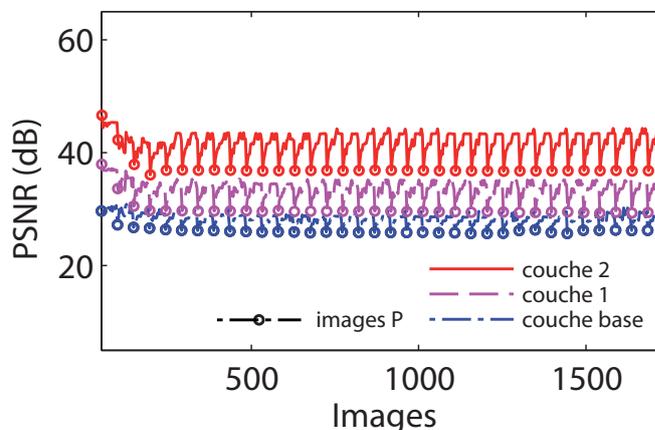


FIGURE 6.1 – Variations du PSNR observées pour la séquence HARBOUR avec la scalabilité en qualité.

TABLE 6.1 – PSNR moyen par type d'image et par couche.

type d'image	P	B1	B2	B3	B4
couche de base	25.99 dB	27.03 dB	27.53 dB	28.08 dB	28.84 dB
1 ^e couche d'amél.	29.62 dB	31.48 dB	32.25 dB	33.40 dB	34.59 dB
2 ^e couche d'amél.	36.87 dB	38.89 dB	40.20 dB	41.54 dB	43.19 dB

6.1 Qualité constante par GOP

Sous une contrainte de débit constant, il n'est pas envisageable d'obtenir une qualité constante sur l'ensemble de la séquence. Les variations d'activité dans la séquence d'origine au cours du temps impliquent des variations de qualité à l'échelle de la séquence, lorsque l'on fixe le débit. Ces variations font partie intégrante des données à coder et doivent être conservées.

En revanche, des variations brusques de qualité à l'échelle des images sont perçues de façon négative par le système visuel humain. En conséquence, minimiser ces variations représente un enjeu majeur pour améliorer la qualité ressentie. C'est pourquoi l'objectif de ce chapitre est de minimiser les variations locales de la qualité à l'échelle d'un groupe d'images.

Dans ce contexte, les variations de la qualité à la jonction entre les groupes d'images peuvent constituer un premier problème. Cependant, le budget alloué à chaque GOP

étant constant, on peut supposer que la qualité moyenne atteinte sur un GOP sera proche de celle obtenue sur les GOP adjacents, et que la transition se fera de manière relativement continue.

La structure de GOP hiérarchique utilisée dans MPEG-4 SVC a des conséquences sur l'efficacité de codage des images en fonction de leur position dans le GOP. Il est primordial de prendre ces différences en compte pour obtenir une qualité constante sur l'ensemble du groupe d'images.

6.2 Efficacité de codage des différents types et niveaux d'images

Le nombre de bits nécessaires pour atteindre une qualité donnée dépend avant tout de l'efficacité de la prédiction spatio-temporelle. En effet, plus la prédiction est mauvaise, plus l'entropie du signal résiduel est élevée et plus le nombre de bits nécessaires pour coder les données est élevé. Les écarts de qualité constatés entre les différents types d'image sont dus à l'utilisation de différents outils de prédiction spatio-temporelle. Les images I utilisent exclusivement la prédiction intra image, dont l'efficacité est limitée. La prédiction inter-images, utilisée pour les images P, est en comparaison beaucoup plus performante. L'utilisation de deux images de référence pour les images B permet d'améliorer encore les performances de codage par rapport aux images P. Par conséquent, à débit égal les images I ont une qualité plus faible que les images P, qui ont une qualité elle même plus faible que les images B.

De plus, la structure de GOP hiérarchique utilisée dans MPEG-4 SVC a des conséquences analogues sur l'efficacité de codage des différents niveaux temporels d'images B. Les images de niveau bas sont codées par rapport à des images relativement éloignées, rendant la prédiction temporelle moins efficace que pour les images du haut de la hiérarchie, codées par rapport à des images proches. À débit égal, la qualité des images B hiérarchiques augmente donc avec leur niveau temporel.

Ainsi, pour compenser ces différences d'efficacité de codage et par conséquent diminuer les variations de la qualité au sein du flux vidéo, l'idée développée ici est de répartir le budget disponible en fonction de l'efficacité de codage des images. Les images nécessitant plus de débit pour atteindre une qualité donnée se verront allouer un plus grand nombre de bits que les images dont les besoins sont plus faibles. Des approches similaires ont été proposées dans [86] et [85]. Une mesure de l'efficacité de codage des images est utilisée pour répartir le budget du GOP parmi les images. Cependant, le premier article

ne montre pas l'impact de la répartition sur les variations de la qualité. Dans le second, les résultats ne montrent pas une atténuation significative des variations du PSNR par rapport à une stratégie de répartition équivalente à celle présentée dans le chapitre 5.

Notre approche se base sur les caractéristiques des images après l'encodage et produit une diminution significative des variations du PSNR, vérifiée sur un ensemble représentatif de configurations. Dans un premier temps nous introduirons une mesure de l'efficacité de codage d'une image, que nous utiliserons ensuite comme un facteur de poids pour répartir le budget disponible entre les différentes images d'un GOP.

6.3 Mesure de l'efficacité de codage d'une image

L'objectif de ce chapitre est de fixer une contrainte de PSNR constant au niveau d'un GOP et d'en déduire le budget à allouer à chaque image. Il est malheureusement difficile d'évaluer la relation entre PSNR et débit directement. Nous allons donc introduire un paramètre intermédiaire dans cette relation, en nous basant sur le processus de quantification utilisé dans MPEG-4 SVC.

Nous avons vu dans le chapitre 1 qu'un coefficient quantifié Z_{ij} peut être exprimé en fonction de la valeur du coefficient transformé Y_{ij} correspondant et du paramètre de quantification q utilisé, selon l'équation (1.6) :

$$Z_{ij} = \text{round} \left(Y_{ij} \times \frac{MF_{ij}}{PF_{ij} \times 2^{qbits}} \right),$$

où PF et MF sont des matrices de décalage définies dans le chapitre 1 et

$$qbits = 15 + \lfloor q/6 \rfloor.$$

Les opérations d'arrondis figurent dans l'équation (1.6) pour conserver des valeurs entières et ont une influence très faible sur la valeur de Z_{ij} . En les ignorant, on obtient

$$Z_{ij} = \frac{Y_{ij}}{2^{q/6}} \times \frac{MF_{ij}}{PF_{ij} \times 2^{15}}. \quad (6.1)$$

Ainsi, en regroupant les termes de décalage constants dans une constante C_{ij} définie par

$$C_{ij} = \frac{MF_{ij}}{2^{15} \times PF_{ij}},$$

on peut écrire

$$Z_{ij} = \frac{Y_{ij} \times C_{ij}}{2^{q/6}}. \quad (6.2)$$

Intuitivement, il est aisé de voir que le contenu du bloc Z de coefficients quantifiés conditionne fortement le débit en sortie de l'encodeur. Si l'entropie de ce bloc est élevée,

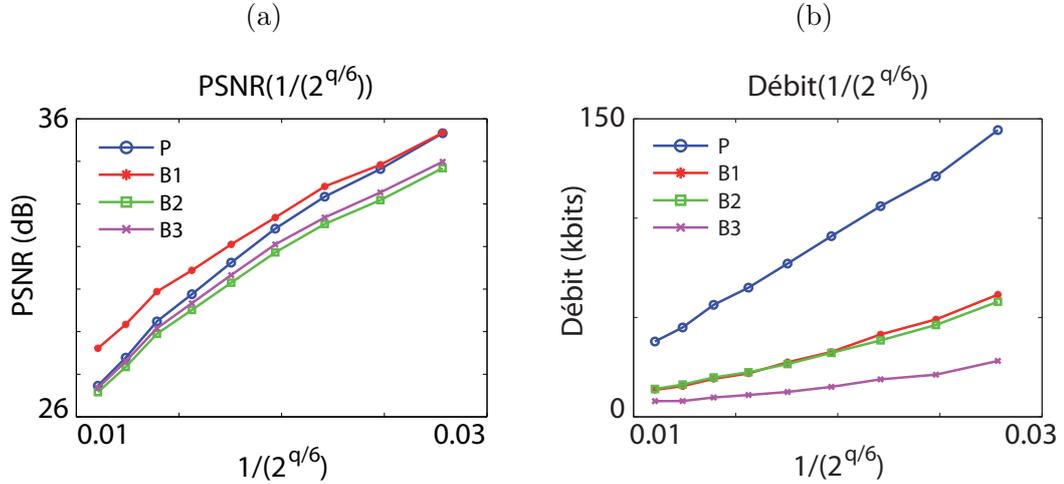


FIGURE 6.2 – (a) relation entre le PSNR et $1/(2^{q/6})$ pour différents types d'images dans un GOP hiérarchique de la séquence HARBOUR, avec des valeurs de QP comprises entre 28 et 36. (b) relation entre le débit et $1/(2^{q/6})$ pour la même configuration.

le nombre de bits nécessaire pour le représenter après le codage entropique est élevé. Inversement, si l'entropie du bloc quantifié est faible, le nombre de bits nécessaire est faible. Ainsi, d'après l'équation (6.2) on peut supposer que le paramètre $Q = 1/(2^{q/6})$ a une influence sur le débit en sortie de l'encodeur. De même, on peut supposer qu'il a une influence sur le PSNR.

La Figure 6.2 trace les relations entre Q , le PSNR et le débit pour différents types d'images dans un GOP, avec des valeurs de QP représentatives des variations observées en pratique. Cet exemple reflète la tendance générale observée sur un ensemble varié de configurations, à savoir des relations relativement linéaires, dont la pente varie en fonction du type et du niveau d'image étudié.

Ainsi, à partir d'une contrainte de PSNR constant, il est possible de déterminer les valeurs de Q correspondantes pour chaque image, puis de réutiliser ces valeurs de Q pour déduire les débits nécessaires pour obtenir une qualité constante sur chaque image.

De plus, l'aspect linéaire de ces relations nous permet de relier le PSNR au débit de manière relativement simple. D'une part, il est possible de modéliser la relation entre le PSNR et Q à l'aide d'un modèle linéaire :

$$PSNR_i = \alpha_i Q_i + \delta_i, \quad (6.3)$$

où α_i et δ_i sont des constantes et $i \in [0, S - 1]$ dénote la i^e image dans un GOP de taille S . Sur la Figure 6.2, on observe que l'ordonnée à l'origine de la droite reliant le

PSNR à la valeur de Q est relativement constant pour tous les types d'images présentés. Cet exemple est représentatif du comportement observé sur un plus large ensemble de configurations. Nous considérerons donc que δ_i est constant pour toutes les images.

Notre objectif étant de fixer un PSNR constant sur l'ensemble du GOP, on peut donc écrire que

$$\frac{Q_i}{Q_j} = \frac{\alpha_j}{\alpha_i}, \quad \forall i, j \in [0, S - 1]. \quad (6.4)$$

En d'autres termes, il existe un rapport constant entre les valeurs de Q permettant d'obtenir un PSNR constant sur toutes les images d'un GOP.

De manière analogue, l'aspect linéaire de la relation entre Q et le débit permet d'écrire

$$R_i = \beta_i Q_i + \epsilon_i, \quad (6.5)$$

où β_i et ϵ_i sont également des constantes et R_i est le débit engendré par l'encodage de l'image i avec la valeur Q_i . De plus, On constate que la relation entre le débit et la valeur de Q passe par un point relativement proche de l'origine. En effet, une valeur faible de Q dénote une valeur élevée de q , et dans ce cas le débit est très faible, quelque soit le type d'image. Nous admettrons donc ici que la valeur ϵ_i est négligeable.

Ainsi, en combinant les équations (6.3) et (6.7), on peut obtenir un lien entre le PSNR et le débit, par l'intermédiaire du paramètre Q .

On peut remarquer que selon les équations (6.7) et (6.4), il est possible d'écrire

$$\frac{R_i}{R_j} = \frac{\beta_i Q_i}{\beta_j Q_j} = \frac{\beta_i \alpha_j}{\beta_j \alpha_i}, \quad \forall i, j \in [0, S - 1]. \quad (6.6)$$

Autrement dit, sous une contrainte de PSNR constant, il existe également un rapport constant entre les débits nécessaires pour chaque image du GOP. Ce rapport traduit le poids relatif qu'il faut accorder à une image lors de la répartition du budget d'un GOP, en comparaison aux autres images.

D'après l'équation (6.7), définissons la constante K_i telle que

$$K_i = R_i / Q_i = R_i \times 2^{q_i/6}. \quad (6.7)$$

Cette constante permet d'évaluer l'efficacité de codage relative d'une image par rapport aux autres. Elle traduit le débit qu'il faut allouer à l'image pour obtenir un PSNR constant au niveau du GOP.

La valeur de K_i peut être calculée pour chaque image après la première passe d'encodage, de la manière suivante :

$$K_i = R_{p_i} \times 2^{q_{p_i}/6}, \quad (6.8)$$

où R_{p_i} et q_{p_i} sont respectivement le débit obtenu lors de la première passe, et le QP utilisé.

Au niveau du GOP, la mesure d'efficacité de codage K_i de chaque image est utilisée pour lui allouer une portion du budget du GOP permettant d'avoir un PSNR constant.

6.4 Répartition du budget du GOP d'après l'efficacité de codage

Le budget G_l alloué au GOP est calculé de la même manière qu'au chapitre 5 en utilisant l'équation (5.2). Pour répartir le budget du GOP entre les différentes images, nous utilisons ensuite la mesure K_i définie par l'équation (6.8). On alloue à chaque image une quantité de bits qui reflète son efficacité de codage par rapport au reste des images du GOP.

Pour chaque image i dans la couche l , la valeur de $K_{i,l}$ est calculée après la première passe d'encodage selon l'équation (6.8). Le budget R_t alloué est ensuite calculé de la manière suivante :

$$R_t = \frac{K_{i,l}}{\sum_{j=0}^{S_l-1} K_{j,l}} \times G_l, \quad (6.9)$$

où S_l est la taille du GOP dans la couche l , G_l est le budget restant pour le GOP avant l'encodage de F_i et F_j désigne la j^e image du GOP. Malheureusement, il n'est pas possible de connaître les valeurs de K_i pour les images qui n'ont pas encore été encodées. On utilise alors comme estimation la valeur $\tilde{K}_{j,l}$ calculée sur la dernière image encodée dans la couche l et ayant le même type⁴. L'équation (6.9) devient alors

$$R_t = \frac{K_{i,l}}{\sum_{j < i} K_{j,l} + \sum_{j \geq i} \tilde{K}_{j,l}} \times G_l. \quad (6.10)$$

Comme nous l'avons montré en introduction de ce chapitre, chaque type d'image a sa propre efficacité de codage. Cette efficacité peut également varier au cours du temps de par les changements d'activité possibles dans les données à coder. La mesure d'efficacité de codage est donc mise à jour après l'encodage de chaque image, de manière à prendre en compte ces variations.

4. On utilisera ici les opérateurs de comparaison “<” et “>” pour symboliser la précedence entre deux images dans l'ordre d'encodage. Si l'image i est encodée avant l'image j , on note $i < j$. Dans le cas contraire, on note $i > j$. Il est à noter que la relation de précedence entre deux images dans l'ordre d'encodage n'implique pas forcément leur précedence dans l'ordre d'affichage (cf. Figure 1.3).

Le reste du mécanisme de régulation est identique à celui présenté dans le chapitre 5. Après le calcul du budget alloué à une image, la valeur de ρ et le QP qui s'en approchent le plus sont calculés. L'image est ensuite encodée lors de la deuxième passe et le budget du GOP est mis à jour en fonction du nombre de bits obtenu. La section suivante présente les résultats obtenus par la méthode de répartition de budget prenant en compte l'efficacité de codage des images. Nous analysons d'une part l'atténuation des variations de la qualité entre les images d'un GOP, et nous validons d'autre part le mécanisme de régulation en mesurant les erreurs de débit observées.

6.5 Résultats expérimentaux

Pour évaluer les améliorations apportées par la méthode de répartition de budget en fonction des types d'images, nous utilisons les mêmes configurations et scénarios d'encodage que dans le chapitre 5 (*cf.* Tableau 5.1). Pour évaluer l'atténuation des variations de la qualité, l'écart type moyen du PSNR par GOP est comparé pour les deux méthodes de répartition. L'évolution du PSNR par image est également tracée, ainsi que les débits par image, par GOP et par seconde. Les Figures 6.3 à 6.8 regroupent les résultats obtenus sur un échantillon représentatif des séquences utilisées. Des résultats complémentaires sont présentés en annexe B.

6.5.1 Comportement du débit au niveau des images

Au niveau des images, on observe clairement que la méthode de répartition de budget alloue davantage de bits aux images P qu'aux images B, conformément à notre postulat de départ. De plus, les images B hiérarchiques se voient allouer une quantité de bits qui diminue quand leur niveau temporel augmente. Ce phénomène est particulièrement visible pour le scénario spatial sur les séquences HARBOUR et SOCCER (Figures 6.3 et 6.4), pour lesquelles les budgets alloués aux images d'un même type forment des paliers par niveau temporel. Pour le scénario en qualité, on constate que pour certains GOP, le budget alloué à l'image P est inférieur à celui alloué à l'image B qui la suit. La différence entre les budgets des niveaux temporels de B hiérarchiques est également plus faible, notamment dans la couche de plus haute qualité. On peut attribuer ce phénomène à l'efficacité de la prédiction inter-couches, qui permet d'obtenir les mêmes performances de codage sur les images P et les premiers niveaux d'images B hiérarchiques. Pour le scénario temporel, les débits ont un aspect plus ordonné par rapport au mécanisme de régulation du chapitre 5.

6.5.2 Atténuation des variations de la qualité

Les variations de la qualité sur un GOP peuvent être mesurées par l'écart type du PSNR sur ce GOP. La comparaison de l'écart type moyen obtenu par la répartition à budget constant au niveau image avec celui obtenu par la répartition en fonction de l'efficacité de codage permet de voir l'apport majeur de ce chapitre. On constate en effet une atténuation des variations comprise entre un facteur 2 dans les configurations les plus problématiques (séquences avec beaucoup d'activité) et un facteur 5 pour les meilleures d'entre elles. Le PSNR est sensiblement plus constant entre deux images P (qui délimitent un GOP), tandis que les variations globales de la qualité sont conservées. La prise en compte de l'efficacité de codage permet donc de façon significative les variations de la qualité à l'intérieur d'un GOP.

Pour les séquences CREW et CITY, la distinction entre les différents types d'images est plus difficile à observer. La raison principale est liée à la présence de variations très élevées de l'activité. Il en résulte une répartition plus désordonnée des débits et une partie des variations du PSNR subsiste. Pour le scénario en qualité, l'écart entre le PSNR des images P et celui des images B (tous niveaux confondus) souffre encore de légères variations. La mesure d'efficacité de codage n'est pas parfaitement adaptée dans ce cas. En revanche, l'atténuation des variations du PSNR est particulièrement remarquable dans le cas du scénario temporel. La couche de base ne contient que des images P et un niveau d'images B hiérarchiques, tandis que la couche intermédiaire contient un niveau hiérarchique de plus. Le peu d'images présentes dans ces deux couches peut expliquer les performances intéressantes de la prise en compte de l'efficacité de codage. Sur la couche la plus haute, les variations du PSNR sont également très faibles. Il est bon de rappeler que la scalabilité temporelle consiste uniquement à ajouter des images B d'une couche à l'autre, sans raffinement de qualité ou de résolution. Elle est donc logiquement plus simple à gérer pour un mécanisme de régulation de débit.

En outre, on constate une légère augmentation de la qualité à l'échelle de la séquence. En effet, les ressources en débit sont mieux réparties entre les images, ce qui améliore les performances de codage de la structure hiérarchique du GOP. Les images P ont une qualité plus élevée, et représentent une meilleure base pour la prédiction des images B. L'amélioration produite se répercute donc sur l'ensemble de la structure d'images B hiérarchiques et la qualité globale augmente.

6.5.3 Évaluation visuelle du gain apporté sur la qualité

Le PSNR ne reflète pas réellement la qualité perçue par un utilisateur. Visuellement, l'atténuation des variations de la qualité est nettement identifiable. Le phénomène d'oscillation de la qualité observé sans la répartition d'après l'efficacité de codage a disparu. Une impression de plus grande fluidité est apportée sur l'ensemble de la séquence.

En outre, on constate que les différences de PSNR encore visibles sur les figures présentées, notamment pour le scénario en qualité, sont difficilement perceptibles à l'œil nu.

6.5.4 Influence de la prédiction inter-couches

La Figure 6.5.4 trace l'évolution de la mesure K_i dans les différentes couches pour les trois scénarios de test. On observe que les images de la couche de base sont globalement les plus coûteuses, ce qui est particulièrement vrai pour les images P. En effet, les images des couches d'amélioration utilisent la prédiction inter-couches et exploitent une partie du débit utilisé pour coder la couche de base. Elles sont par conséquent moins coûteuses à coder, dans le sens où il faut un débit moins important pour obtenir une qualité égale à celle de la couche de base.

Ces résultats montrent clairement l'apport de la prédiction inter-couches pour mettre à profit les redondances entre les différents niveaux de scalabilité.

6.5.5 Conclusion

Dans ce chapitre nous avons proposé une nouvelle méthode de répartition de budget pour minimiser les variations de la qualité entre les images à l'intérieur d'un GOP. Dans un premier temps, une mesure de l'efficacité de codage au niveau image a été introduite pour symboliser les différences entre les types et niveaux d'images. L'efficacité de codage de chaque image est calculée après la phase de pré-encodage puis utilisée comme facteur de poids pour répartir le budget du GOP. Les résultats montrent que cette méthode permet d'atténuer les variations du PSNR de façon significative pour les trois types de scalabilités. Ils mettent clairement en évidence l'effet de lissage de la qualité permis par la méthode. Les travaux présentés dans ce chapitre ont notamment fait l'objet d'une publication dans la conférence internationale *SPIE Electronic Imaging : Video Communications and Image Processing*, à San Jose (USA), en Janvier 2009 [64].

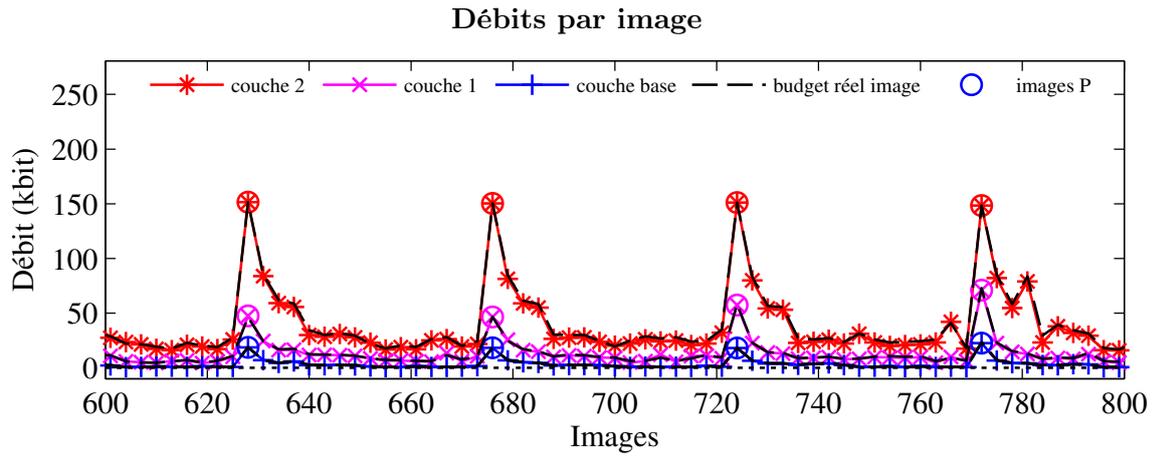
L'impression visuelle est grandement améliorée, avec la suppression de l'effet de scintillement de la qualité. Les différences entre l'efficacité de codage des images P et

des images B hiérarchiques sont clairement mises en évidence. La répartition prend en compte uniquement la mesure K_i , sans exercer de contrôle sur le rapport entre le budget des images P face au budget des images B. En résultat, les images P se voient allouer un budget plus important du fait de leurs besoins plus élevés en débit pour obtenir la même qualité que les images B. De plus, on constate un phénomène analogue entre les différents niveaux d'images B hiérarchiques. Les images de bas niveau ont une efficacité de codage moins élevée que les images de haut niveau. L'influence de la prédiction inter-couches est finalement visible, en permettant dans certains cas à des images P d'avoir une efficacité plus grande que l'image B qui les suit.

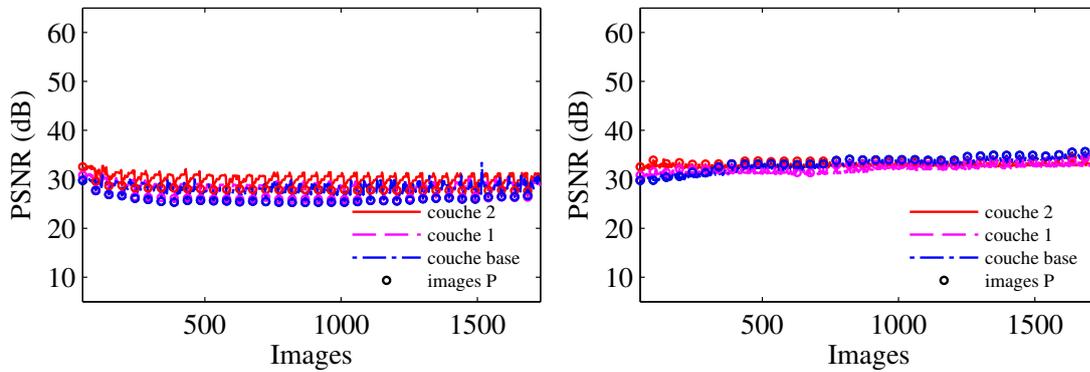
Malgré ses performances intéressantes à la fois en termes de précision de la régulation exercée et de variations réduites de la qualité, notre mécanisme souffre d'un défaut majeur. En effet, la phase de pré-encodage nécessaire pour initialiser le modèle de débit pénalise fortement la rapidité du mécanisme (*cf.* Tableau 5.2).

Afin de diminuer la complexité en termes de calculs du mécanisme, il est primordial de se passer de la première passe d'encodage sur chaque image. Pour cela, l'idée développée dans le chapitre suivant est d'exploiter les corrélations entre images successives. En effet, le contenu d'une séquence d'images varie de façon relativement continue, et deux images adjacentes sont empreintes de fortes dépendances. De façon à retirer la première passe d'encodage, le modèle de débit peut donc se baser sur les informations obtenues sur les images précédentes. Chaque image sera ainsi encodée une seule fois et l'impact du mécanisme sur la durée d'encodage devrait être fortement réduit.

FIGURE 6.3 – Améliorations apportées sur la qualité pour le scénario **spatial** sur la séquence **HARBOUR**.



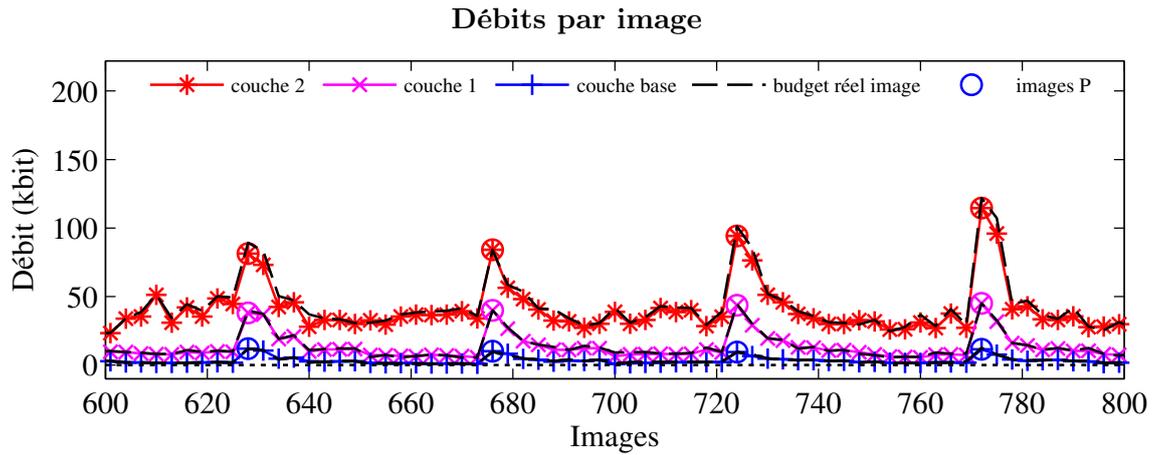
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



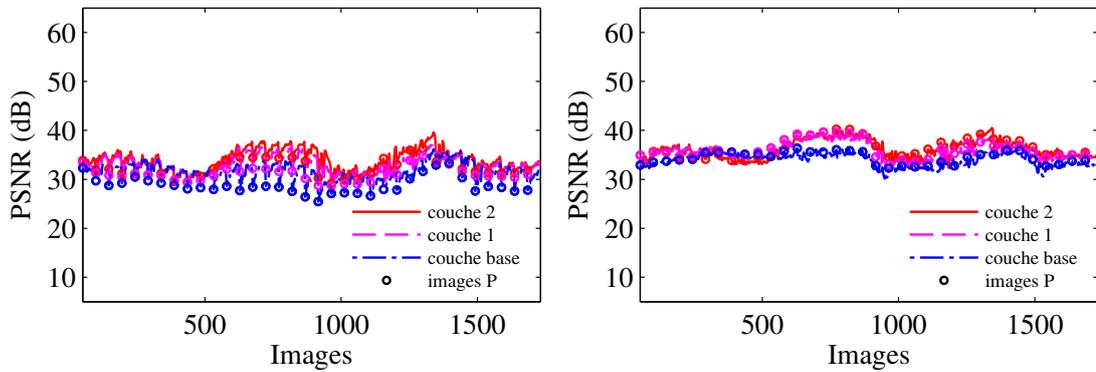
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.96	0.85	0.85
avec l'efficacité de codage	0.35	0.33	0.32

FIGURE 6.4 – Améliorations apportées sur la qualité pour le scénario **spatial** sur la séquence **SOCCER**.



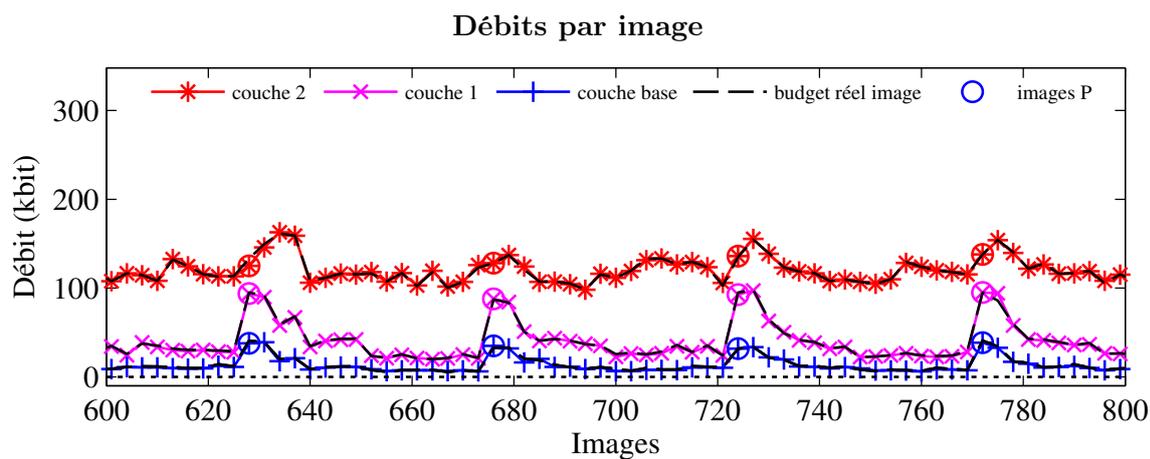
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



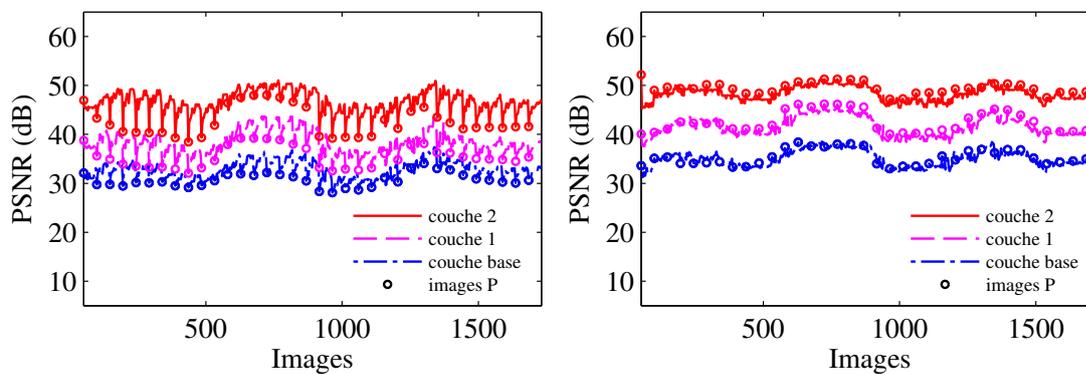
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	1.15	0.92	0.90
avec l'efficacité de codage	0.44	0.47	0.55

FIGURE 6.5 – Améliorations apportées sur la qualité pour le scénario **qualité** sur la séquence **SOCCER**.



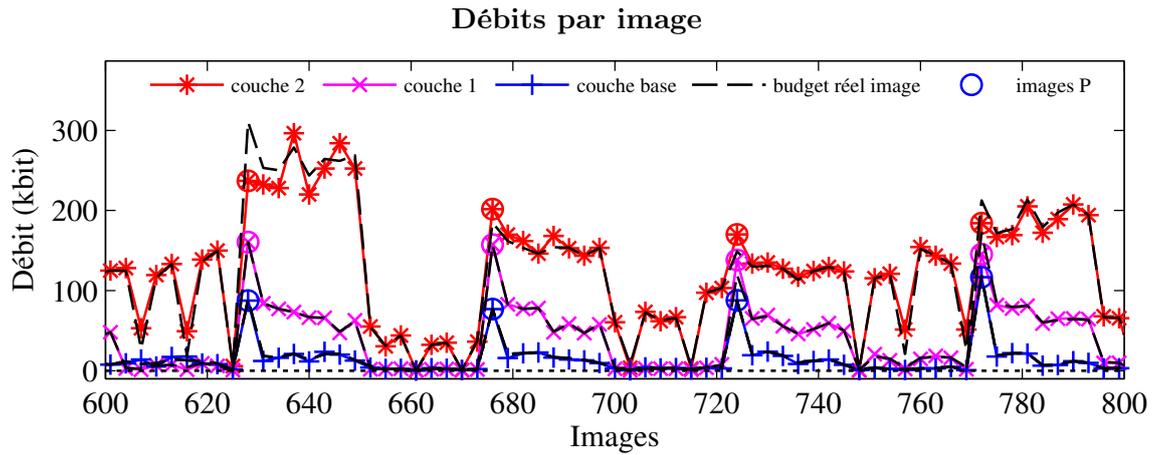
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



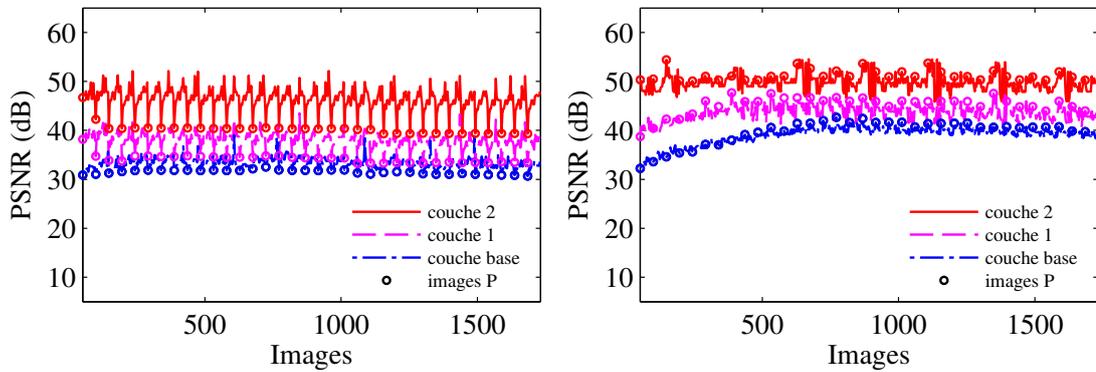
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.98	1.28	1.54
avec l'efficacité de codage	0.49	0.60	0.61

FIGURE 6.6 – Améliorations apportées sur la qualité pour le scénario **qualité** sur la séquence **CITY**.



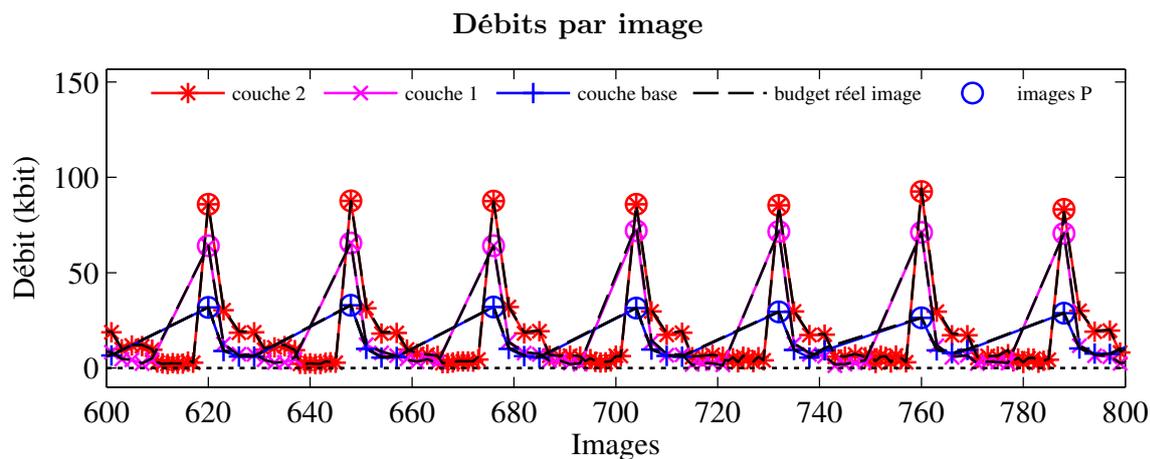
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



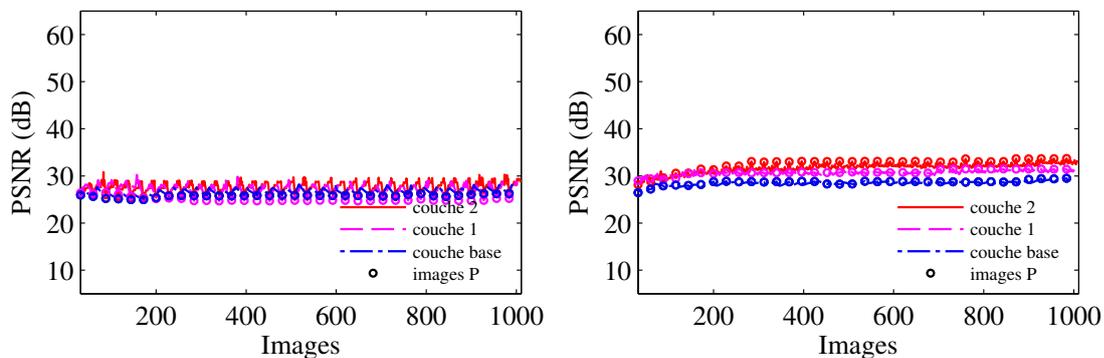
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.88	1.42	2.16
avec l'efficacité de codage	0.56	1.12	1.33

FIGURE 6.7 – Améliorations apportées sur la qualité pour le scénario **temporel** sur la séquence **HARBOUR**.



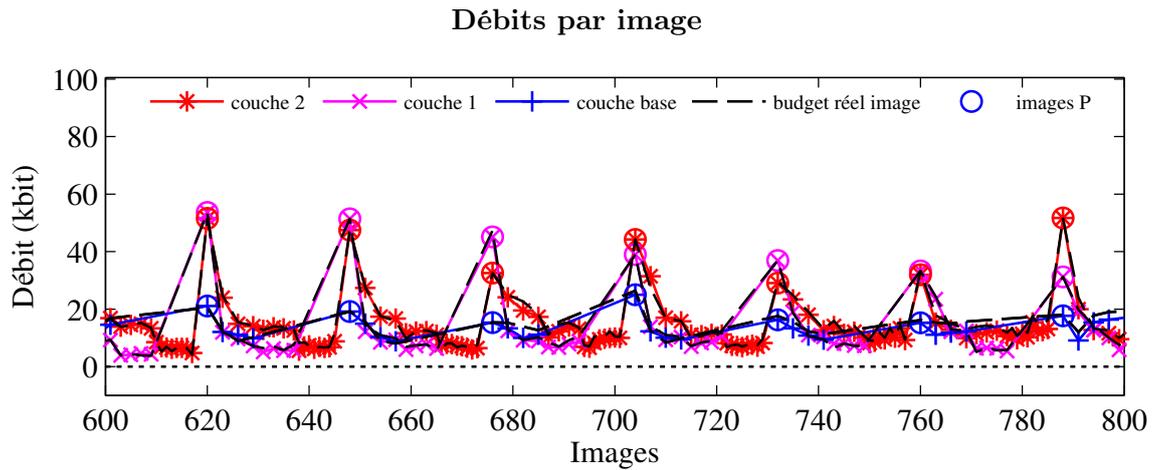
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



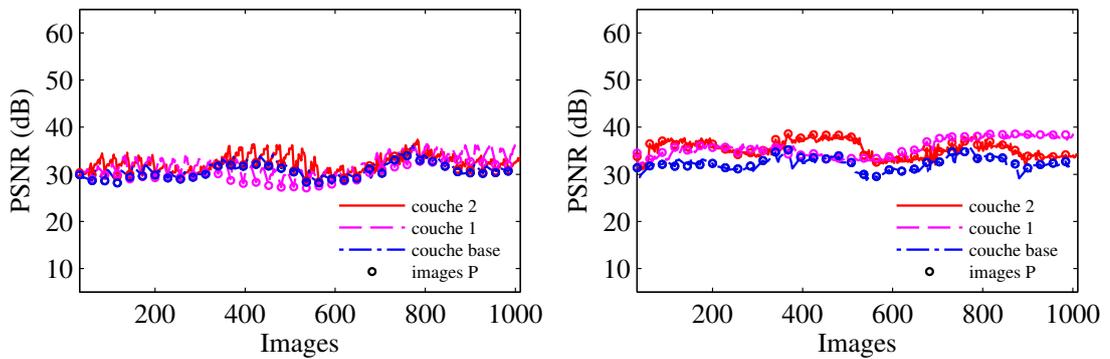
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.64	1.04	0.85
avec l'efficacité de codage	0.21	0.21	0.38

FIGURE 6.8 – Améliorations apportées sur la qualité pour le scénario **temporel** sur la séquence **SOCCER**.



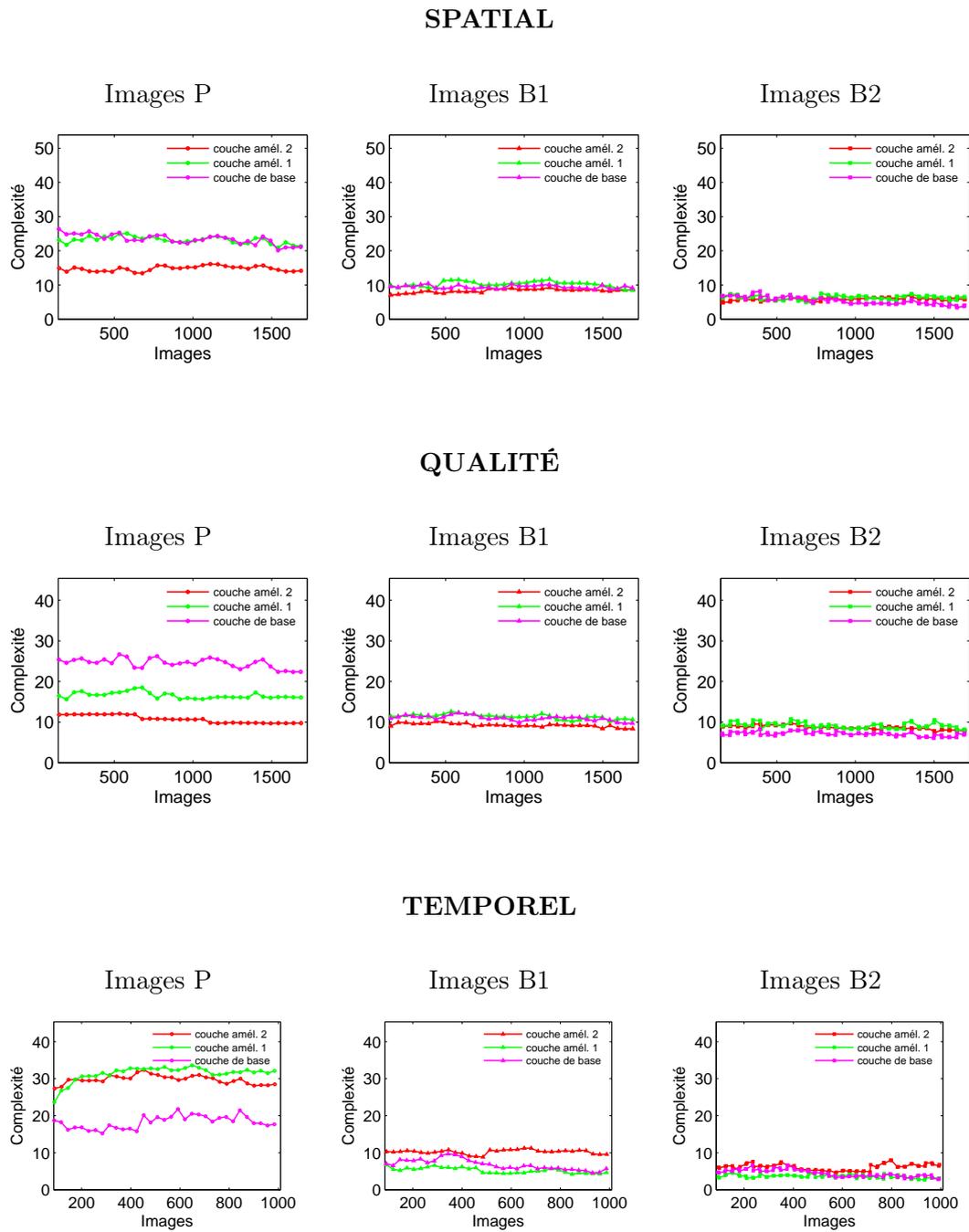
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.67	1.12	0.95
avec l'efficacité de codage	0.51	0.31	0.49

FIGURE 6.9 – Influence de la prédiction inter-couches sur l'efficacité de codage des images pour la séquence HARBOUR.



Chapitre 7

Contrôle de débit en une passe pour MPEG-4 SVC

Le processus d'encodage de MPEG-4 SVC a en lui même une complexité élevée en termes de calculs. En effet, le processus de prédiction spatio-temporelle requiert un nombre important d'opérations. Par rapport à un encodage indépendant des couches, les mécanismes de prédiction inter-couches augmentent considérablement la quantité de calculs effectués pour trouver la meilleure prédiction pour un macrobloc dans les couches d'amélioration. En comparaison avec MPEG-4 AVC, la durée d'un encodage peut devenir prohibitive dès lors que plusieurs couches sont impliquées. Il est donc d'autant plus primordial de prendre en compte la complexité calculatoire dans l'élaboration d'une méthode de contrôle de débit.

Le mécanisme présenté dans le chapitre 6 effectue deux passes d'encodage sur chaque image. La première passe est utilisée pour initialiser le modèle de débit, et la seconde passe encode l'image avec la valeur de QP choisie. Comme nous l'avons montré dans le Tableau 5.2, cela se traduit par une forte augmentation du temps d'encodage. Cette initialisation du modèle de débit sur les données de l'image elle même permet d'atteindre une grande précision dans le contrôle effectué, mais elle pénalise fortement la rapidité du processus, en doublant le temps d'encodage.

Pour être précis, il est important que le modèle de débit soit initialisé avec des données proches de celles de l'image à encoder. Or, il existe de fortes dépendances dans les contenus de deux images proches (exception faite des cas où l'activité change brusquement, comme lors des coupures entre deux plans dans une séquence). Si leur contenu est proche, une des images peut représenter une base statistique fiable pour effectuer

la régulation de débit sur l'autre image. L'idée développée dans ce chapitre est par conséquent d'utiliser les informations obtenues après l'encodage des images précédentes comme base pour le contrôle de débit sur l'image en cours. La première passe d'encodage est ainsi supprimée, et peut être remplacée par une phase d'initialisation beaucoup moins coûteuse.

On peut s'attendre à observer une diminution de la précision de la régulation, car les changements de contenu entre les images, aussi proches soient-elles, introduisent un biais dans le calcul du QP optimal. Néanmoins, l'impact du mécanisme de régulation de débit sur le temps d'encodage est fortement réduit. Le mécanisme proposé dans ce chapitre représente donc un compromis entre performances et complexité. L'image précédente est ainsi utilisée comme base pour le modèle de débit de l'image en cours. Nous verrons que le choix de cette image est déterminant dans l'efficacité du contrôle, et qu'il peut poser problème dans les cas où les changements d'activité sont trop importants.

7.1 Initialisation du modèle de débit

Le modèle de débit nécessite une phase d'initialisation, assurée par l'utilisation d'une première passe d'encodage. Il doit en effet s'adapter aux caractéristiques de chaque image pour être en mesure de prédire le comportement du débit en fonction du paramètre de quantification. Concrètement, le modèle a besoin de connaître la pente de la droite symbolisant la relation entre ρ et le débit, qui change pour chaque image (équation (7.1)). Il a également besoin de connaître l'image des coefficients transformés, sur laquelle se base la relation entre ρ et le QP (équation (7.2)).

Comme nous l'avons vu au chapitre 5, le modèle de débit formule la relation entre le débit et le paramètre de quantification de la manière suivante :

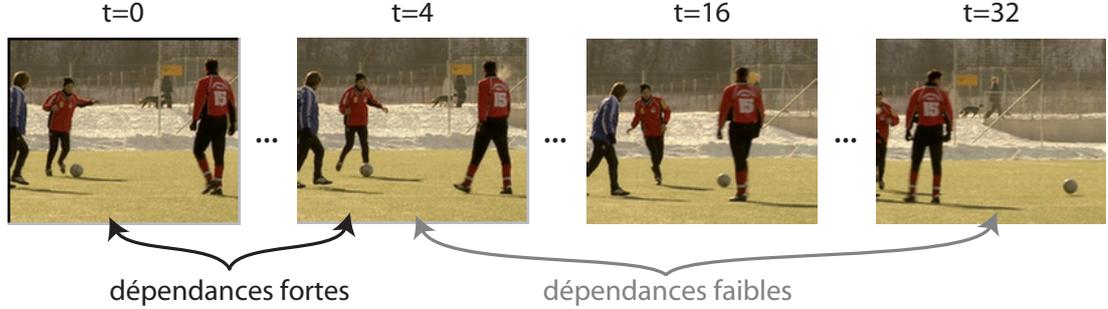
$$R(q) = R(\rho(q)) = \frac{R_0 (1 - \rho(q))}{1 - \rho_0}, \quad (7.1)$$

avec

$$\rho(q) = \sum_{m \in f} \left(\frac{1}{M} \sum_{(i,j) \in m} z(c_{ij}^m, i, j, q) \right). \quad (7.2)$$

Pour fonctionner, le modèle requiert les valeurs de ρ_0 et R_0 , ainsi que les coefficients transformés c_{ij}^m de l'image. Ces paramètres dépendent de la valeur de QP utilisé, malgré le fait que la régulation de débit prenne place avant le choix du QP. Par conséquent, ces paramètres ne sont pas disponibles au moment du contrôle de débit. Dans les chapitres 5 et 6, les valeurs de ρ_0 , R_0 et les coefficients transformés sont donc obtenus en encodant

FIGURE 7.1 – Dépendances entre les contenus d’images proches.



l’image une première fois. Cette approche n’étant pas optimale du point de vue de la complexité de calculs, nous proposons dans ce chapitre de réutiliser ces paramètres à partir de l’image précédemment encodée.

Ce choix est justifié par le fait que l’on observe des corrélations fortes entre des images successives (Figure 7.1). De fait, la prédiction inter-images utilise ces redondances, en codant les blocs d’une image par rapport aux blocs d’une image de référence. Ces corrélations impliquent que le contenu des images et leur comportement en termes de débit et de qualité sont relativement proches. Il est donc possible d’utiliser les informations d’une image pour prédire le comportement d’une image proche.

Ainsi, les valeurs initiales ρ_0 et R_0 peuvent être simplement remplacées par les valeurs calculées sur une image précédente :

$$R(q) = \frac{R_p (1 - \rho(q))}{1 - \rho_p}, \quad (7.3)$$

où p est l’image précédente choisie comme référence pour la régulation, ρ_p est la valeur de ρ mesurée après l’encodage de l’image p et R_p est le débit obtenu sur cette image.

De manière analogue, l’équation (7.2) devient :

$$\rho(q) = \sum_{m \in p} \left(\frac{1}{M} \sum_{(i,j) \in m} z(c_{ij}^m, i, j, q) \right). \quad (7.4)$$

Le choix de l’image p utilisée comme référence pour le modèle de débit est primordial. En effet, plus les corrélations avec l’image en cours sont fortes, plus le contrôle de débit peut être précis. La section suivante décrit la méthode que nous avons adoptée pour le choix de cette image de référence.

7.2 Choix de l'image de référence pour le modèle de débit

Il est important que l'image utilisée comme base pour le contrôle de débit ait des caractéristiques très proches de l'image en cours pour assurer une régulation fiable. Or, dans le processus d'encodage de MPEG-4 SVC, deux images consécutives dans l'ordre d'affichage ne sont pas toujours encodées l'une après l'autre (*cf.* Figure 1.3). Par conséquent, l'image immédiatement précédente dans l'ordre de lecture n'est pas nécessairement du même type que l'image courante. Dans le chapitre 6, nous avons vu que chaque type d'image (et chaque niveau d'images B hiérarchiques) a un comportement propre en termes de débit et de qualité. Ces différences de comportement ne permettent pas d'utiliser une image d'un type différent pour initialiser le modèle de débit.

L'image qui possède le comportement le plus proche de celui de l'image en cours de régulation est la dernière image encodée ayant le même type, le même niveau temporel et située dans la même couche. C'est cette image que nous utilisons pour obtenir les valeurs initiales pour les paramètres du modèle de débit. Ce choix laisse d'ores et déjà apparaître quelques problèmes éventuels. Pour certains types d'images, deux images du même type peuvent être relativement éloignées dans le temps, et les dépendances dans leurs contenus par conséquent peu élevées (*e. g.* : deux images P sont espacées de la taille d'un GOP, deux images B du premier niveau temporel sont espacées de la moitié de la taille du GOP). Nous discuterons de la perte d'efficacité causée par la distance entre deux images de même type dans la section présentant les résultats expérimentaux. La section suivante présente la méthode de calcul du budget des images dans le mécanisme en une passe.

7.3 Calcul du budget au niveau image

Le budget de l'image est calculé de manière similaire à celle présentée au chapitre 6. Néanmoins, l'absence de première passe implique quelques changements dans l'évaluation de l'efficacité de codage de l'image. En effet, la valeur de K_i pour l'image en cours est désormais estimée simplement à partir de celle de l'image précédente :

$$K_i = 2^{q_{p_i}/6} \times R_{p_i}, \quad (7.5)$$

où p_i désigne la dernière image encodée de même type et de même niveau temporel que l'image en cours. Le budget alloué à l'image est ensuite calculé en utilisant l'équation (6.10). Cette solution est volontairement simple de manière à minimiser les opérations nécessaires pour calculer le budget alloué à une image. Des améliorations peuvent y être apportées, mais ne font pas l'objet de ce chapitre. Le paragraphe suivant présente les

résultats obtenus expérimentalement pour évaluer les performances du mécanisme de régulation en une passe.

7.4 Résultats

Pour tester les performances du mécanisme de contrôle de débit en une passe, nous utilisons les scénarios présentés dans les chapitres précédents. Notre analyse s'appuie sur les temps d'encodage et les débits par image, par GOP et par seconde obtenus.

7.4.1 En termes d'erreur de débit

Les Figures 7.2 à 7.7 présentent les débits obtenus au niveau des images, des GOP et des secondes, ainsi que les erreurs δ_t et $\hat{\delta}_t$, calculées comme au chapitre 5 (des résultats complémentaires sont disponibles en annexe C).

Globalement, on constate une augmentation au niveau image de l'erreur entre le budget alloué et le débit obtenu. L'écart type augmente également, ce qui montre que les performances varient de manière plus importante. Ceci est une conséquence logique de l'utilisation des informations de l'image précédente, qui ne sont pas aussi fiables que les informations obtenues lors d'une première passe d'encodage sur la même image.

Cependant, pour les séquences comme HARBOUR, SOCCER et HOCKEY, dans lesquelles les changements locaux d'activité sont relativement faibles, l'écart avec les performances affichées par le mécanisme en deux passes est relativement peu élevé, malgré la simplicité de l'approche proposée. L'analyse des débits obtenus au niveau des images confirme un comportement similaire sur les deux approches. Sur les séquences comme CREW ou CITY, les changements d'activité entre images proches sont très importants (*cf.* Figure 3.1), et il est difficile d'estimer de manière efficace le comportement d'une image par rapport à l'image précédente. C'est pourquoi on observe des performances moins bonnes au niveau des images, et particulièrement sur les images B pour lesquelles le plus grand nombre de modes de prédiction pour les macroblocs provoque davantage d'imprécisions dans la régulation. Ce phénomène est amplifié à haut débit, comme le montrent les résultats pour le scénario en qualité (Figure C.6). La couche la plus haute étant encodée avec un débit élevé (*i. e.* 3600kbps pour des images de taille 352*288 à 30Hz), les valeurs de QP utilisées sont très faibles (*e. g.* inférieures à 15). Les erreurs dans le choix du QP ont par conséquent un impact beaucoup plus important sur le débit produit.

Les erreurs moyennes et leurs écarts types ne sont pas représentatifs pour ces deux séquences. Le budget alloué aux images B de haut niveau hiérarchique est en effet très

faible car le budget restant pour le GOP est faible, du fait des erreurs effectuées sur les images précédentes. L'erreur engendrée est donc plus élevée car elle est calculée relativement au budget alloué. Toutefois, comme le montre le Tableau 7.2, l'erreur exprimée de manière absolue en kilobits par seconde reste faible pour l'ensemble des séquences encodées.

Au niveau des GOP et du débit par seconde, la régulation reste néanmoins très précise. L'erreur entre le débit cible et le débit obtenu est très faible pour les trois scalabilités, grâce à la compensation des erreurs d'allocation au niveau des GOP. Pour des GOP de grande taille, nous avons observé une perte de précision du mécanisme en une passe sur les images P, et dans une moindre mesure sur les images B de niveau temporel bas. L'intervalle entre deux images P dans la même couche est alors élevé et les changements d'activité entre les deux images ne permettent pas de réguler le débit aussi efficacement sur la deuxième image. Il en va de même pour les images B de niveau temporel bas, qui ne sont pas beaucoup représentées dans le flux. Un moyen de remédier à ce problème est de réduire la taille des GOP, au prix d'une perte d'efficacité de codage du flux car il contient alors davantage d'images P.

7.4.2 En termes de qualité

On observe des variations du PSNR plus importantes qu'avec le mécanisme en deux passes, notamment pour le scénario en qualité (Figures C.5 et C.6). L'allocation de budget au niveau des images est effectuée en se basant sur l'efficacité de codage des images précédentes, dont les caractéristiques visuelles ne sont pas forcément aussi proches. La répartition du budget du GOP est donc potentiellement faite à partir d'informations en partie obsolètes, et la méthode présentée au chapitre 6 ne parvient pas dans tous les cas à obtenir une qualité constante.

Visuellement, la qualité demeure néanmoins globalement constante, et les variations du PSNR observées sur les figures sont difficilement perceptibles. Une adaptation de la méthode de répartition de débit est présentée dans le chapitre suivant, en tentant de corriger la mesure de l'efficacité de codage à partir des différences de PSNR mesurées entre les images d'un GOP.

7.4.3 En termes de temps d'encodage

Le Tableau 7.1 compare le temps moyen d'encodage par image pour les mécanismes de contrôle en une et en deux passes au temps d'encodage sans contrôle de débit. Alors que le mécanisme en deux passes provoque la multiplication du temps d'encodage no-

TABLE 7.1 – Comparaison du temps d’encodage moyen par image des mécanismes en une et deux passes face à l’encodage sans régulation de débit.

	SPATIAL	QUALITÉ	TEMPOREL
Sans régulation	1.19 s.	0.56 s.	0.37 s.
Régulation 1 passe	1.22 s. (+2%)	0.56 s. (+1%)	0.40 s. (+6%)
Régulation 2 passes	2.46 s. (+105%)	1.13 s. (+102%)	0.78 s. (+108%)

minal (sans régulation de débit) par deux, l’augmentation produite par le mécanisme en une passe est inférieure à 6% du temps total d’encodage. Le fait d’exploiter les dépendances inter-images permet par conséquent d’obtenir un mécanisme aux performances intéressantes, en n’influant que très légèrement sur la durée d’encodage.

Cette réduction de la complexité du mécanisme se fait au prix d’une relative baisse de la précision du contrôle de débit. Cependant, ses performances demeurent comparables et la contrainte de débit au niveau de la seconde est respectée avec une grande précision.

7.5 Conclusion

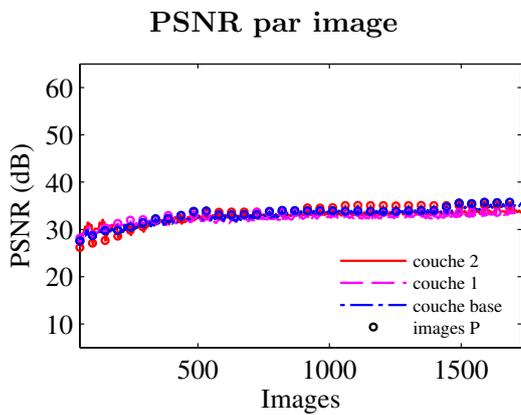
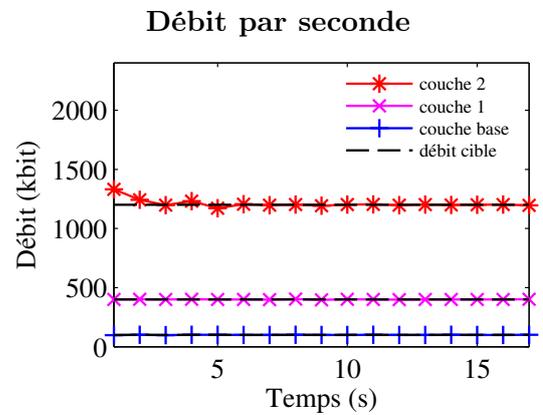
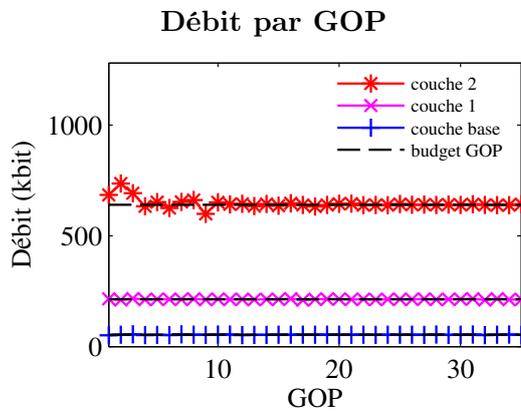
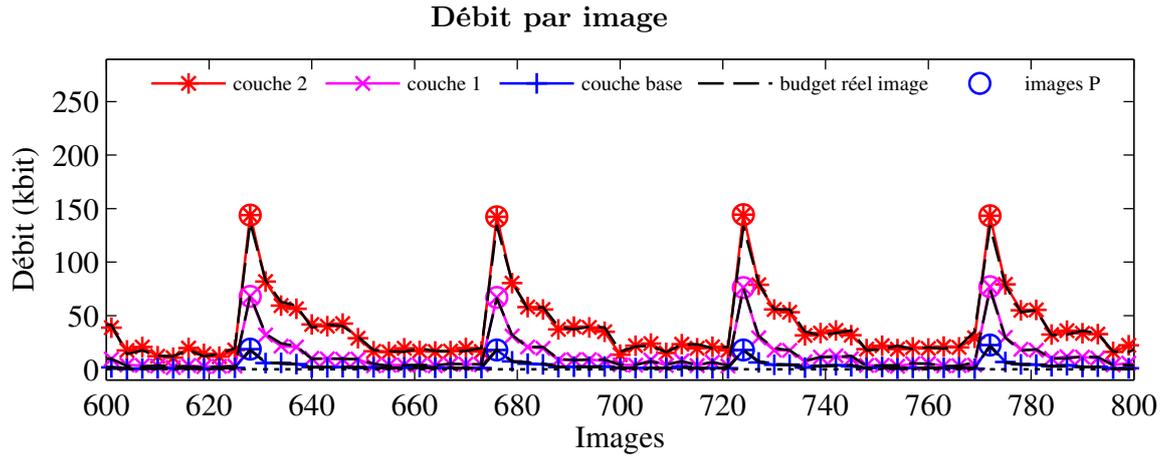
Dans ce chapitre nous avons proposé un mécanisme de régulation de débit en une passe pour MPEG-4 SVC. La réutilisation des informations obtenues sur la dernière image de même type nous a permis d’initialiser de manière très simple le modèle de débit pour l’image courante. Les résultats montrent que les performances du nouveau mécanisme restent relativement similaires à celles du mécanisme en deux passes. En revanche, le temps d’encodage est fortement réduit, et l’impact de la régulation n’excède pas 6% du temps d’encodage total. L’apport de ce chapitre est donc majeur, car il propose un mécanisme de contrôle de débit efficace, ayant un coût très faible en termes de complexité de calculs. Il a notamment donné lieu à la publication d’un article dans la conférence internationale *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, à Bilbao (Espagne), en Mai 2009 [62]. Un article de journal a également été soumis pour la revue *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)* en Avril 2009 [63].

Sur les séquences d’images complexes contenant de grandes variations de mouvement et de détails, les performances du mécanisme en une passe sont toutefois moins bonnes. La forte activité provoque en effet des changements importants sur des intervalles courts

d'images, et les dépendances inter-images ne sont pas assez élevées pour obtenir une bonne initialisation du modèle de débit et une régulation très précise. En particulier, les performances sur les images P, qui sont les moins représentées dans le flux encodé, souffrent de variations importantes. Le raccourcissement des groupes d'images permet de résoudre en partie ce problème, en réduisant l'intervalle entre deux images P.

Dans ce type de séquences contenant une grande activité, nous avons observé que les dépendances temporelles entre images d'une même couche peuvent être faibles, ce qui ne permet pas de les utiliser systématiquement pour initialiser le modèle de débit dans le mécanisme en une passe. En revanche, le mécanisme n'exploite pas les dépendances inter-couches. Deux images situées au même instant mais dans deux couches distinctes ont pourtant des caractéristiques proches. L'exploitation de ces similitudes pour anticiper les variations brusques de débit dans les couches d'amélioration à partir de leur couche de base sera donc abordé dans les perspectives.

FIGURE 7.2 – Performances du contrôle de débit en **une passe** pour le scénario **spatial** avec des GOP de 16 images sur la séquence **HARBOUR**.



Erreur de débit par image

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-1.80 %	17.84	-0.03 %	15.00
couche 1	-1.71 %	16.66	-0.25 %	14.16
couche 2	-3.33 %	16.70	-0.65 %	13.00

FIGURE 7.3 – Performances du contrôle de débit en **une passe** pour le scénario **spatial** avec des GOP de 16 images sur la séquence **SOCCER**.

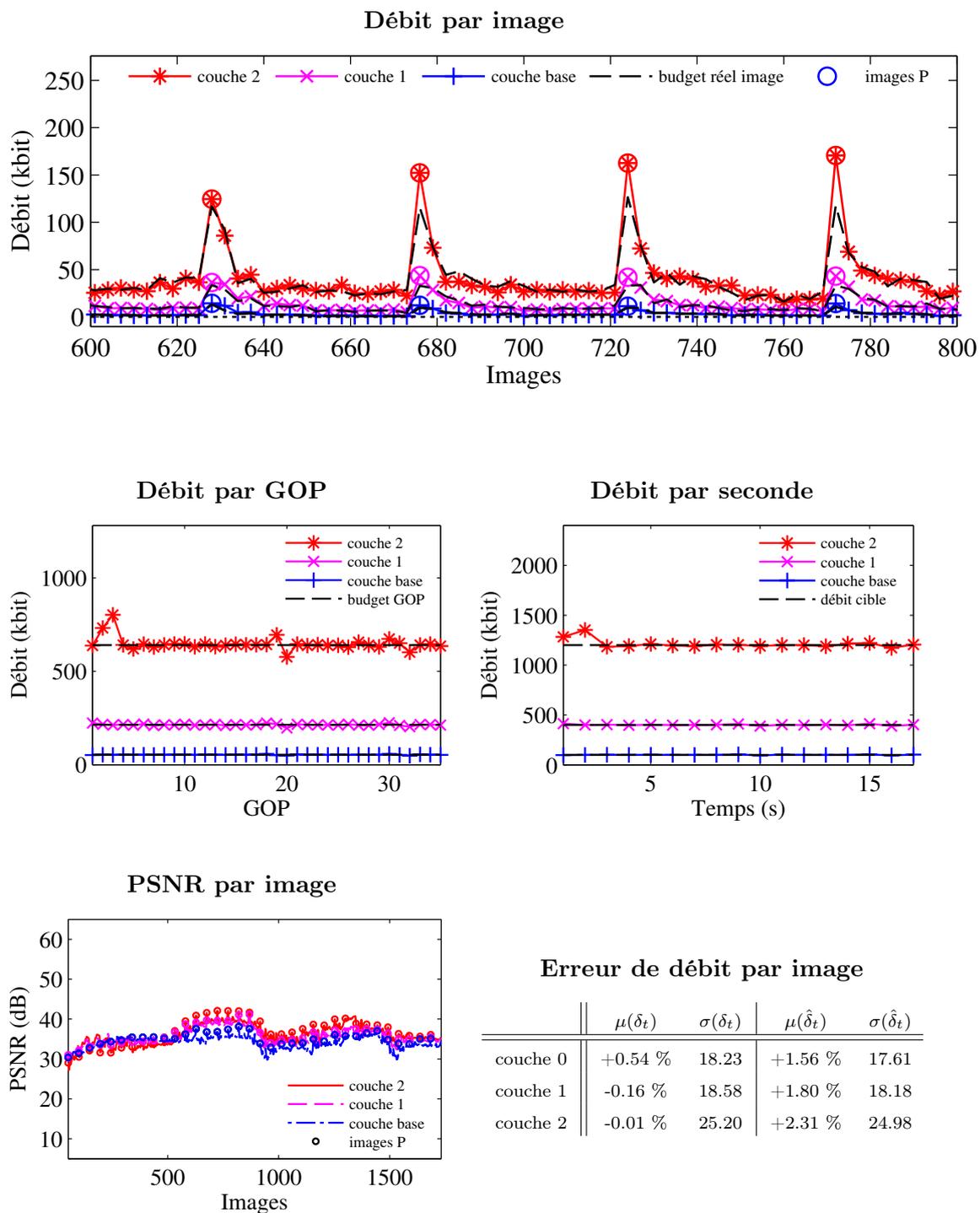
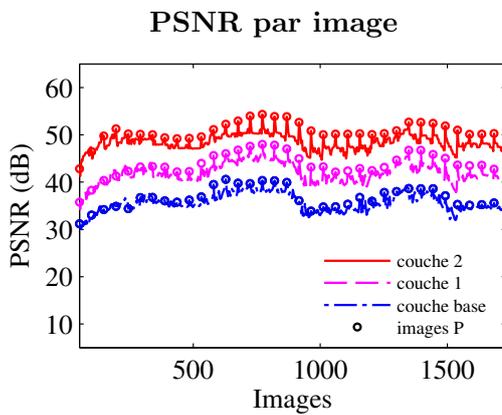
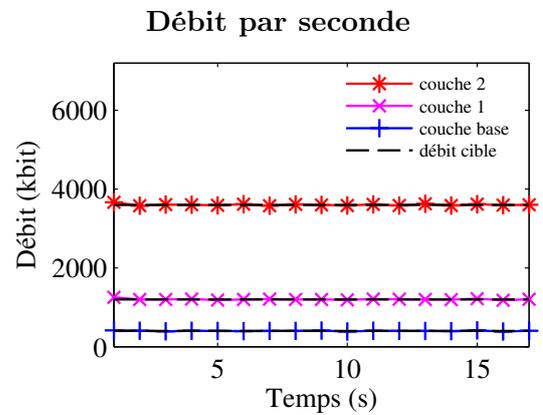
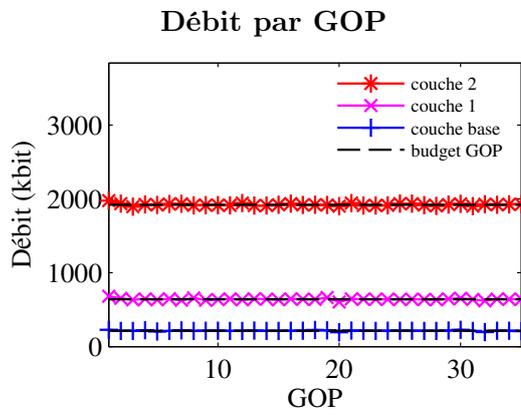
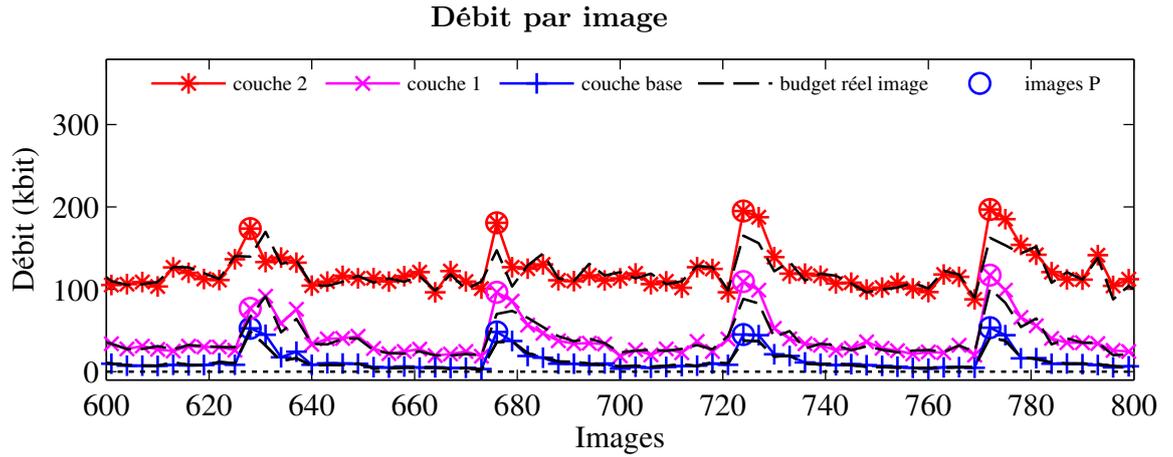


FIGURE 7.4 – Performances du contrôle de débit en **une passe** pour le scénario **qualité** avec des GOP de 16 images sur la séquence **SOCCER**.



Erreur de débit par image

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	+1.18 %	20.47	+2.76 %	20.15
couche 1	+0.70 %	20.90	+2.35 %	19.41
couche 2	+1.09 %	14.48	+1.80 %	13.03

FIGURE 7.5 – Performances du contrôle de débit en **une passe** pour le scénario **qualité** avec des GOP de 16 images sur la séquence **CREW**.

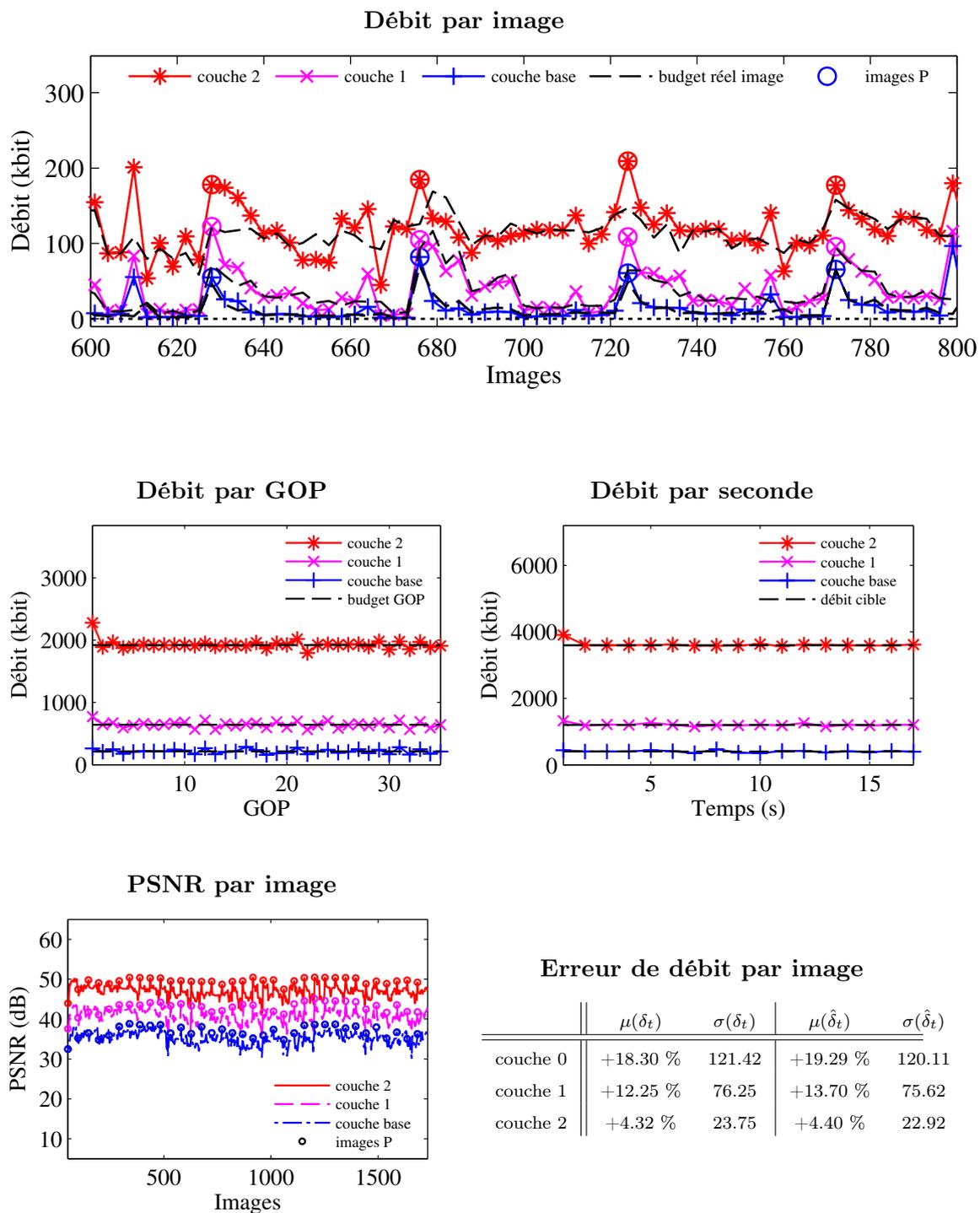


FIGURE 7.6 – Performances du contrôle de débit en **une passe** pour le scénario **temporel** avec des GOP de 16 images sur la séquence **HARBOUR**.

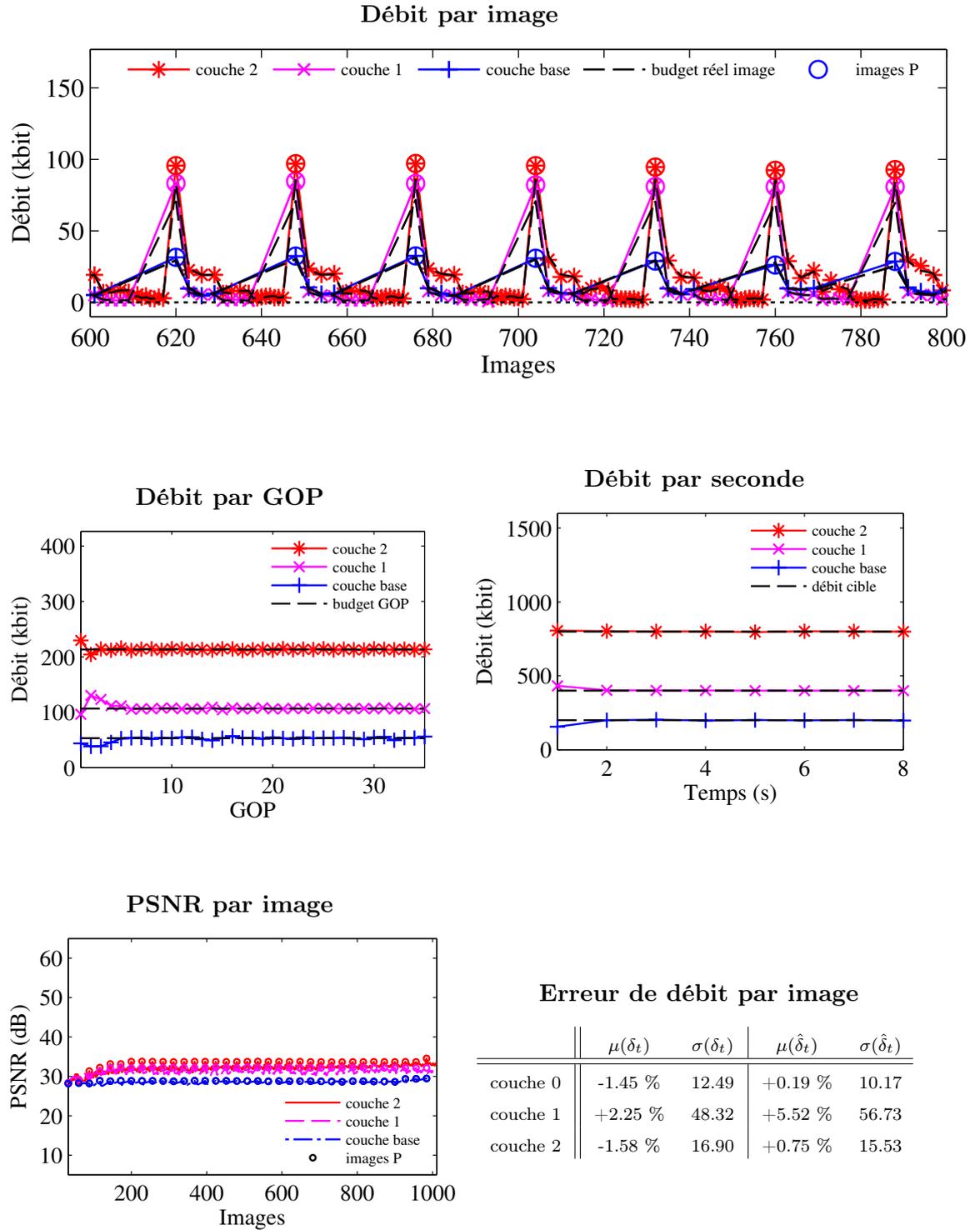
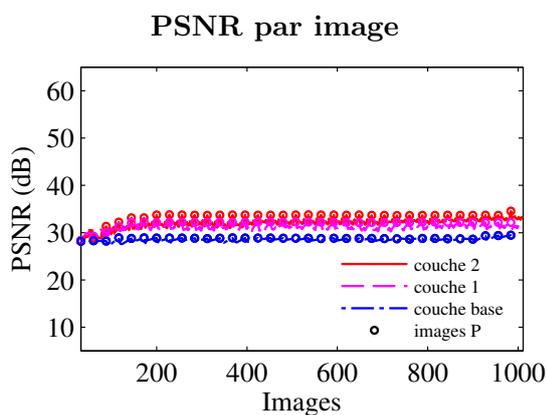
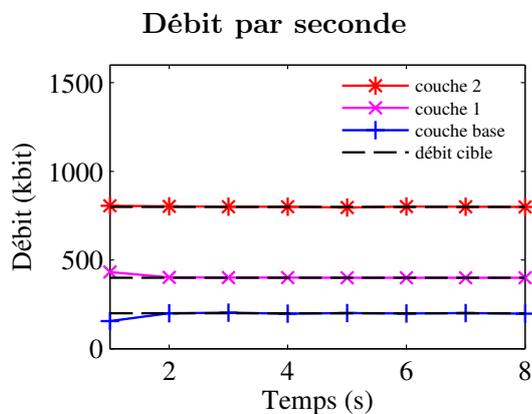
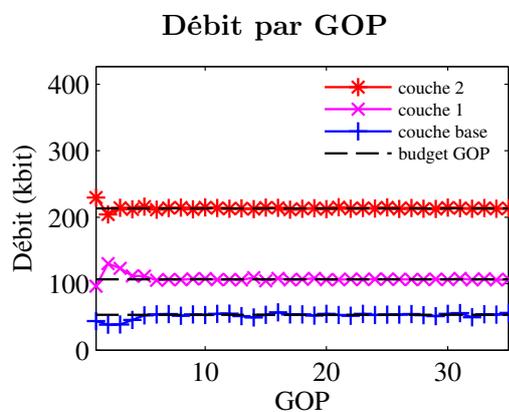
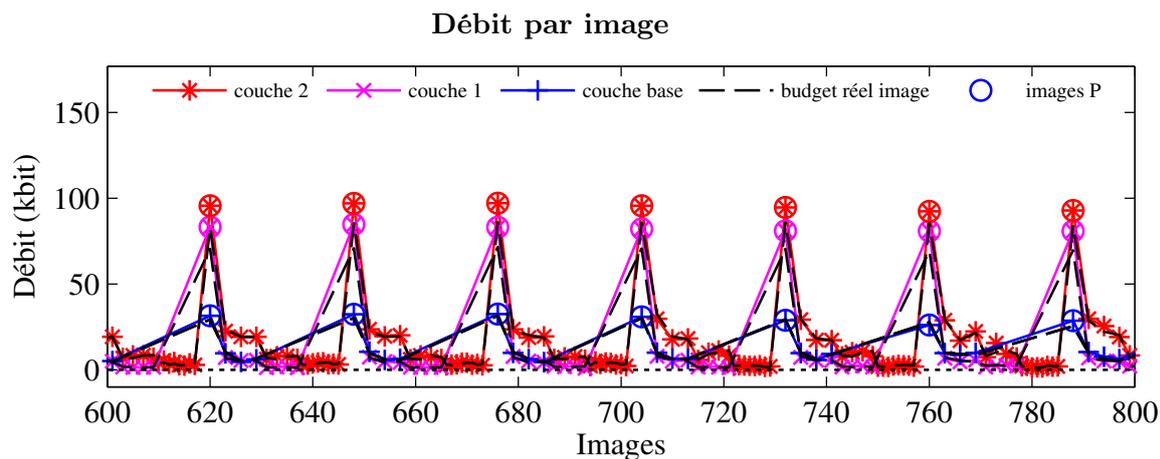


FIGURE 7.7 – Performances du contrôle de débit en **une passe** pour le scénario **temporel** avec des GOP de 16 images sur la séquence **SOCCER**.



Erreur de débit par image

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-1.45 %	12.49	+0.19 %	10.17
couche 1	+2.25 %	48.32	+5.52 %	56.73
couche 2	-1.58 %	16.90	+0.75 %	15.53

TABLE 7.2 – Erreur de débit avec le mécanisme en une passe **exprimée en kilobits**.

		SPATIAL			
		$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
CITY	couche 0	-0.15 kb	1.25	-0.09 kb	1.07
	couche 1	-0.45 kb	4.07	-0.25 kb	3.53
	couche 2	-1.64 kb	13.24	-0.79 kb	11.02
CREW	couche 0	+0.20 kb	2.49	+0.23 kb	2.51
	couche 1	+0.56 kb	7.98	+0.76 kb	7.97
	couche 2	+1.69 kb	20.49	+2.22 kb	20.23
HARBOUR	couche 0	-0.12 kb	0.67	-0.04 kb	0.53
	couche 1	-0.34 kb	1.98	-0.11 kb	1.48
	couche 2	-1.23 kb	6.42	-0.05 kb	4.92
HOCKEY	couche 0	-0.05 kb	0.98	-0.04 kb	0.97
	couche 1	-0.23 kb	3.61	-0.08 kb	3.61
	couche 2	-0.33 kb	12.85	+0.46 kb	11.82

		QUALITÉ			
		$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
CITY	couche 0	-0.21 kb	4.96	-0.08 kb	4.49
	couche 1	+0.03 kb	13.67	+0.15 kb	13.05
	couche 2	+0.25 kb	54.30	+0.52 kb	53.98
CREW	couche 0	+0.90 kb	9.37	+0.87 kb	9.34
	couche 1	+2.09 kb	16.35	+2.65 kb	16.19
	couche 2	+4.40 kb	23.68	+4.54 kb	22.90
HARBOUR	couche 0	-0.27 kb	1.92	+0.00 kb	1.59
	couche 1	-0.81 kb	4.93	-0.02 kb	4.14
	couche 2	-0.45 kb	8.45	-0.36 kb	5.53
SOCCER	couche 0	+0.47 kb	4.89	+0.61 kb	4.81
	couche 1	+0.75 kb	11.31	+1.40 kb	10.79
	couche 2	+1.30 kb	18.21	+2.12 kb	16.82

		TEMPOREL			
		$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
CITY	couche 0	-0.49 kb	2.84	-0.29 kb	2.51
	couche 1	+2.43 kb	6.95	+2.53 kb	6.90
	couche 2	-0.36 kb	2.92	-0.02 kb	2.46
CREW	couche 0	+0.76 kb	5.36	+1.14 kb	5.47
	couche 1	+2.02 kb	9.35	+2.05 kb	9.43
	couche 2	+0.85 kb	9.07	+0.97 kb	9.16
HOCKEY	couche 0	-0.11 kb	4.41	+0.04 kb	4.47
	couche 1	+0.61 kb	4.15	+0.70 kb	4.11
	couche 2	-0.01 kb	4.07	+0.16 kb	3.96
SOCCER	couche 0	+1.00 kb	4.29	+1.14 kb	4.32
	couche 1	+1.25 kb	5.67	+1.39 kb	5.51
	couche 2	+0.68 kb	4.91	+0.83 kb	4.81

Chapitre 8

Schémas avancés : travaux préliminaires

Notre approche de la régulation en une passe bénéficie de performances intéressantes, notamment du point de vue de la rapidité d'exécution. En revanche, cette complexité très faible est obtenue au prix d'une relative perte de précision, et la qualité perceptuelle est quelque peu altérée avec la réapparition de variations du PSNR dans les GOP.

Dans ce chapitre, nous présentons donc des travaux préliminaires en vue d'améliorer notre approche de régulation sur ces deux aspects. Une adaptation de la méthode de répartition du budget au niveau des images sera proposée dans un premier temps, suivie par une technique se basant sur les relations présentes entre les couches pour effectuer une régulation en une passe plus précise.

8.1 Poids d'images adaptatifs

La mesure de l'efficacité de codage des images présentée dans le chapitre 6 permet de réduire les variations de qualité à l'intérieur d'un groupe d'images de manière significative. L'effet de scintillement est fortement diminué et l'impression visuelle nettement améliorée. Cependant, des différences de qualité subsistent entre certaines images, notamment entre les images P et les images B hiérarchiques constituant le reste du GOP.

Dans cette section, des travaux préliminaires en vue d'une amélioration de la méthode de répartition de budget au niveau des GOP sont présentés. Une approche prenant en compte l'écart de qualité observé entre les images du GOP après encodage permet d'adapter la répartition aux changements dans les données. Le rapport entre le budget alloué aux images des différents types est ainsi corrigé au fur et à mesure pour obtenir une qualité plus constante sur le GOP.

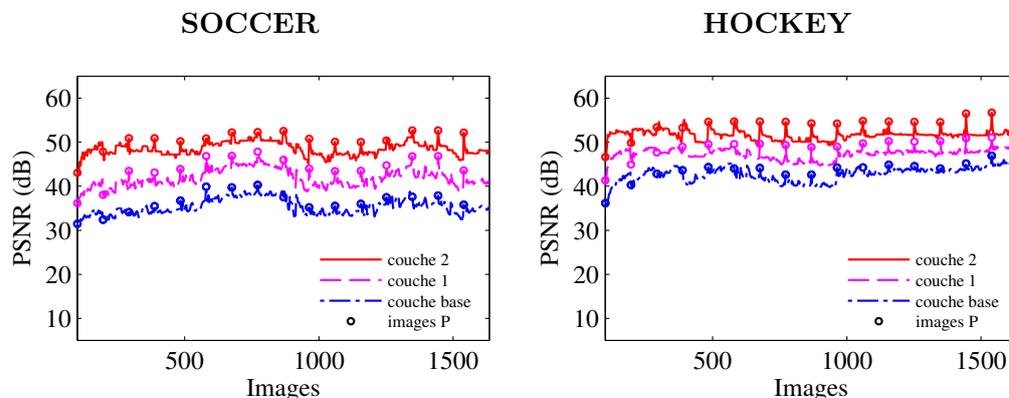


FIGURE 8.1 – Variations de la qualité observées dans le scénario en qualité avec la méthode de répartition de budget présentée dans le chapitre 6.

8.1.1 Limites de la mesure d'efficacité de codage

L'idée développée au chapitre 6 consiste à allouer à une image une portion du budget de son GOP relative à son efficacité de codage par rapport aux autres images du GOP. Ainsi les différences d'efficacité de codage entre les images P et les niveaux d'images B hiérarchiques sont compensées, ce qui permet d'atténuer de manière efficace les variations de qualité à l'intérieur d'un GOP.

Cependant, cette mesure n'est pas remise en question au cours de l'encodage, et la répartition du budget du GOP est faite sans prendre en compte la qualité effective obtenue sur les différents types d'images au cours de l'encodage. Comme le montre la Figure 8.1, les écarts résiduels de PSNR sont particulièrement visibles pour le scénario en qualité, avec des GOP de grande taille. Ces écarts montrent les limites de notre mesure de l'efficacité de codage, qui ne suffit pas toujours à répartir le budget de manière équitable au sens de la qualité entre les différentes images.

L'idée développée dans ces travaux préliminaires est par conséquent de modifier la répartition du budget du GOP entre les images en prenant en compte les différences de qualité observées entre les images au cours de l'encodage. Le poids affecté à chaque image est adapté de cette manière afin de corriger *a posteriori* la répartition de budget et obtenir une qualité plus constante. La section suivante décrit comment les poids des types d'images sont modifiés.

8.1.2 Évolution du poids des types d'images

Les poids relatifs des différents types d'image employés pour répartir le budget du GOP sont modifiés pour converger vers une répartition du budget permettant d'avoir une qualité constante. Les poids des niveaux successifs d'images B hiérarchiques sont ajustés en fonction de l'écart de qualité observé avec le PSNR de l'image P située au début du GOP, utilisée comme référence.

On affecte un poids $p(T, l)$ à chaque type d'image $T \in \{P, B1, \dots, B8\}$ dans chaque couche de scalabilité l , qui détermine le budget qui sera alloué aux images de ce type. Soit une image f de type T_f dans une couche de scalabilité l . L'image P située au début du GOP qui contient l'image f est notée P_f . Après l'encodage de l'image f , l'écart de PSNR $\Delta(P_f, f) = PSNR(P_f) - PSNR(f)$ est utilisé pour faire évoluer le poids du type d'image T_f de la manière suivante :

$$p(T_f, l) = \left(2^{q_f/6} \times b_f\right) \cdot evol(\Delta(P_f, f)), \quad (8.1)$$

où $evol(\cdot)$ est une fonction d'évolution, explicitée dans la section suivante.

Le budget R_t alloué à l'image f est ensuite calculé de manière similaire au chapitre 6, selon

$$R_t = \frac{p(T_f, l)}{\sum_{i \in GOP} p(T_i, l)} \times G_l, \quad (8.2)$$

où G_l est le budget restant pour le GOP. Naturellement, le choix de la fonction d'évolution est déterminant pour l'efficacité de la méthode. Les points importants à prendre en compte à cet effet sont discutés dans la section suivante.

8.1.2.1 Choix de la fonction d'évolution des poids d'images

Le choix de la fonction d'évolution détermine la capacité de la méthode de répartition à s'adapter aux changements d'efficacité de codage entre les types d'images. La rapidité et la stabilité de la convergence représentent les deux principaux enjeux de ce choix. Elle doit respecter les trois critères suivants :

1. **Sens de variation** : le sens de variation est dicté par le signe de $\Delta(P_f, f)$. Si l'image f a un PSNR plus faible que l'image P_f , son poids sera augmenté, si f a un PSNR plus élevé que P_f , son poids sera diminué.
2. **Amplitude** : l'amplitude de la correction dépend de l'écart de qualité entre les deux images. Les différences importantes de PSNR seront fortement corrigées, les faibles différences seront peu affectées.

3. **Borne maximale** : afin d'éviter la divergence du système, l'amplitude de la correction sera limitée à une valeur V_{max} à définir.

Au regard de ces trois critères, nous définissons la fonction d'évolution pour l'image f dans la couche l de manière itérative, selon

$$evol(\Delta(P_f, f)) = evol(\Delta(P_{f_{-1}}, f_{-1})) + signe(\Delta(P_f, f)) \cdot \left(e^{\alpha|\Delta(P_f, f)|} - 1 \right), \quad (8.3)$$

où f_{-1} est l'image précédant l'image f , et α est une constante à définir pour respecter le troisième critère.

Pour contraindre davantage le comportement de la fonction d'évolution, la valeur de $\Delta(P_f, f)$ peut être limitée elle aussi à une valeur Δ_{max} , telle que

$$\Delta(P_f, f) = \max(\Delta(P_f, f), \Delta_{max}). \quad (8.4)$$

La constante α peut alors être déterminée pour que $evol(\Delta_{max}) = V_{max}$, soit

$$\alpha = \frac{\log(V_{max} + 1)}{\Delta_{max}}. \quad (8.5)$$

La Figure 8.2 illustre le comportement de la fonction d'évolution. Conformément à sa définition, ses valeurs sont bornées entre les points $(0, 0)$ et (Δ_{max}, V_{max}) .

Les valeurs de Δ_{max} et V_{max} doivent être déterminées empiriquement. Lors de tests préliminaires, nous avons constaté qu'en fixant la valeur de Δ_{max} à 6, le système converge de manière satisfaisante, car la différence maximale de qualité observée entre deux images ne dépasse pas cette valeur.

Le choix de la valeur de V_{max} demande davantage de réflexion. En effet, cette valeur dicte la vitesse d'évolution du poids des types d'images. Une valeur trop forte occasionnera donc une divergence du système, alors qu'une valeur trop faible pénalisera sa rapidité d'adaptation. Une première idée pour la définition de V_{max} est d'utiliser le poids de l'image en cours comme référence :

$$V_{max} = \beta * p(T_f), \quad (8.6)$$

où le facteur β est une constante inférieure à 1, dont la valeur peut être adaptée à la séquence encodée et à la scalabilité utilisée.

En outre, il peut être intéressant de moduler la valeur de V_{max} en fonction du niveau temporel des images, de manière à compenser les différentes vitesses d'évolution. Cet aspect est développé dans la section suivante.

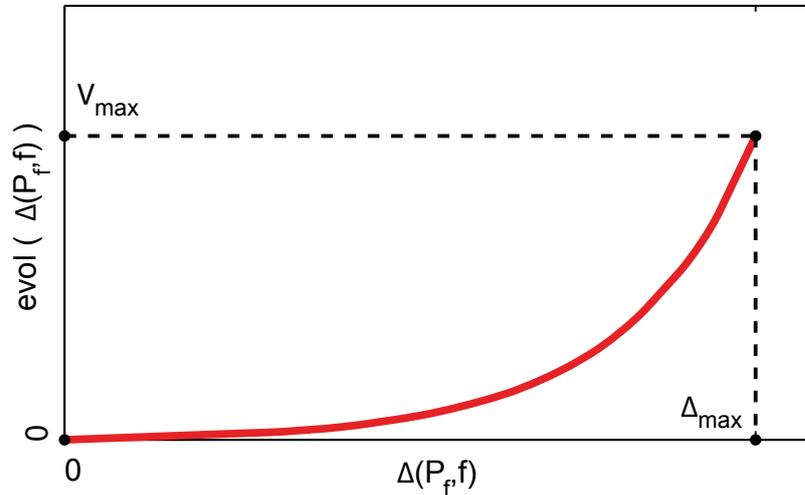


FIGURE 8.2 – Fonction d'évolution des poids d'image.

8.1.2.2 Adaptation de la vitesse d'évolution en fonction du niveau temporel

Du fait de la structure de GOP hiérarchique utilisée dans MPEG-4 SVC, la fréquence d'apparition des différents niveaux temporels d'images n'est pas la même (voir Figure 1.2). Étant donné que le poids d'un type d'images est attribué à chaque occurrence d'une image de ce type, le poids des images de niveau temporel élevé est mis à jour plus souvent que le poids des images de niveau temporel bas. Naturellement, le poids des images de niveau temporel élevé évoluera plus rapidement que celui des images de niveau temporel bas.

Pour compenser ce phénomène, il serait donc intéressant de faire varier la vitesse d'évolution en fonction du niveau temporel. Par exemple, on peut poser

$$V_{max} = p(T_f) * (0.8 - \frac{t_f}{10}), \quad (8.7)$$

où $t_f \in 1, \dots, 8$ est le niveau temporel de l'image f .

Ainsi, à différence de PSNR égale, le poids des images de niveau temporel bas évolue plus rapidement que celui des images de niveau élevé.

8.1.3 Conclusion

Dans cette section, nous avons introduit une méthode pour adapter la répartition du budget du GOP parmi les images aux différences de qualité observées au cours de l'encodage. Une fonction d'évolution a été définie, d'après l'écart de qualité entre une image et l'image P qui se situe au début du GOP. Les travaux sur cette méthode d'évolution sont encore en cours au moment de la rédaction de ce manuscrit, et feront l'objet de recherches futures.

Le choix de la fonction d'évolution est naturellement l'enjeu principal. Entre autres, différents points de vues sont envisagés, issus du domaine de l'Automatique et des régulateurs en *Proportionnel Intégral Dérivé* (PID), ou la caractérisation de la relation liant la différence de poids entre deux images à l'écart de qualité engendré.

8.2 Exploitation des dépendances inter-couches

Le mécanisme présenté au chapitre 7 exploite les dépendances temporelles entre images dans une même couche pour initialiser le modèle de débit. Les informations de la dernière image encodée du même type que l'image en cours sont utilisées, ce qui permet de s'affranchir de la première passe d'encodage. Ainsi, le mécanisme a une complexité calculatoire très faible et ne ralentit pas le processus d'encodage de manière significative.

Malheureusement, dans certaines configurations, les dépendances temporelles ne sont pas suffisamment fortes pour offrir une initialisation efficace. C'est notamment le cas lors de changements brutaux d'activité dans la séquence, causés par exemple par des flashes lumineux (*e. g.* la séquence CREW) ou des changements de scène. Cela est aussi vrai pour les GOP de grande taille, lorsque pour certains niveaux temporels les images sont relativement éloignées les unes des autres (*i. e.* images P et images B de bas niveau). La contrainte d'utiliser la dernière image de même type pour la régulation oblige alors le mécanisme à se baser sur une image dont les caractéristiques ne sont pas assez proches de l'image à encoder.

Dans ces deux cas, les dépendances entre l'image à encoder et la même image dans la couche de base sont donc vraisemblablement plus fortes que les dépendances temporelles. L'image de la couche de base peut alors représenter une base plus fiable pour la régulation. Dans cette section, nous étudions la faisabilité d'une telle approche où l'image précédente dans la même couche est remplacée par l'image de la couche de base pour initialiser le modèle de débit. À partir de cette étude préliminaire, des perspectives de recherche seront élaborées.

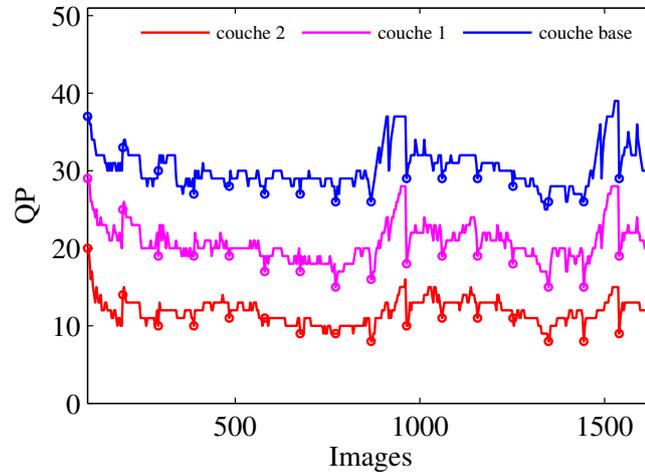


FIGURE 8.3 – Valeurs de QP choisies pour les trois couches du scénario en qualité avec la séquence SOCCER.

8.2.1 Comportements débit-distorsion similaires entre les couches

Comme le montre la Figure 8.3, le QP choisi par la régulation de débit varie globalement de manière équivalente sur l'ensemble des couches de scalabilité. Cette observation se généralise pour les trois scénarios de test et l'ensemble des séquences d'images utilisées. Le même type de contrainte étant appliqué aux trois couches, on peut donc entrevoir un comportement semblable en termes de débit et de distorsion. Cependant, les valeurs de QP choisies ne sont pas identiques pour autant. On observe plutôt un biais relativement constant entre les QP des différentes couches, qui proscrit une utilisation directe du QP choisi dans une couche pour la couche supérieure. On assiste ainsi à un problème de différence d'échelle entre les couches.

8.2.2 Problème d'échelle entre les couches

Dans des configurations scalables réalistes telles que celles étudiées tout au long de cette thèse, les caractéristiques des couches diffèrent par leur taille, leur fréquence temporelle ou le débit qui leur est alloué. Ces différences amènent naturellement des comportements singuliers de la régulation sur chacune des couches. Ainsi, la relation entre le débit et le paramètre de quantification change d'une couche à l'autre, et il n'est pas envisageable d'utiliser de manière directe les informations d'une couche pour effectuer la régulation sur une autre.

Une étape de mise à l'échelle s'impose alors, afin de faire le lien entre les couches.

Cette mise à l'échelle peut être rendue possible par la modélisation de la relation qui existe entre le QP des différentes couches. À l'aide de cette relation, on peut en effet convertir la valeur du QP choisie dans la couche de base en une valeur correspondante pour la couche d'amélioration.

De manière attendue, la relation entre les QP des différentes couches est relativement simple à représenter, comme le montre la Figure 8.4. Elle trace la valeur du QP dans une couche en fonction de celui de sa couche de base, pour les deux couches d'amélioration. On observe une relation fortement linéaire pour la plupart des configurations. Cette tendance est confirmée par les coefficients de corrélation élevés entre les QP des différentes couches, présentés dans le tableau 8.1. Par conséquent, la relation qui lie les QP de deux couches peut être approchée de manière relativement efficace par une méthode de régression linéaire. Ce processus sera éclairci dans la suite de cette section, qui décrit le nouveau mécanisme de régulation envisagé.

Le mécanisme intégrant les informations de la couche de base pour réguler le débit dans les couches d'amélioration diffère relativement peu du mécanisme en une passe du chapitre 7. La méthode de répartition de budget reste inchangée, car c'est au niveau du choix du QP qu'interviennent les modifications. Le processus de respect de la consigne au niveau image est décrit dans cette section.

8.2.3 Choix du QP

Le choix du QP dans la couche de plus basse résolution ne change naturellement pas. Les informations de l'image précédente de même type est utilisé comme base pour le calcul du QP de l'image en cours, de la même manière que dans le chapitre 7.

Dans les couches d'amélioration, le QP est calculé en fonction de celui de leur couche de base. Soit q_{base} la valeur de QP choisie dans la couche de base. La valeur de QP pour la couche d'amélioration, notée q_{amel} , est calculée de la manière suivante :

$$q_{amel} = A.q_{base} + B, \quad (8.8)$$

où A et B sont les paramètres de la droite de régression linéaire calculés par la méthode des moindres carrés [23]. Ces coefficients sont calculés à partir des valeurs observées des QP dans la couche de base et la couche d'amélioration pour les images précédentes.

De manière à rendre la régulation plus précise, il est intéressant d'exécuter une seconde fois le calcul du QP sur l'image de la couche de base. En effet, la valeur de q_{base} est calculée dans un premier temps à l'aide des informations de l'image précédente dans la couche de base, qui peut être relativement éloignée dans le temps. En effectuant à nouveau le choix du QP pour l'image de la couche de base après son encodage, on retrouve un fonctionnement équivalent à celui du mécanisme en deux passes présenté au chapitre 5. Cependant, contrairement au mécanisme en deux passes, l'image de la couche de base n'est pas ré-encodée, mais il s'agit plutôt la nouvelle valeur de q_{base} qui est simplement utilisée pour calculer la valeur de q_{amel} .

La régulation est par conséquent effectuée à deux reprises sur la couche de base, ce qui permet de corriger les erreurs de contrôle constatées après la première régulation. On peut donc naturellement espérer une atténuation des erreurs de régulation au fur et à mesure que l'on monte dans les couches d'amélioration.

8.2.4 Conclusion

Les travaux sur le mécanisme présenté dans ce chapitre sont toujours en cours au moment de la rédaction de ce manuscrit. L'efficacité de l'utilisation du QP de la couche de base dans les couches d'amélioration n'a pas encore été vérifiée de manière rigoureuse à l'heure actuelle. Cependant, les fortes corrélations entre les QP des différentes couches laissent présager de la précision de l'étape de régression linéaire, qui conditionne principalement la précision de la régulation. L'approche semble donc très prometteuse et les recherches dans cette direction seront poursuivies dans notre travail futur.

Pour la valider, plusieurs aspects des dépendances inter-couches devront être étudiés. Dans un premier temps, l'influence de la différence de résolution (spatiale, temporelle ou finesse de qualité) entre deux couches sur les corrélations entre leurs QP devra être éclaircie. On peut s'attendre à ce qu'elles diminuent lorsque l'écart est grand entre les couches, ce qui peut handicaper la régulation inter-couches.

Dans un second temps, il est envisageable de la combiner avec l'approche en une passe décrite au chapitre 7. En effet, dans une grande partie des cas, ce mécanisme est en effet performant, et des erreurs significatives n'apparaissent que dans les contextes où la régulation devient difficile. Un critère de choix entre les deux approches permettrait alors de décider quelle image utiliser comme base pour la régulation. Ce critère pourrait être basé sur l'observation de l'erreur de débit dans la couche de base, indiquant une baisse des dépendances temporelles.

FIGURE 8.4 – Relations entre les QP des différentes couches pour les scénarios spatial et en qualité.

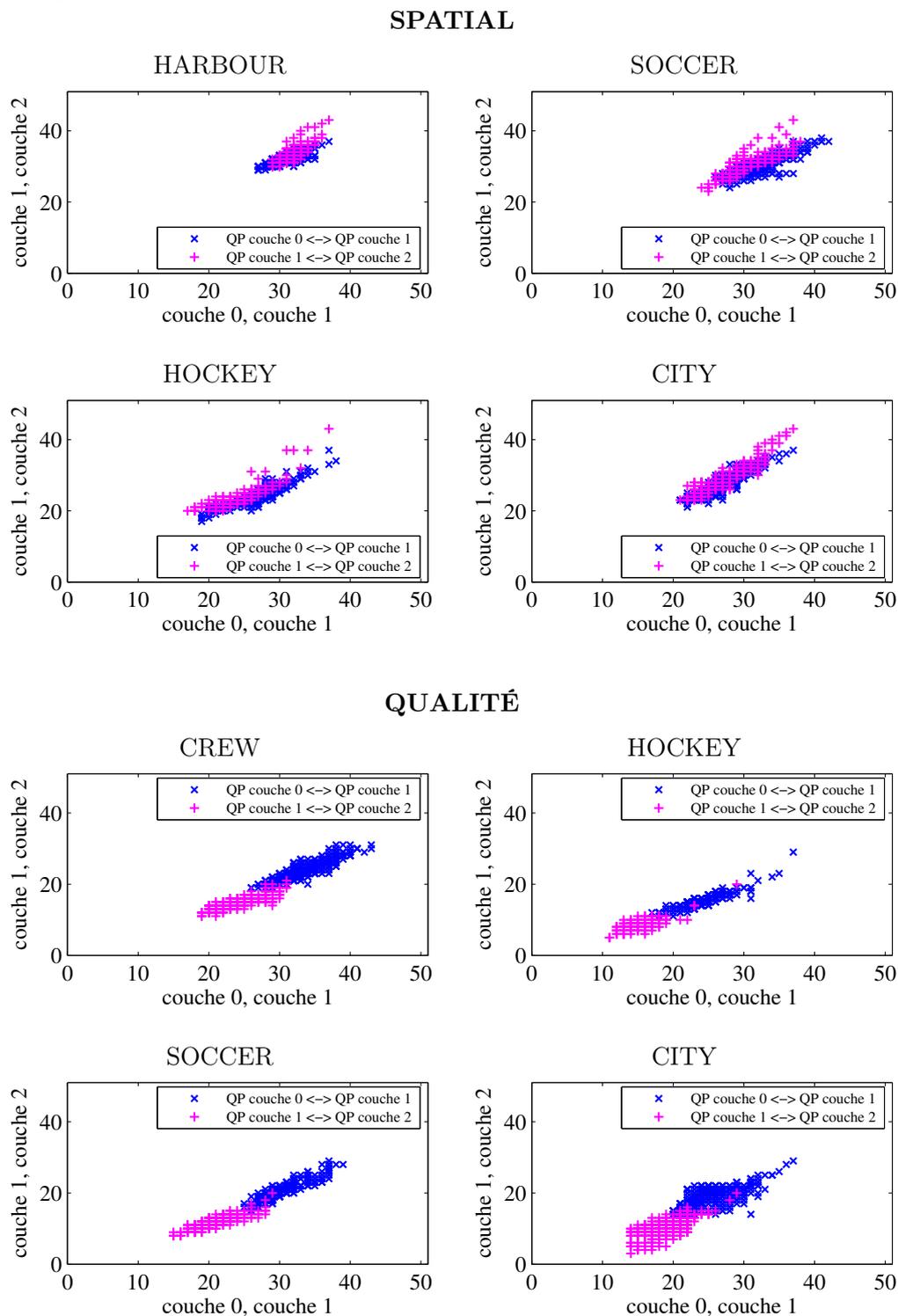


TABLE 8.1 – Corrélations entre les QP des différentes couches par type d'image, pour les scénarios spatial et en qualité.

		SPATIAL					
		P	B1	B2	B3	B4	B5
HARBOUR	<i>couches 0 et 1</i>	0.97	0.98	0.96	0.94	0.87	0.64
	<i>couches 1 et 2</i>	0.97	0.94	0.94	0.89	0.88	0.54
SOCCER	<i>couches 0 et 1</i>	0.87	0.80	0.80	0.81	0.89	0.82
	<i>couches 1 et 2</i>	0.93	0.93	0.91	0.75	0.75	0.90
HOCKEY	<i>couches 0 et 1</i>	0.94	0.95	0.95	0.95	0.95	0.88
	<i>couches 1 et 2</i>	0.96	0.97	0.92	0.92	0.81	0.87
CITY	<i>couches 0 et 1</i>	0.96	0.98	0.93	0.90	0.92	0.83
	<i>couches 1 et 2</i>	0.99	0.99	0.96	0.96	0.94	0.90
CREW	<i>couches 0 et 1</i>	0.93	0.96	0.97	0.96	0.90	0.92
	<i>couches 1 et 2</i>	0.80	0.75	0.83	0.84	0.92	0.90

		QUALITÉ					
		P	B1	B2	B3	B4	B5
HARBOUR	<i>couches 0 et 1</i>	0.97	0.94	0.95	0.93	0.87	0.76
	<i>couches 1 et 2</i>	0.97	0.96	0.96	0.94	0.85	0.70
SOCCER	<i>couches 0 et 1</i>	0.97	0.98	0.94	0.82	0.78	0.94
	<i>couches 1 et 2</i>	0.94	0.94	0.91	0.84	0.81	0.81
HOCKEY	<i>couches 0 et 1</i>	0.98	0.98	0.95	0.94	0.90	0.88
	<i>couches 1 et 2</i>	0.99	0.95	0.66	0.48	0.63	0.75
CITY	<i>couches 0 et 1</i>	0.81	0.89	0.84	0.80	0.84	0.28
	<i>couches 1 et 2</i>	0.86	0.88	0.67	0.72	0.66	0.58
CREW	<i>couches 0 et 1</i>	0.94	0.95	0.95	0.92	0.88	0.91
	<i>couches 1 et 2</i>	0.91	0.93	0.80	0.87	0.89	0.83

Conclusion

L'objectif de cette thèse consistait en l'élaboration de stratégies d'encodage permettant de faciliter le transport de flux vidéos scalables. Tout au long de ce manuscrit, cette problématique a été abordée par la construction de plusieurs techniques de régulation de débit pour le standard MPEG-4 SVC.

La première partie de notre travail a consisté à valider un modèle de débit pour le standard MPEG-4 SVC, à partir d'une approche originellement introduite pour la vidéo monocouche. Le modèle validé a par la suite permis de construire une stratégie simple de régulation en deux passes au niveau image pour les trois types de scalabilités proposés par le standard étudié. La qualité du flux a ensuite été intégrée à notre mécanisme de régulation, par le biais d'une méthode de répartition des ressources entre les différents types images. Suite à ce travail sur la qualité, les dépendances entre images successives dans une couche de scalabilité ont été exploitées pour développer une approche en une seule passe.

Parmi les travaux en cours actuellement, nous avons abordé une amélioration de la méthode de répartition du budget en la rendant plus adaptée aux variations de la qualité au cours de l'encodage. Nous avons également étudié la faisabilité d'une approche de régulation en une passe exploitant les dépendances inter-couches, à l'aide d'un mécanisme de mise à l'échelle entre les couches.

La suite de notre conclusion reprend chacun de ces aspects d'un point de vue critique. Les perspectives de recherche mises en évidence par cette analyse seront ensuite énumérées et argumentées.

1.1 Bilan sur le travail effectué

1.1.1 Validation du ρ -domaine sur MPEG-4 SVC

Le point de départ de notre travail a consisté en la validation d'un modèle de débit pour le standard MPEG-4 SVC, motivé par l'étude des différentes approches existantes pour le codage vidéo monocouche traditionnel. Le modèle retenu, appelé le ρ -domaine, bénéficie à la fois d'une grande précision et d'un coût très faible en termes de calculs. L'étude de son comportement sur un ensemble représentatif de configurations scalables nous a amené à en établir la validité dans le contexte de SVC et permis de construire une première approche de régulation de débit, décrite dans le paragraphe suivant.

1.1.2 Régulation de débit en deux passes pour MPEG-4 SVC

Nous avons proposé une première approche simple de la régulation de débit, se basant sur un pré-encodage de chaque image pour initialiser le modèle de débit. La grande précision de la régulation produite a été montrée sur les trois scalabilités fournies par la norme, avec une erreur de contrôle inférieure à 6% du budget alloué sur l'ensemble des configurations testées.

Il faut souligner l'efficacité de la méthode, malgré sa grande simplicité. En effet, le respect de la consigne de débit n'est assuré qu'à l'échelle de l'image, sans raffinement au niveau des macroblocs. L'intérêt d'une granularité plus fine paraît par conséquent assez limité, au vu des performances observées.

Un désavantage majeur de cette méthode est inhérent à la simplicité de la répartition du budget utilisée au niveau des groupes d'images. En effet, le budget disponible est réparti de manière uniforme entre les images, sans prendre en compte leurs différences de types. Un phénomène de scintillement de la qualité est donc introduit, provoquant une impression visuelle désagréable. Une méthode de répartition plus élaborée a par la suite été développée, comme le rappelle le prochain paragraphe.

1.1.3 Régulation de débit et qualité constante

Du fait des différents types de prédiction spatio-temporelle et de la structure de groupes d'images (GOP) hiérarchiques utilisés dans MPEG-4 SVC, les images n'ont pas toutes les mêmes besoins en termes de débit pour atteindre une qualité donnée.

Une analyse des processus de prédiction et de quantification nous a donc conduit à proposer une mesure de l'efficacité de codage pour chaque image. Cette mesure est ensuite mise à profit afin d'allouer à chaque image le budget nécessaire pour atteindre une qualité égale à celle des autres images du GOP.

Cette méthode de répartition permet de réduire significativement les variations de la qualité au cours des GOP, ce qui offre une impression visuelle améliorée de façon très nette. On constate également une légère augmentation de la qualité moyenne de l'encodage, car les ressources disponibles sont distribuées de manière à mieux exploiter la structure de GOP hiérarchique. Dans certains cas, quelques variations de la qualité subsistent cependant, dû au manque de souplesse de la mesure d'efficacité de codage, qui n'est pas remise en question au cours de l'encodage pour s'adapter aux particularités de chaque séquence. Une amélioration de cette technique a été abordée dans les perspectives de travail, et sera décrite dans la deuxième section de notre conclusion.

Hormis ces variations résiduelles de la qualité, l'inconvénient majeur du mécanisme proposé est la nécessité de pré-encoder chaque image pour initialiser le modèle de débit. Nous avons par conséquent élaboré un mécanisme de régulation en une seule passe.

1.1.4 Régulation de débit en une passe pour MPEG-4 SVC

Dans la dernière partie de nos recherches, nous avons proposé un mécanisme de régulation basé sur l'exploitation des dépendances entre images d'une même couche, de manière à éliminer la phase de pré-encodage et obtenir un mécanisme à faible complexité algorithmique.

La dernière image encodée représente une base relativement fidèle pour l'initialisation du modèle de débit de l'image en cours d'encodage. Les informations jusqu'alors obtenues pas une phase de pré-encodage sont donc remplacées par celles de l'image précédente. Du fait des différences de comportement entre les types d'images évoquées plus tôt, le choix de l'image de référence pour la régulation s'est porté sur la dernière image ayant le même type et niveau temporel que l'image en cours.

Le mécanisme qui en résulte offre une régulation ayant une complexité algorithmique très faible. L'exécution de l'intégralité de la régulation ne représente en effet pas plus de 8% du temps d'encodage. Ses performances en termes de précision de contrôle sont naturellement moins élevées que celles des mécanismes en deux passes présentés dans les paragraphes précédents. On constate notamment des erreurs relativement élevées pour des séquences contenant de fortes variations d'activité, ainsi que pour des GOP de grande taille. Dans ces deux cas, les faibles dépendances temporelles entre images d'une même couche ne permettent pas d'avoir une régulation très précise. Cependant, dans la majorité des configurations testées, l'erreur de contrôle demeure dans des valeurs comparables aux approches en deux passes, ce qui confère à ce mécanisme une bonne attractivité, particulièrement dans les contextes applicatifs requérant une grande rapidité d'exécution.

Une approche permettant d'améliorer la précision de la régulation de ce mécanisme a été évoquée dans nos perspectives de travail. Elle sera abordée dans la section suivante.

Un aspect important de notre travail est le fait que les mécanismes développés pour la vidéo scalable sont également compatibles avec le standard de codage monocouche MPEG-4 AVC/H.264. La prise en compte des aspects inter-couches est en effet optionnelle, et l'ensemble des approches est applicable à un encodage avec une couche unique.

1.2 Travaux en cours et perspectives de recherche

Certains aspects de la régulation restent à aborder pour proposer une approche complète sur la vidéo scalable. Comme le lecteur pourra le remarquer, ces perspectives sont nombreuses. Loin de traduire un travail inachevé, elles montrent le potentiel de recherche que représente la problématique de régulation de débit pour la vidéo scalable. Cette section passe en revue chacun de ces aspects, en commençant par les travaux en cours, pour finir sur les pistes de recherches envisagées.

1.2.1 Qualité constante et poids adaptatifs

La méthode de répartition utilisée pour réduire les variations de la qualité souffre à l'heure actuelle d'un manque de souplesse car elle ne prend pas en compte les résultats obtenus au cours de l'encodage. Nous avons évoqué dans le chapitre 8.1 l'ajout d'un terme d'évolution du poids affecté à chaque image, en fonction de l'écart de qualité observé avec les images voisines. L'utilisation de ce terme d'évolution doit permettre d'adapter la répartition du budget du GOP au contenu des images à coder et d'obtenir une qualité réellement constante au cours de l'encodage.

En outre, l'étude de mesures de qualité perceptuelles adaptées au codage vidéo est envisagée pour remplacer le PSNR, qui malgré son utilisation majoritaire manque de refléter l'impression visuelle ressentie par le spectateur, à qui s'adresse avant tout l'ensemble des travaux dans le domaine de la compression de vidéos.

1.2.2 Prise en compte du buffer de transmission

La prise en compte du taux d'occupation du buffer de transmission et le respect des contraintes liées au HRD, abordés au chapitre 3, représentent des enjeux importants de la régulation de débit, notamment dans les contextes d'encodage à bas débit et en temps réel.

L'intégration de ces contraintes dans la méthode d'allocation de budget est donc envisagée, de manière à faciliter davantage la transmission des flux scalables. La question du nombre de buffers à utiliser sera plus particulièrement étudiée, en fonction des scénarios d'encodage. Un passage à la dernière version de l'encodeur de référence de MPEG-4 SVC est notamment en projet, car elle fournit les outils nécessaires pour le HRD sur la vidéo scalable.

1.2.3 Régulation de débit inter-couches

En vue de fournir une régulation en une passe plus précise, nous avons étudié la faisabilité d'une approche utilisant les informations de la couche de base pour remplacer celles de l'image précédente, qui dans certaines configurations ne représente pas une référence fiable.

Malgré les problèmes de mise à l'échelle entre les couches, cette approche semble très prometteuse, du fait de la relation notablement linéaire qui existe entre les paramètres de quantification des différents niveaux. On peut également envisager sa combinaison avec l'approche en une passe développée, à l'aide d'un critère de pertinence permettant de choisir l'une ou l'autre des approches en fonction du comportement des données à coder.

Cette perspective de travail représente l'enjeu majeur de nos futures recherches. Une méthode telle que celle abordée ici offrirait en effet une régulation plus précise, en ajoutant une quantité de calculs très faible par rapport à la régulation en deux passes. L'optimisation du débit pour l'ensemble des couches sera également envisagée.

1.2.4 Régulation globale du débit

Pour aborder un aspect davantage lié à l'évaluation des possibilités apportées par le codage vidéo scalable, l'étude de stratégies d'optimisation globale du débit s'impose comme un enjeu majeur.

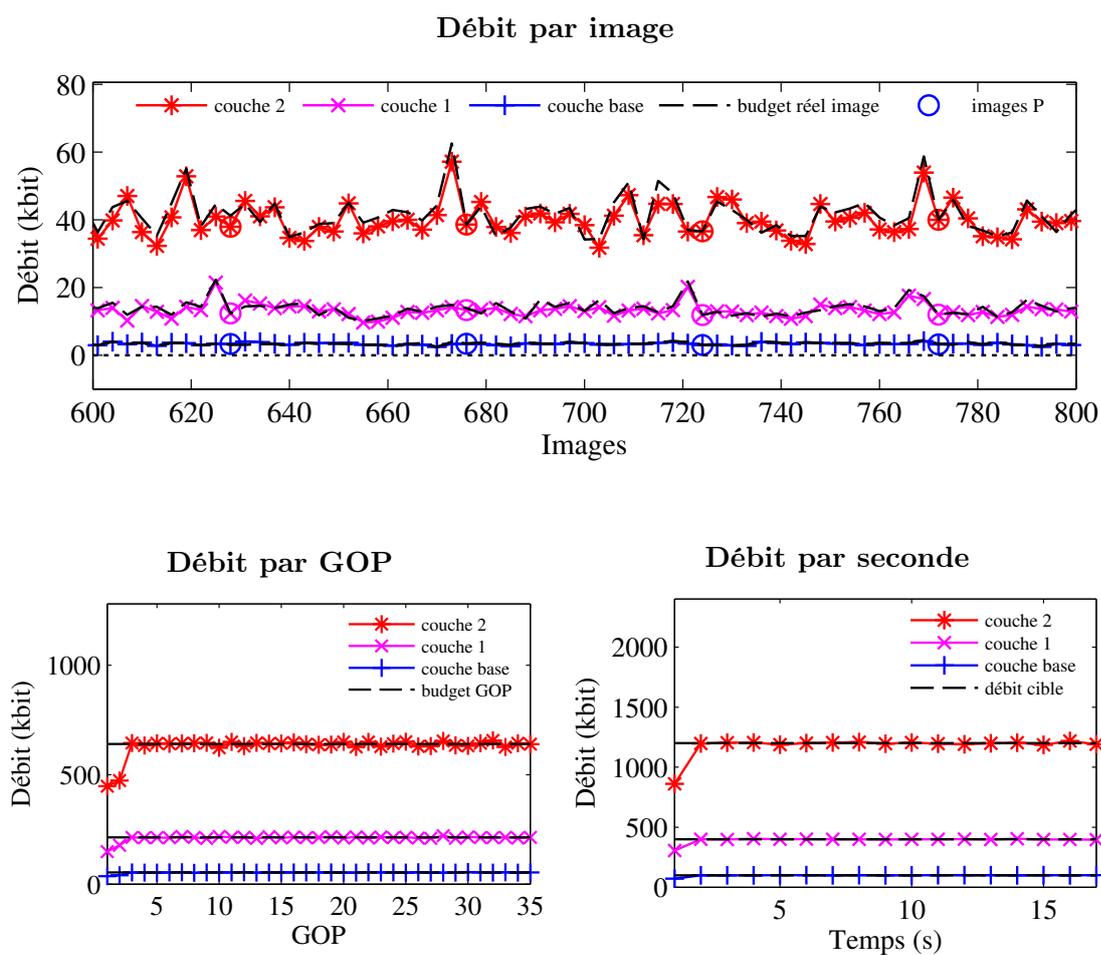
L'analyse des différents contextes applicatifs et des approches de la répartition du débit entre les niveaux de scalabilité permettraient d'évaluer leur impact sur la qualité et l'adaptabilité des flux encodés. La question du nombre de couches à utiliser devra notamment être explorée, ainsi que celle de la priorité à accorder à la couche de base face aux couches d'amélioration. Cette étude représenterait elle aussi une perspective ambitieuse de nos recherches futures, par le nombre de paramètres à prendre en compte, et de l'ampleur des configurations envisageables.

Annexes

Annexe A

Résultats complémentaires pour la régulation en deux passes

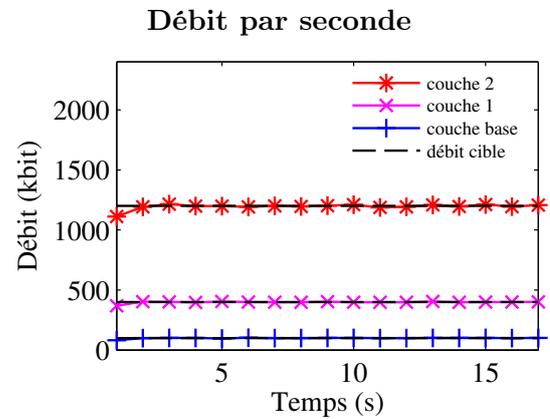
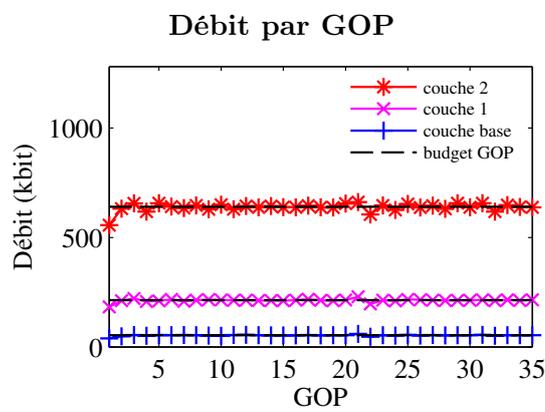
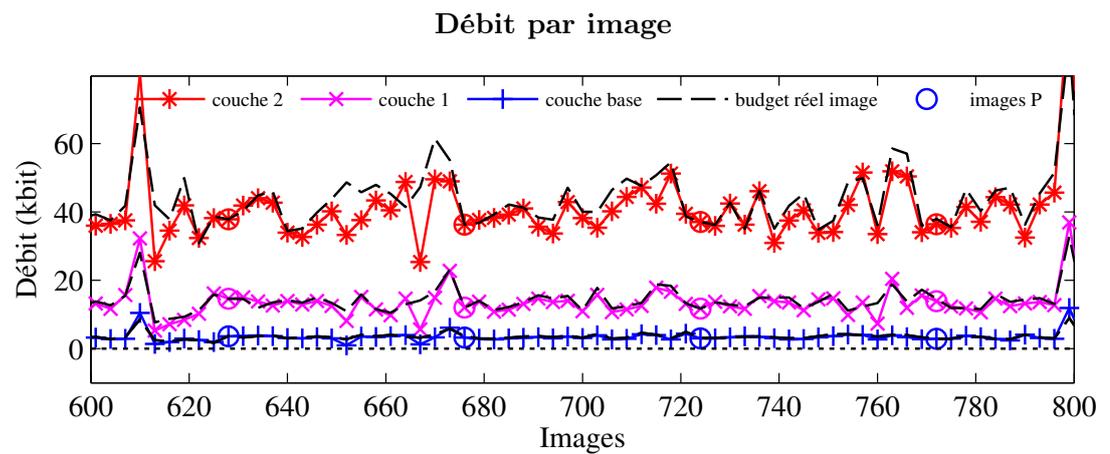
FIGURE A.1 – Performances du contrôle de débit pour le scénario **spatial** avec des GOP de 16 images sur la séquence **SOCCER**.



Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-2.62 %	9.93	-2.75 %	6.99
couche 1	-5.05 %	10.98	-4.77 %	6.07
couche 2	-8.70 %	12.20	-7.62 %	7.30

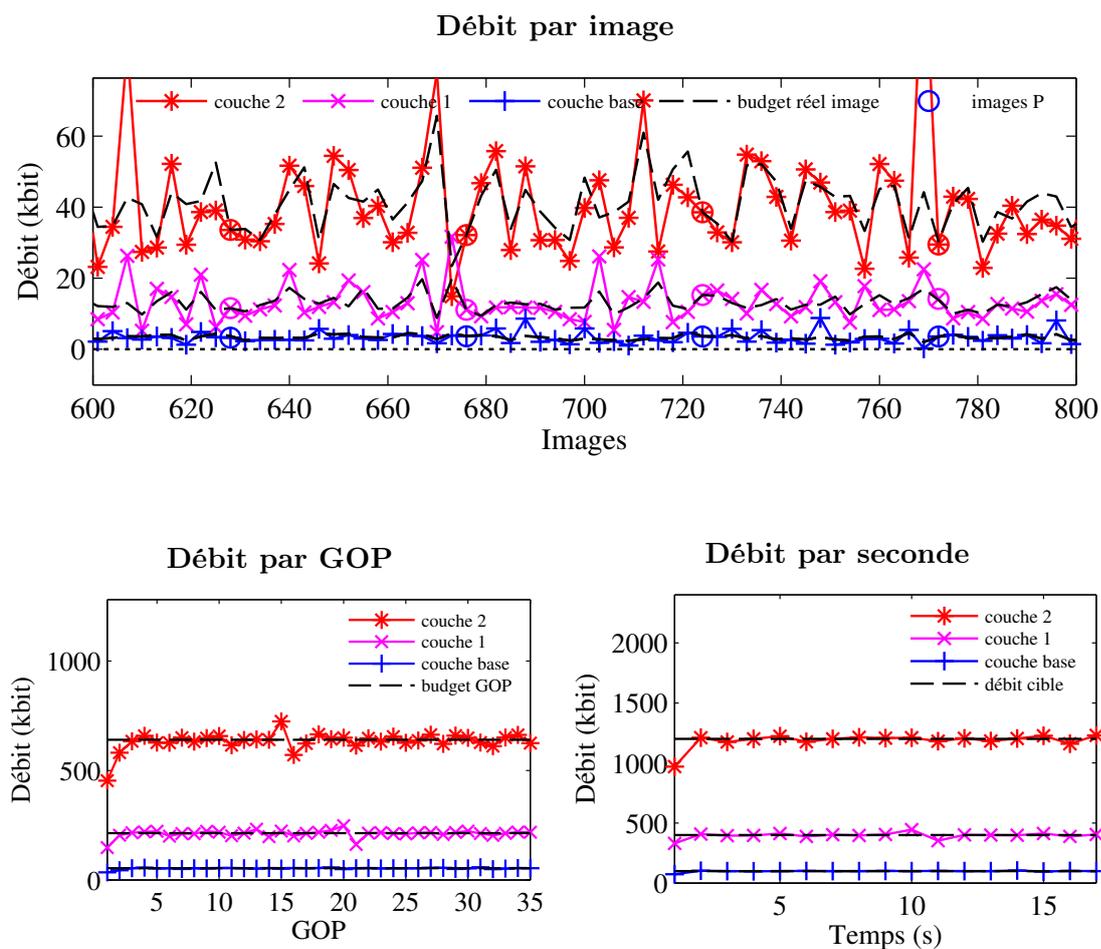
FIGURE A.2 – Performances du contrôle de débit pour le scénario **spatial** avec des GOP de 16 images sur la séquence **CREW**.



Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-4.77 %	20.15	-5.00 %	12.94
couche 1	-6.18 %	17.97	-5.91 %	12.79
couche 2	-8.13 %	15.69	-7.30 %	11.89

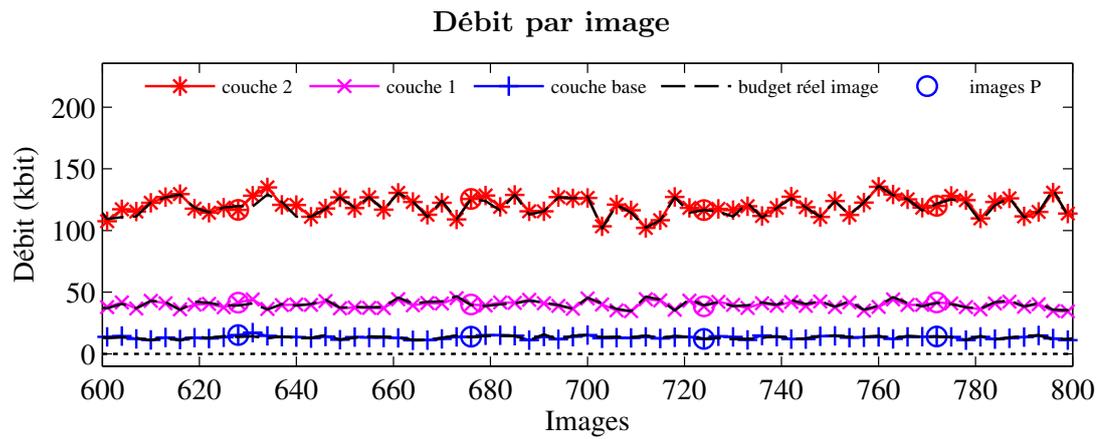
FIGURE A.3 – Performances du contrôle de débit pour le scénario **spatial** avec des GOP de 16 images sur la séquence **CITY**.



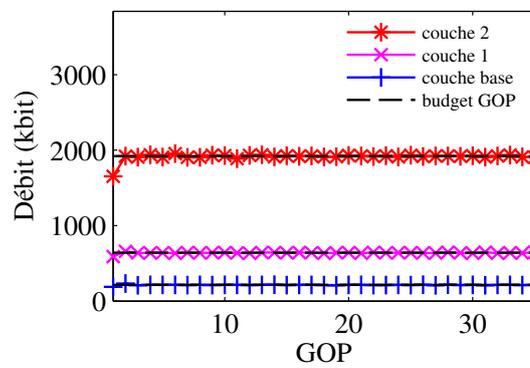
Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-2.65 %	32.39	-2.00 %	29.58
couche 1	-1.41 %	32.74	-0.80 %	29.15
couche 2	-1.16 %	45.12	-2.26 %	33.96

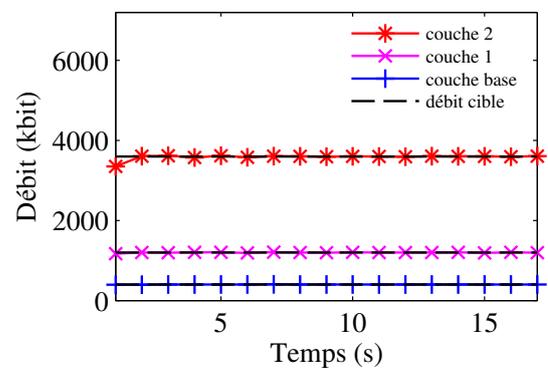
FIGURE A.4 – Performances du contrôle de débit pour le scénario en **qualité** avec des GOP de 16 images sur la séquence **HOCKEY**.



Débit par GOP

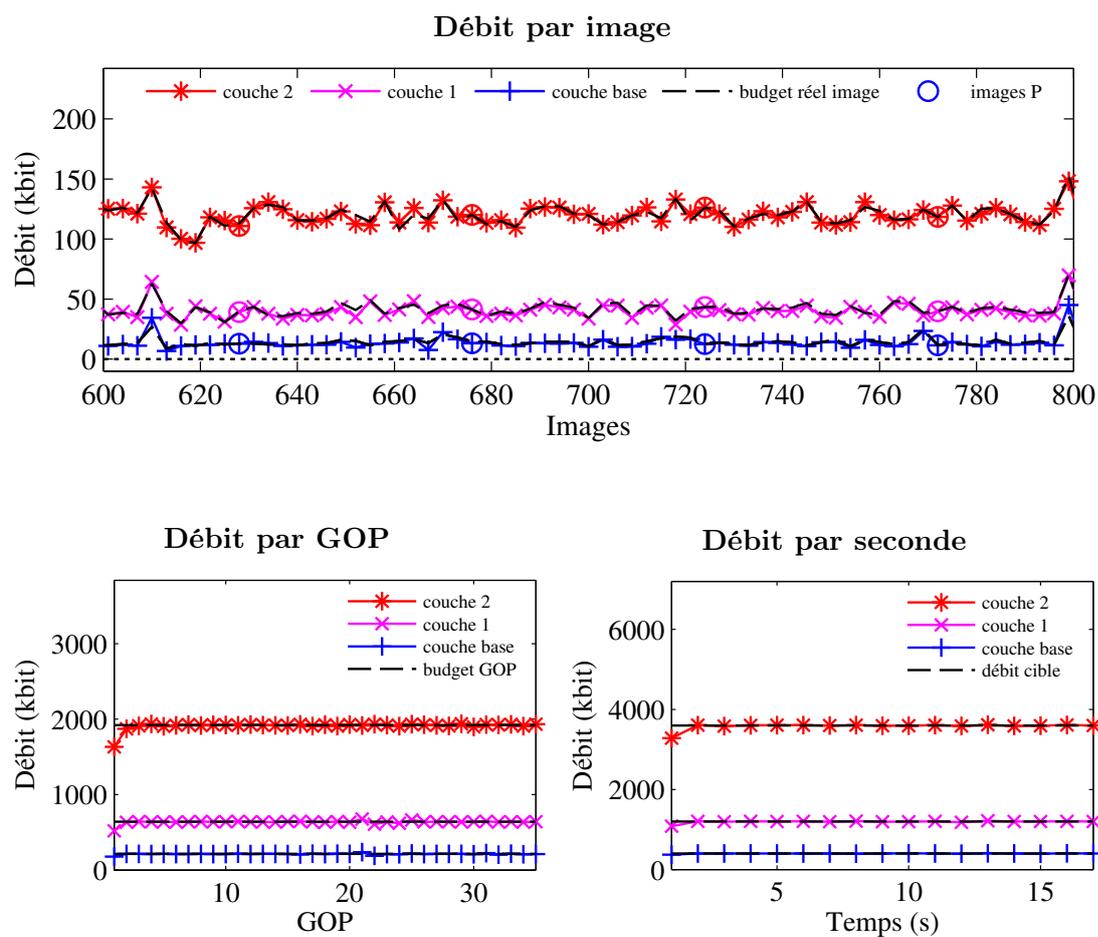


Débit par seconde



Erreur de débit par image pour chaque couche.

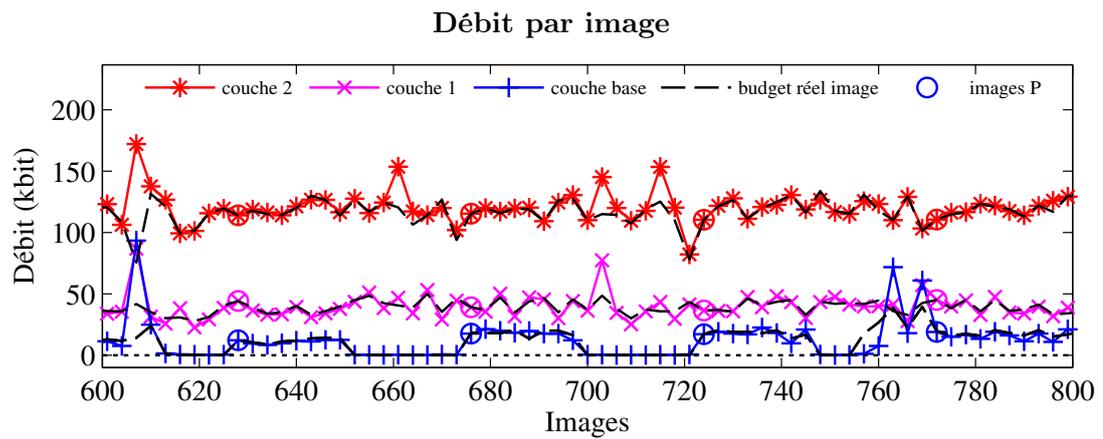
	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-2.74 %	8.94	-1.90 %	5.29
couche 1	-1.06 %	6.54	-0.95 %	2.46
couche 2	-0.44 %	6.49	+0.51 %	3.03

FIGURE A.5 – Performances du contrôle de débit pour le scénario en **qualité** avec des GOP de 16 images sur la séquence **CREW**.

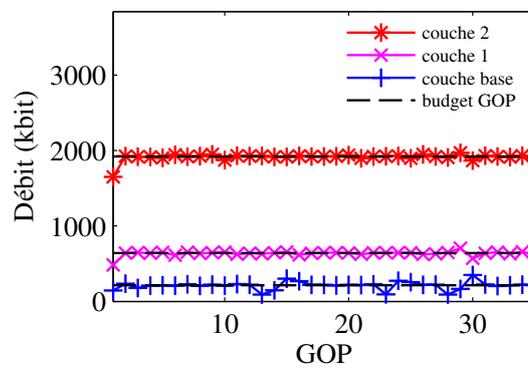
Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-6.38 %	21.69	-6.22 %	14.33
couche 1	-3.83 %	10.74	-3.71 %	6.40
couche 2	+0.10 %	5.47	+0.07 %	2.43

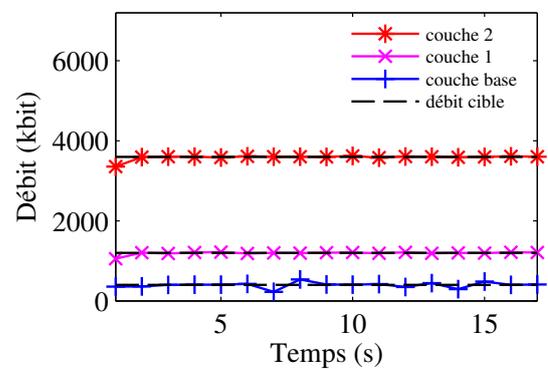
FIGURE A.6 – Performances du contrôle de débit pour le scénario en **qualité** avec des GOP de 16 images sur la séquence **CITY**.



Débit par GOP



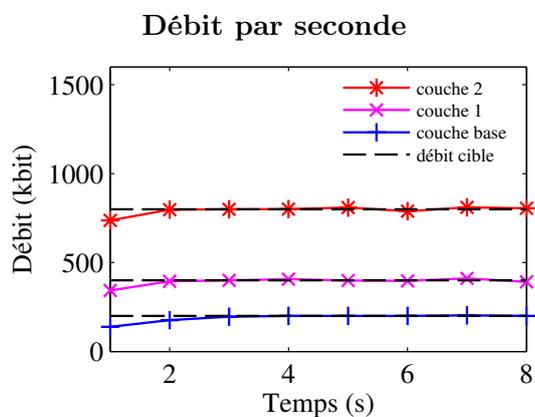
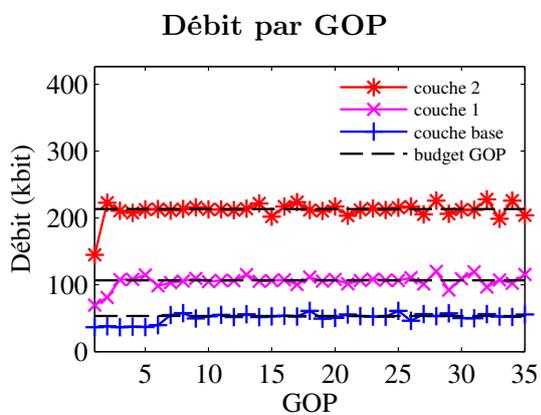
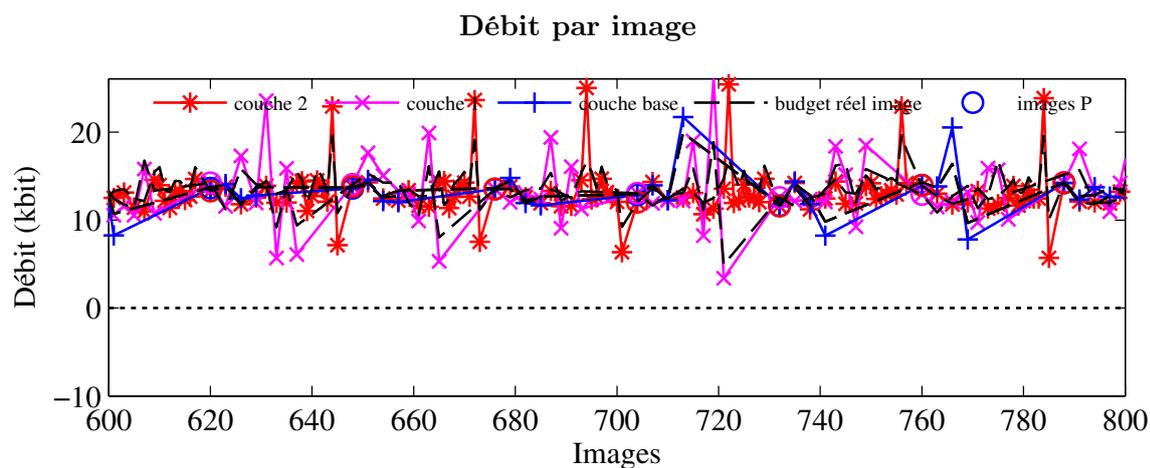
Débit par seconde



Erreur de débit par image pour chaque couche.

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	+5.98 %	147.77	+5.39 %	58.81
couche 1	+12.88 %	222.97	+1.80 %	21.93
couche 2	+3.06 %	15.71	+5.49 %	25.77

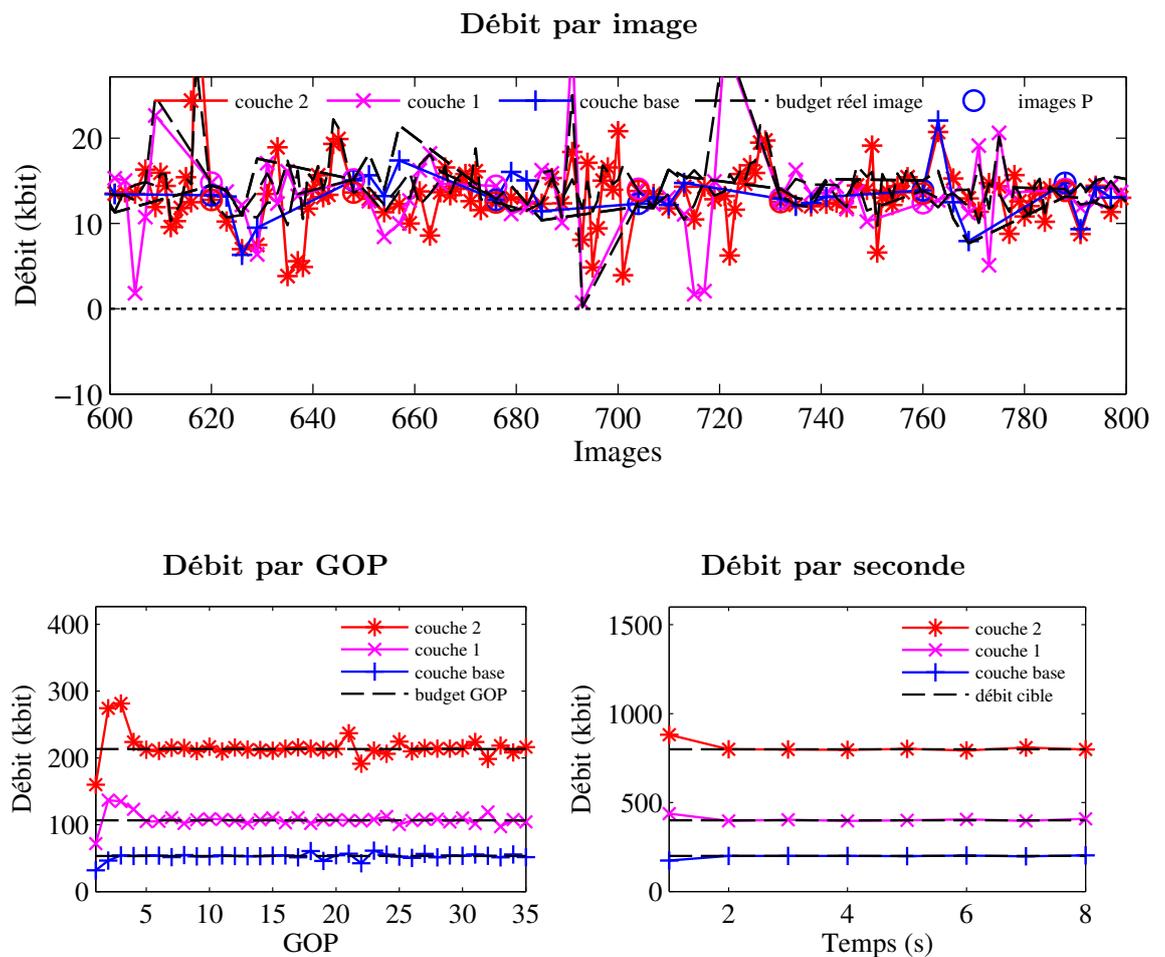
FIGURE A.7 – Performances du contrôle de débit pour le scénario **temporel** avec des GOP de 16 images sur la séquence **HARBOUR**.



Erreur de débit par image pour chaque couche.

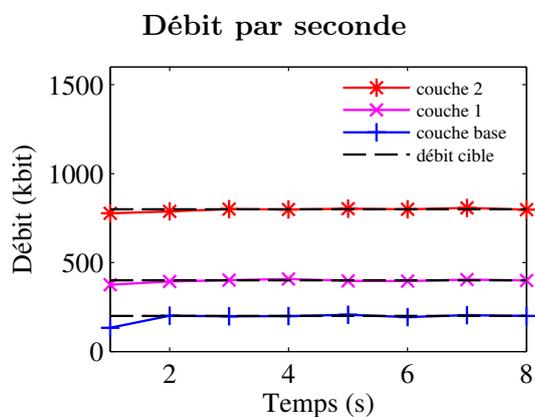
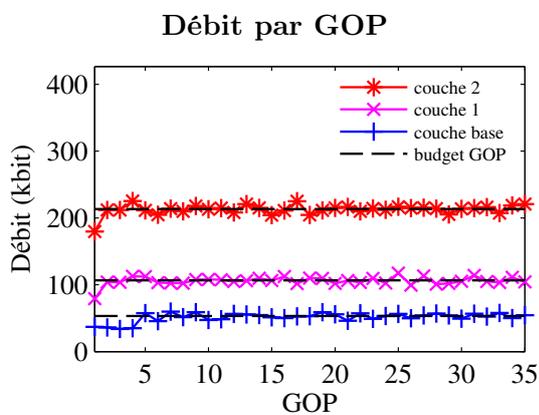
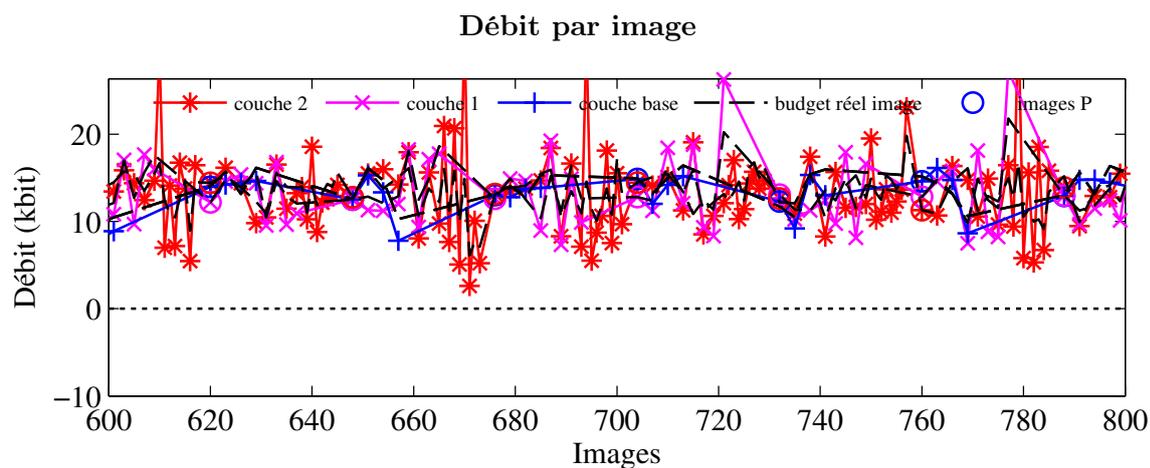
	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-1.26 %	12.92	-0.65 %	7.00
couche 1	-2.06 %	22.51	-0.83 %	16.90
couche 2	-3.76 %	21.47	-3.22 %	16.54

FIGURE A.8 – Performances du contrôle de débit pour le scénario **temporel** avec des GOP de 16 images sur la séquence **CREW**.



Erreur de débit par image pour chaque couche.

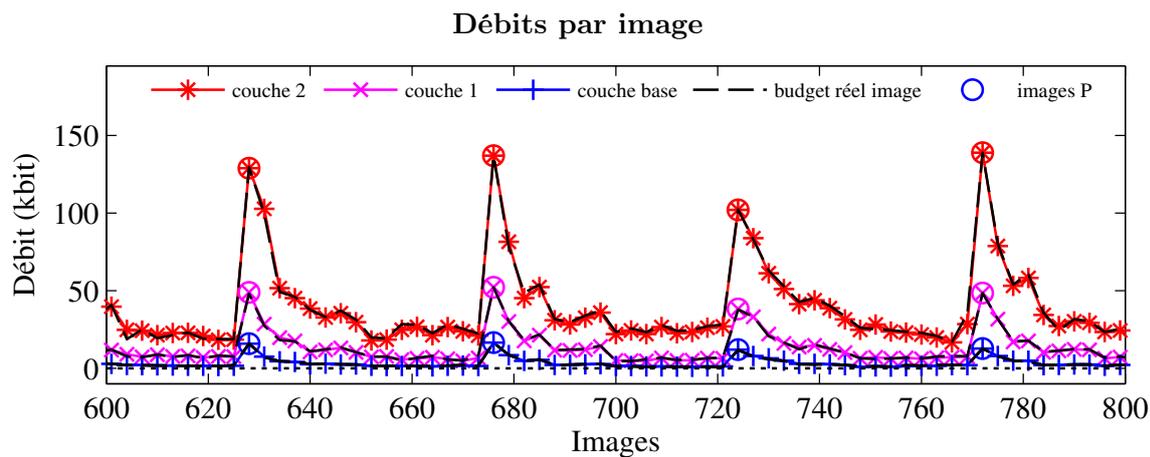
	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-2.55 %	13.20	-2.37 %	12.60
couche 1	-4.69 %	27.25	-4.40 %	17.90
couche 2	-5.71 %	22.05	-5.89 %	15.58

FIGURE A.9 – Performances du contrôle de débit pour le scénario **temporel** avec des GOP de 16 images sur la séquence **CITY**.**Erreur de débit par image pour chaque couche.**

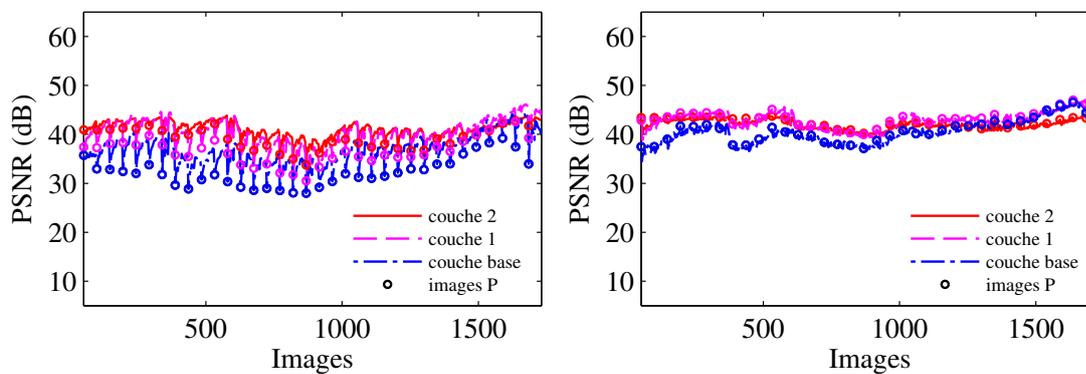
	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-3.42 %	15.60	-2.70 %	9.57
couche 1	-3.85 %	20.22	-2.38 %	17.09
couche 2	-0.62 %	29.50	-1.81 %	22.92

Annexe B

Résultats complémentaires pour la régulation à qualité constante

FIGURE B.1 – Améliorations apportées sur la qualité pour le scénario **spatial** sur la séquence **HOCKEY**.

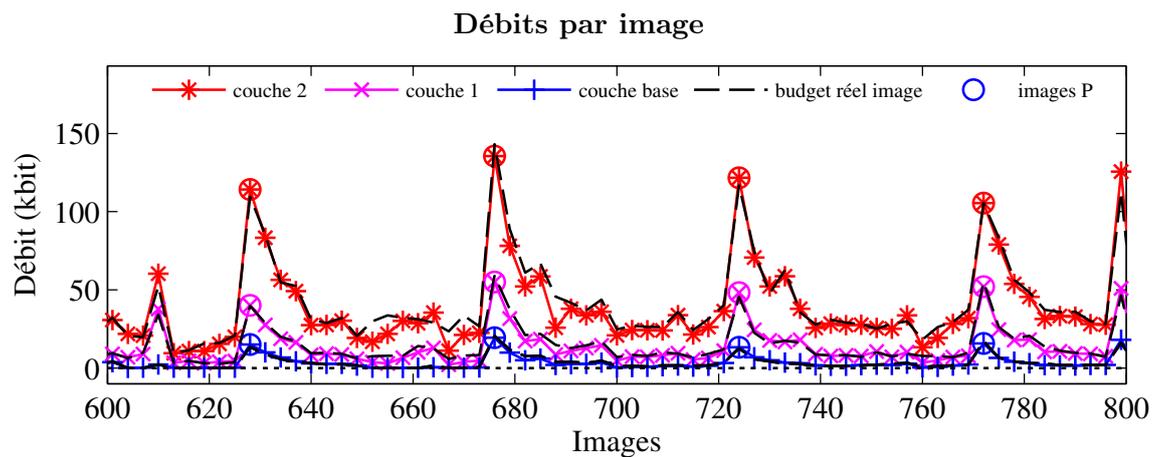
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



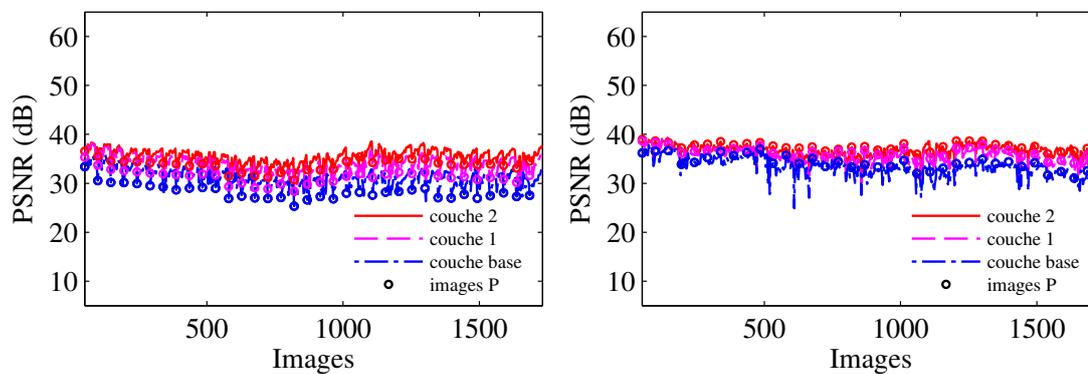
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	1.68	1.46	0.81
avec l'efficacité de codage	0.53	0.46	0.29

FIGURE B.2 – Améliorations apportées sur la qualité pour le scénario **spatial** sur la séquence **CREW**.

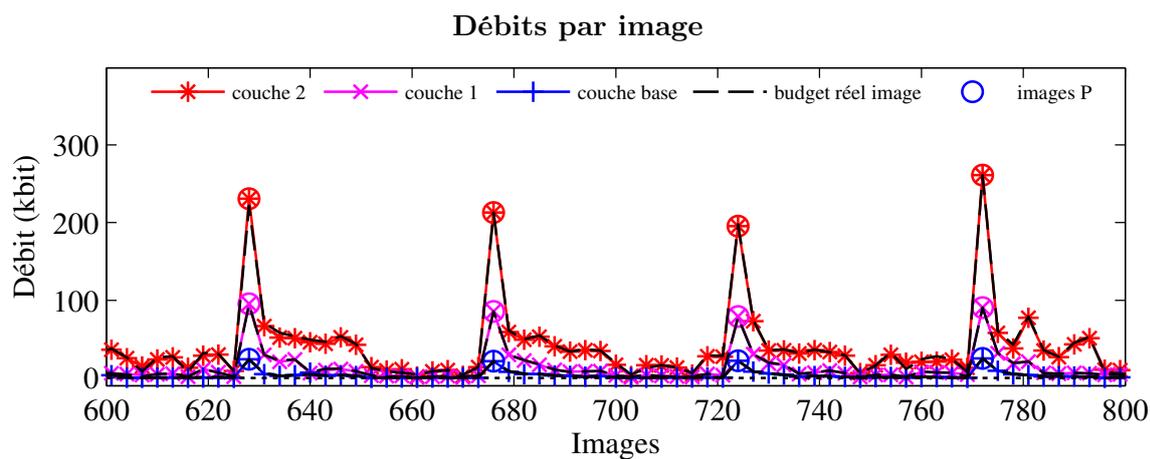


PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage

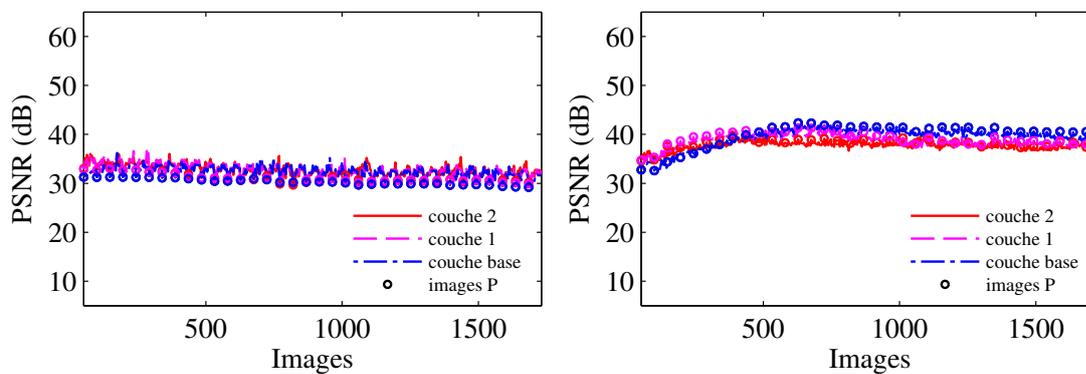


écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	1.31	1.22	1.01
avec l'efficacité de codage	1.03	0.91	0.75

FIGURE B.3 – Améliorations apportées sur la qualité pour le scénario **spatial** sur la séquence **CITY**.

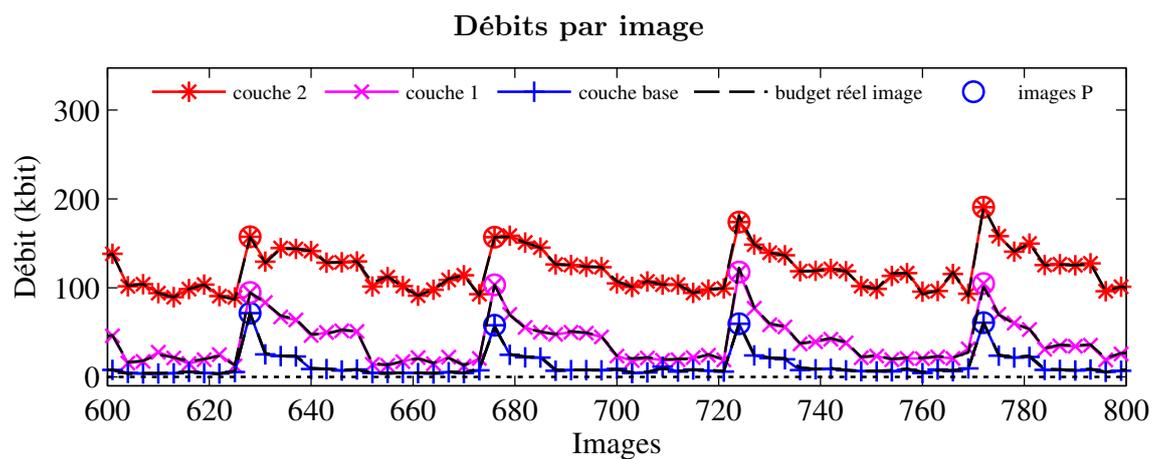
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



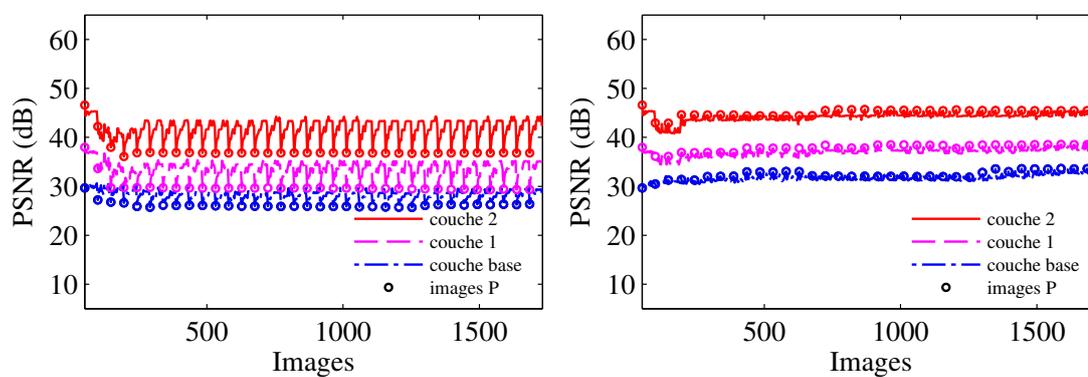
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.90	0.76	0.77
avec l'efficacité de codage	0.43	0.37	0.39

FIGURE B.4 – Améliorations apportées sur la qualité pour le scénario **qualité** sur la séquence **HARBOUR**.

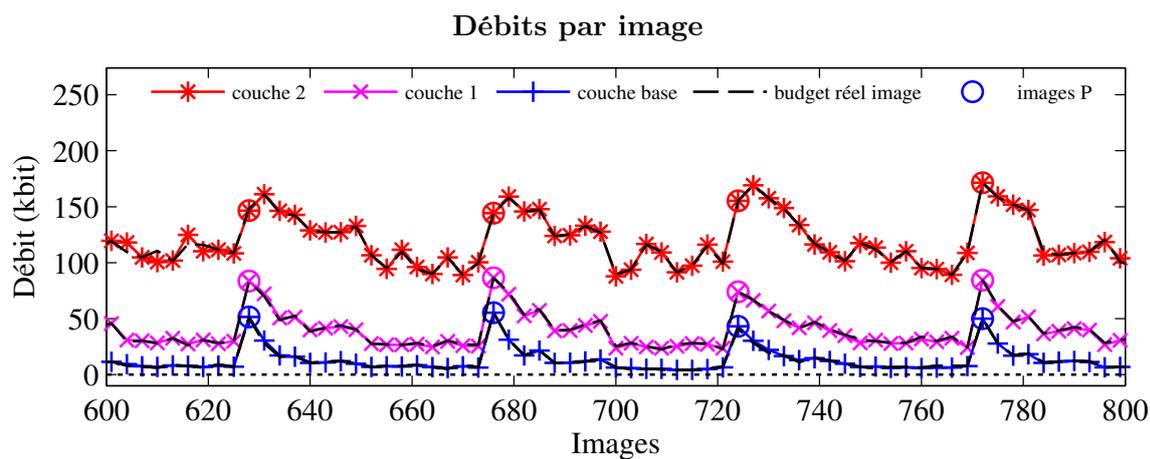


PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage

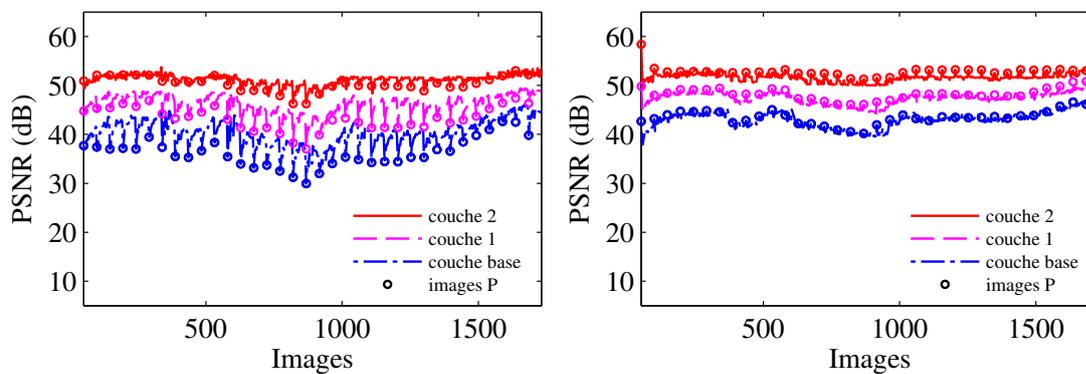


écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.84	1.40	1.79
avec l'efficacité de codage	0.39	0.41	0.44

FIGURE B.5 – Améliorations apportées sur la qualité pour le scénario **qualité** sur la séquence **HOCKEY**.

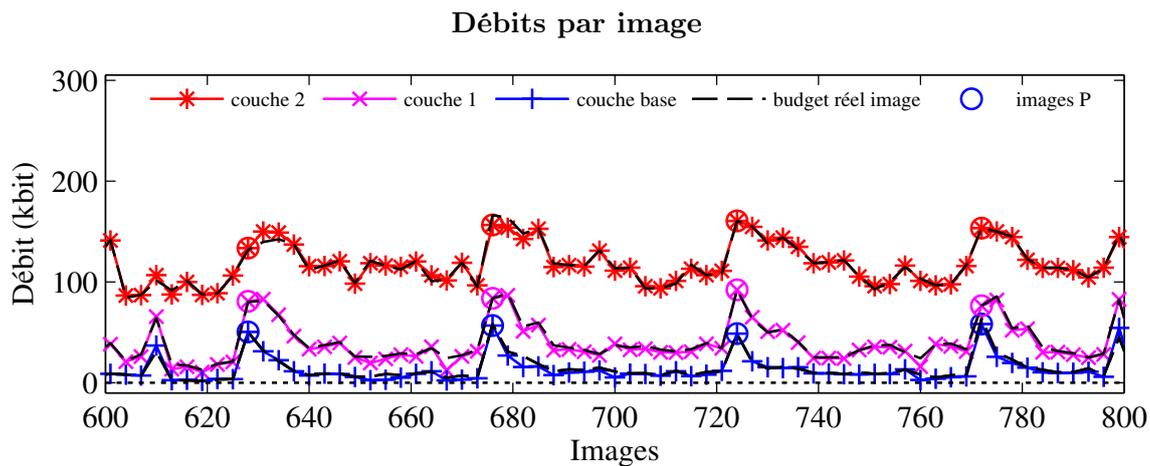
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



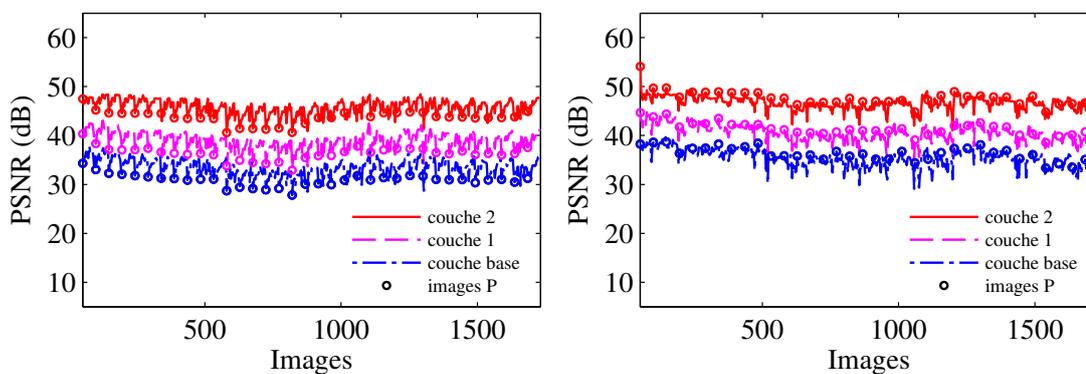
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	1.58	1.13	0.59
avec l'efficacité de codage	0.46	0.41	0.53

FIGURE B.6 – Améliorations apportées sur la qualité pour le scénario **qualité** sur la séquence **CREW**.

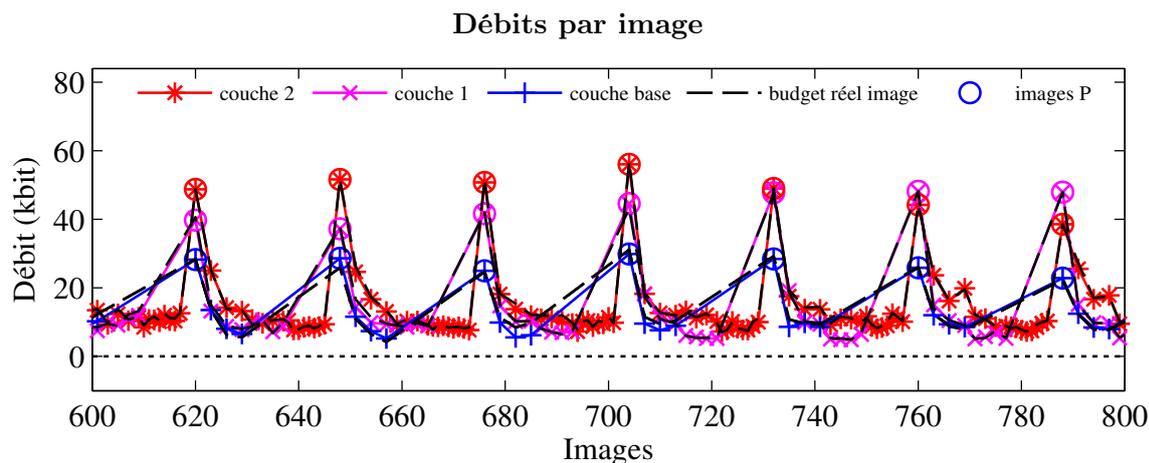


PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage

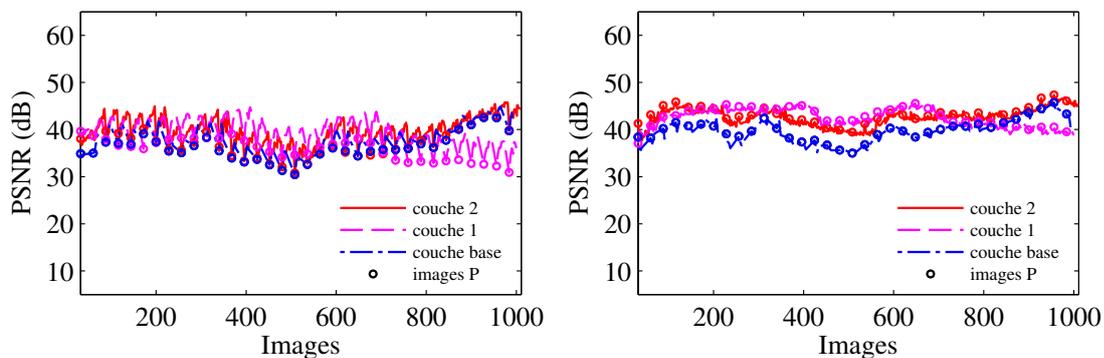


écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	1.28	1.29	1.19
avec l'efficacité de codage	1.02	0.99	0.96

FIGURE B.7 – Améliorations apportées sur la qualité pour le scénario **temporel** sur la séquence **HOCKEY**.

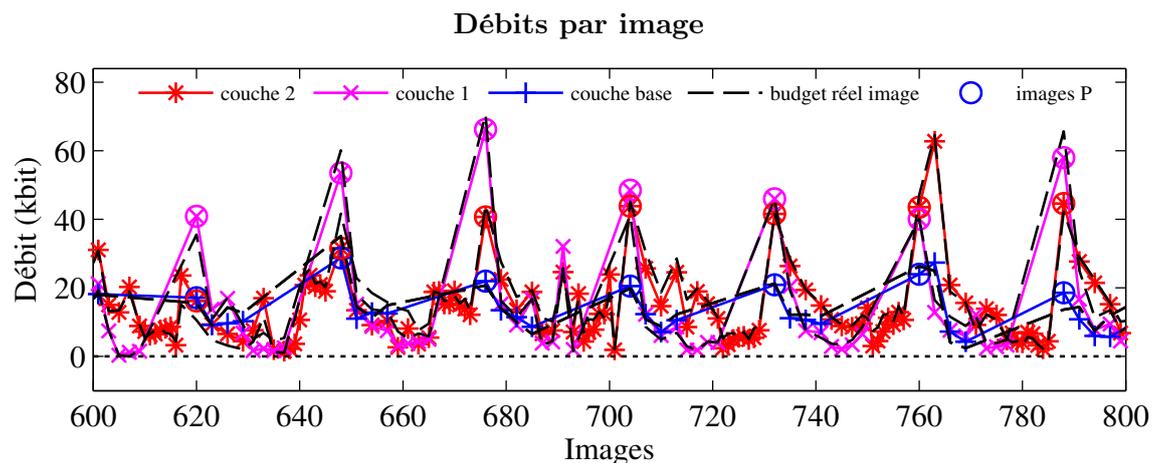
PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



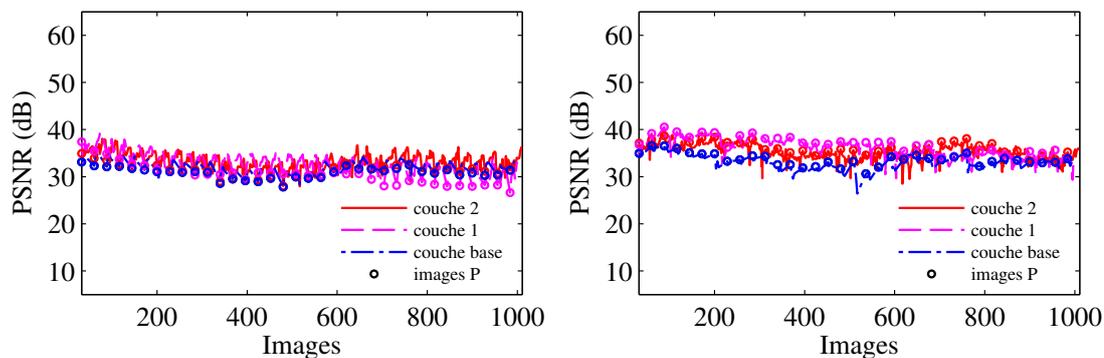
écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	1.20	1.73	1.42
avec l'efficacité de codage	0.59	0.47	0.48

FIGURE B.8 – Améliorations apportées sur la qualité pour le scénario **temporel** sur la séquence **CREW**.

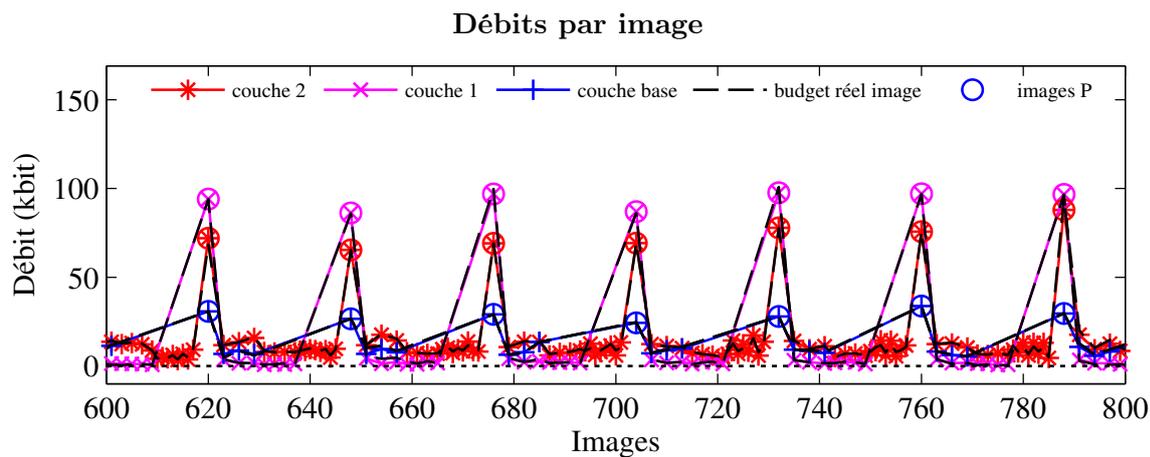


PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage

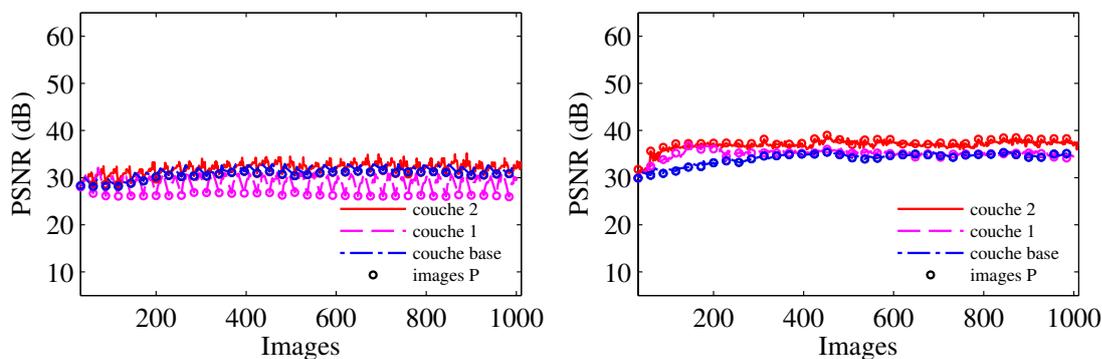


écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.90	1.36	1.20
avec l'efficacité de codage	0.63	0.88	0.97

FIGURE B.9 – Améliorations apportées sur la qualité pour le scénario **temporel** sur la séquence **CITY**.

PSNR par image sans l'efficacité de codage PSNR par image avec l'efficacité de codage



écart type moyen du PSNR par GOP pour chaque couche.

	couche 0	couche 1	couche 2
sans l'efficacité de codage	0.63	1.32	0.78
avec l'efficacité de codage	0.22	0.23	0.35

Annexe C

Résultats complémentaires pour la régulation en une passe

FIGURE C.1 – Performances du contrôle de débit en **une passe** pour le scénario **spatial** avec des GOP de 16 images sur la séquence **HOCKEY**.

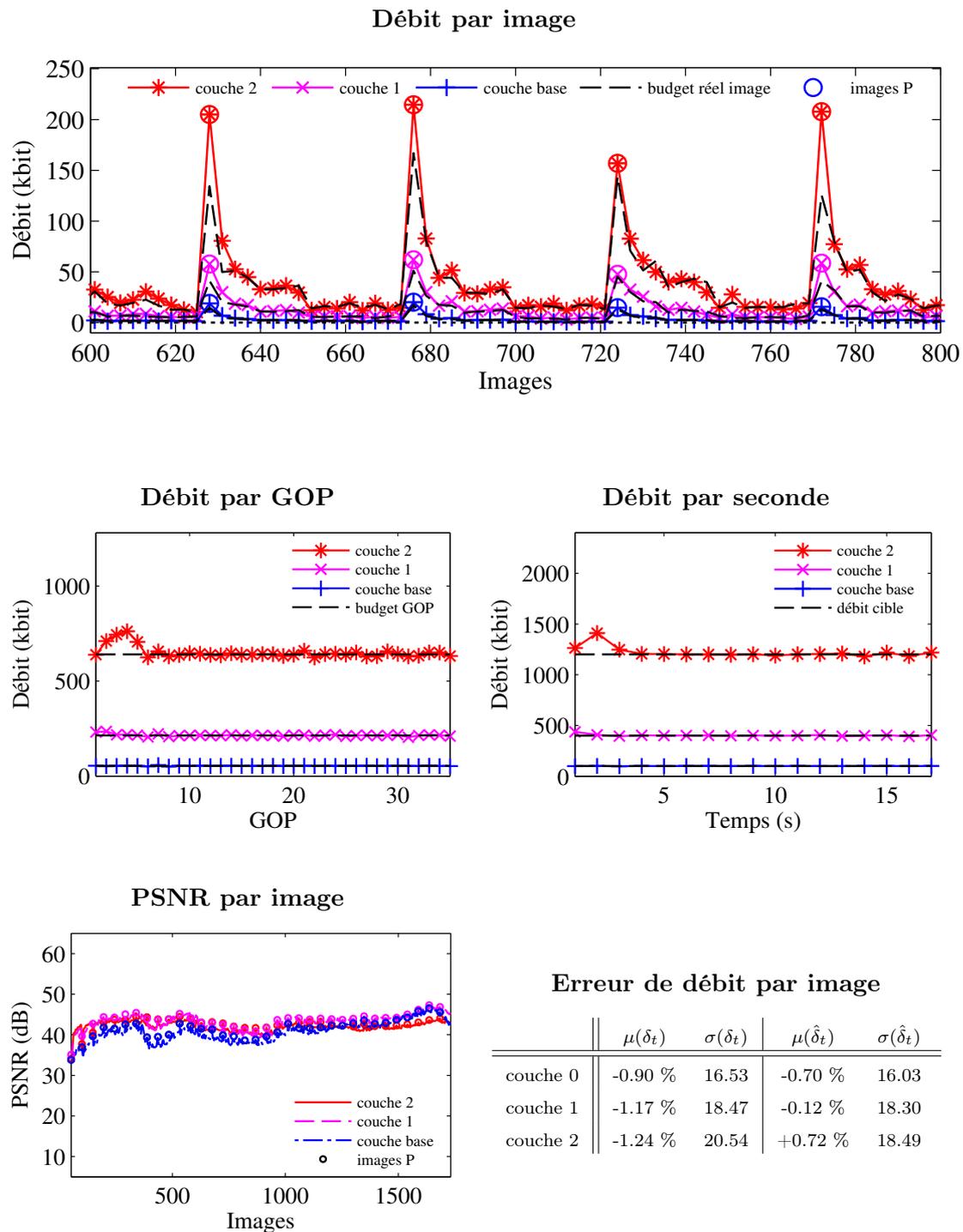
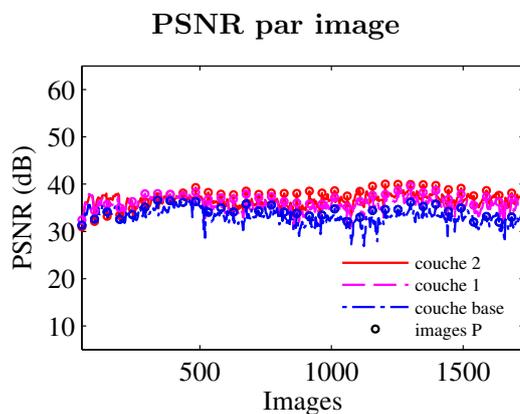
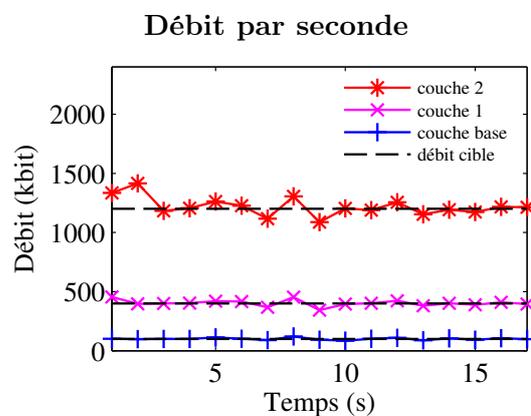
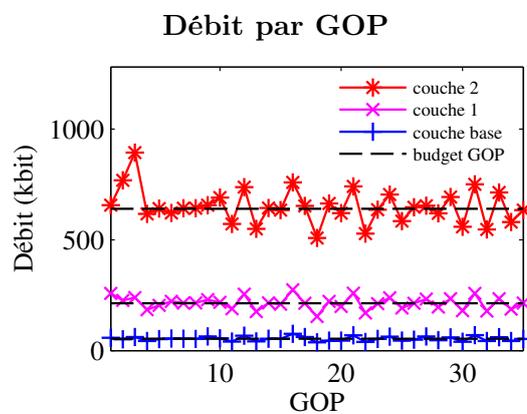
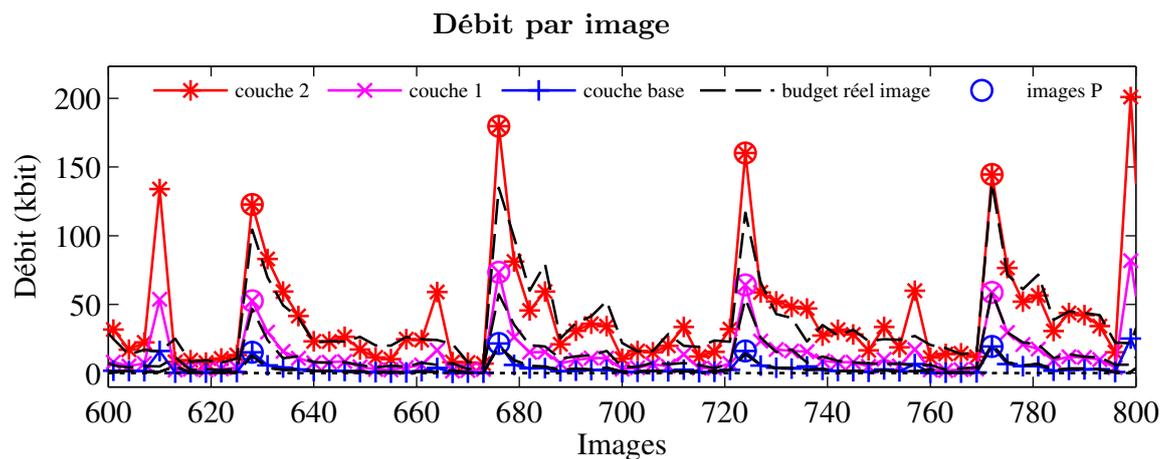


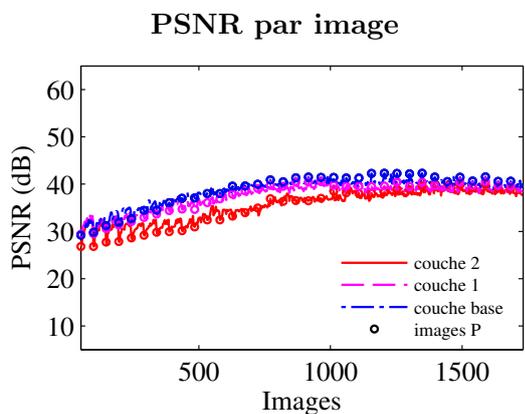
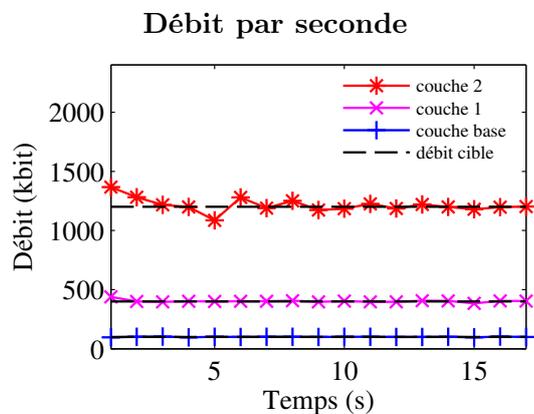
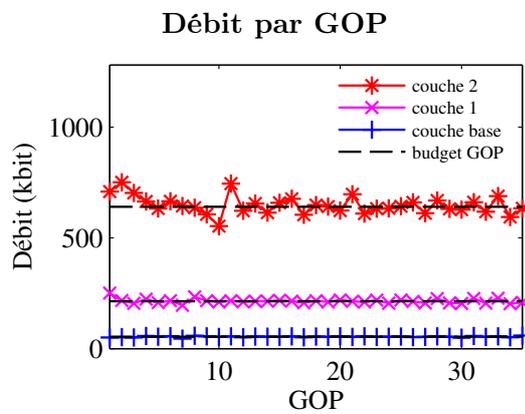
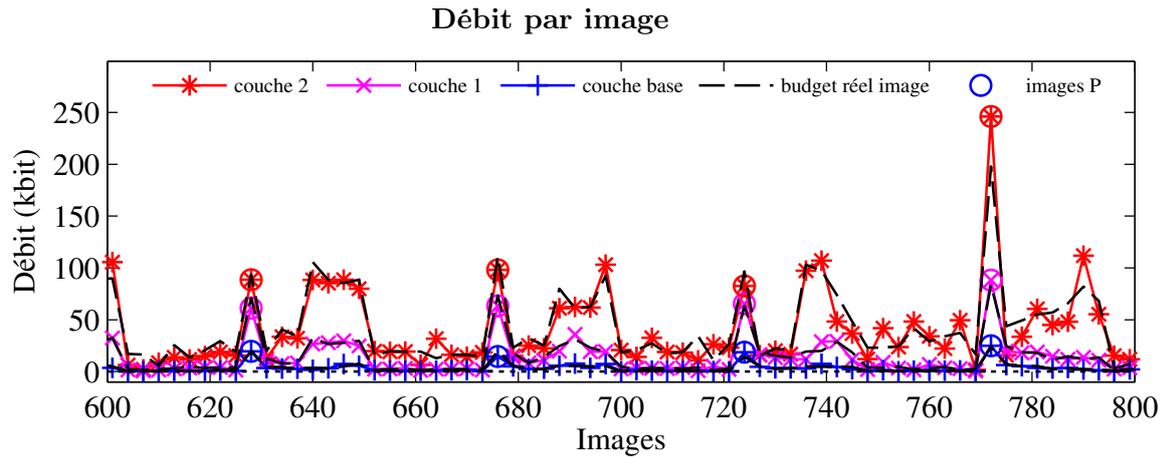
FIGURE C.2 – Performances du contrôle de débit en **une passe** pour le scénario **spatial** avec des GOP de 16 images sur la séquence **CREW**.



Erreur de débit par image

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	+11.37 %	92.40	+13.93 %	94.60
couche 1	+12.59 %	94.13	+15.67 %	98.43
couche 2	+11.09 %	83.93	+12.54 %	80.32

FIGURE C.3 – Performances du contrôle de débit en **une passe** pour le scénario **spatial** avec des GOP de 16 images sur la séquence **CITY**.



Erreur de débit par image

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-0.81 %	37.49	+2.17 %	36.58
couche 1	+6.61 %	64.15	+8.59 %	62.46
couche 2	+8.67 %	77.06	+11.66 %	78.25

FIGURE C.4 – Performances du contrôle de débit en **une passe** pour le scénario **qualité** avec des GOP de 16 images sur la séquence **HARBOUR**.

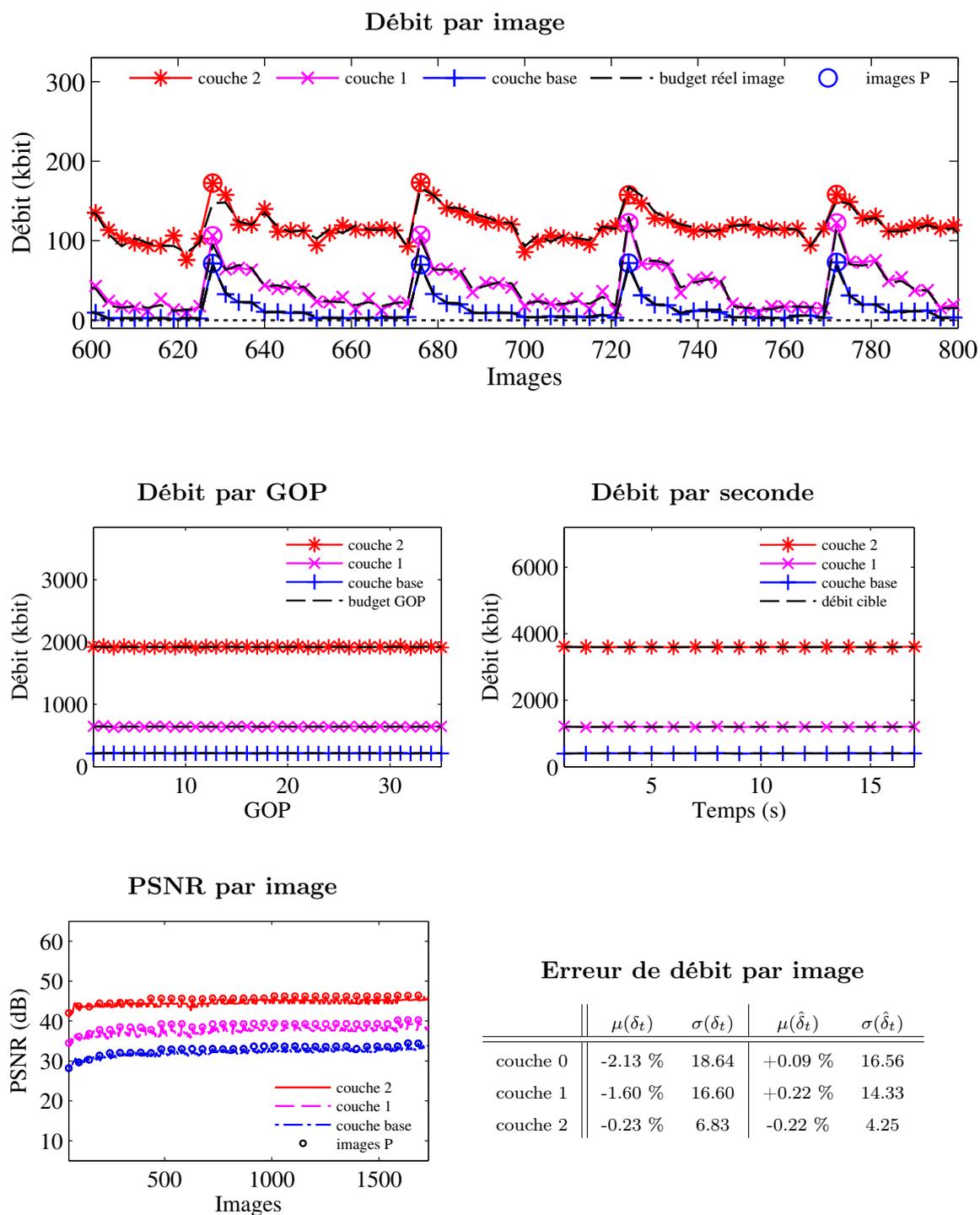


FIGURE C.5 – Performances du contrôle de débit en **une passe** pour le scénario **qualité** avec des GOP de 16 images sur la séquence **SOCCER**.

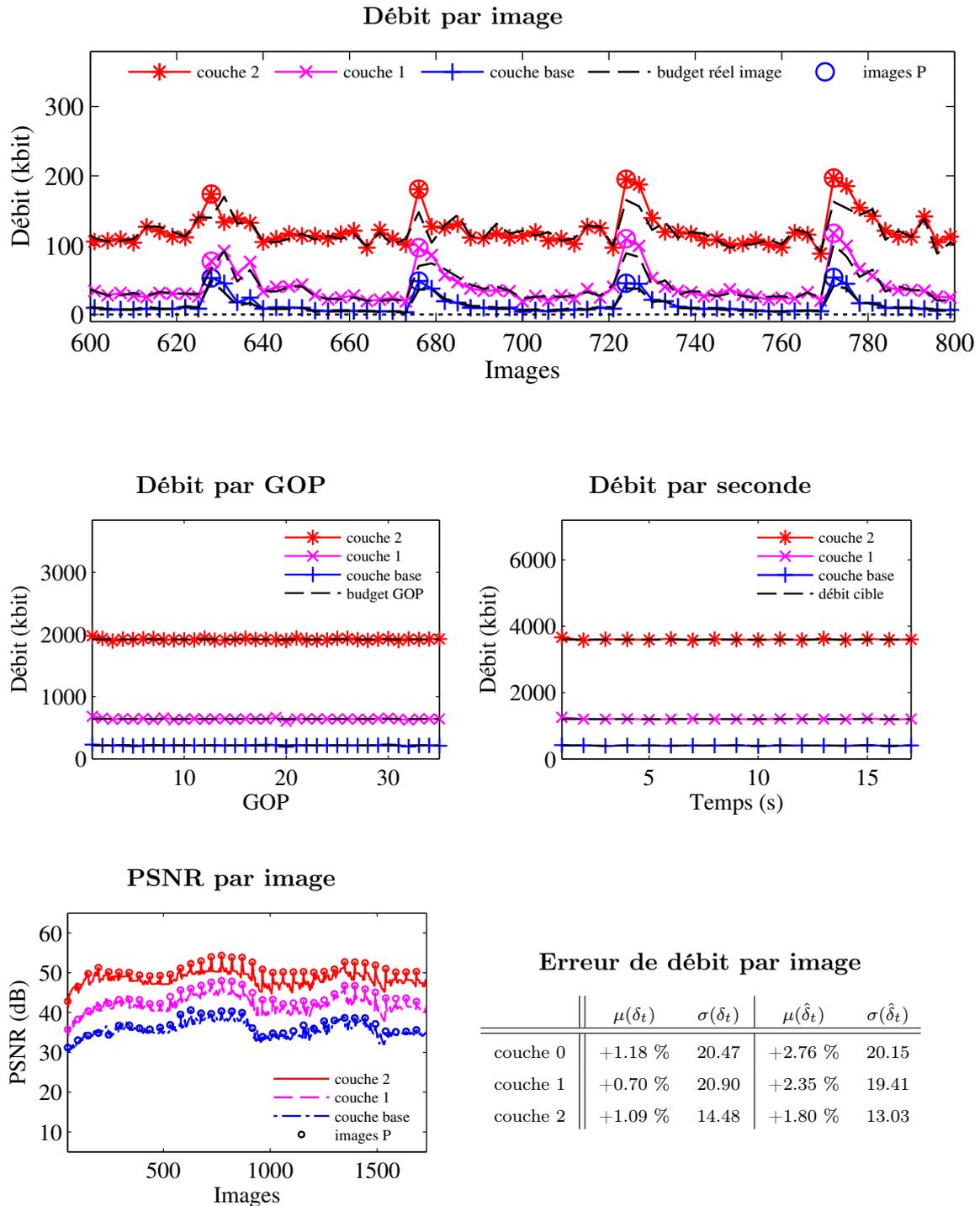


FIGURE C.6 – Performances du contrôle de débit en **une passe** pour le scénario **qualité** avec des GOP de 16 images sur la séquence **CITY**.

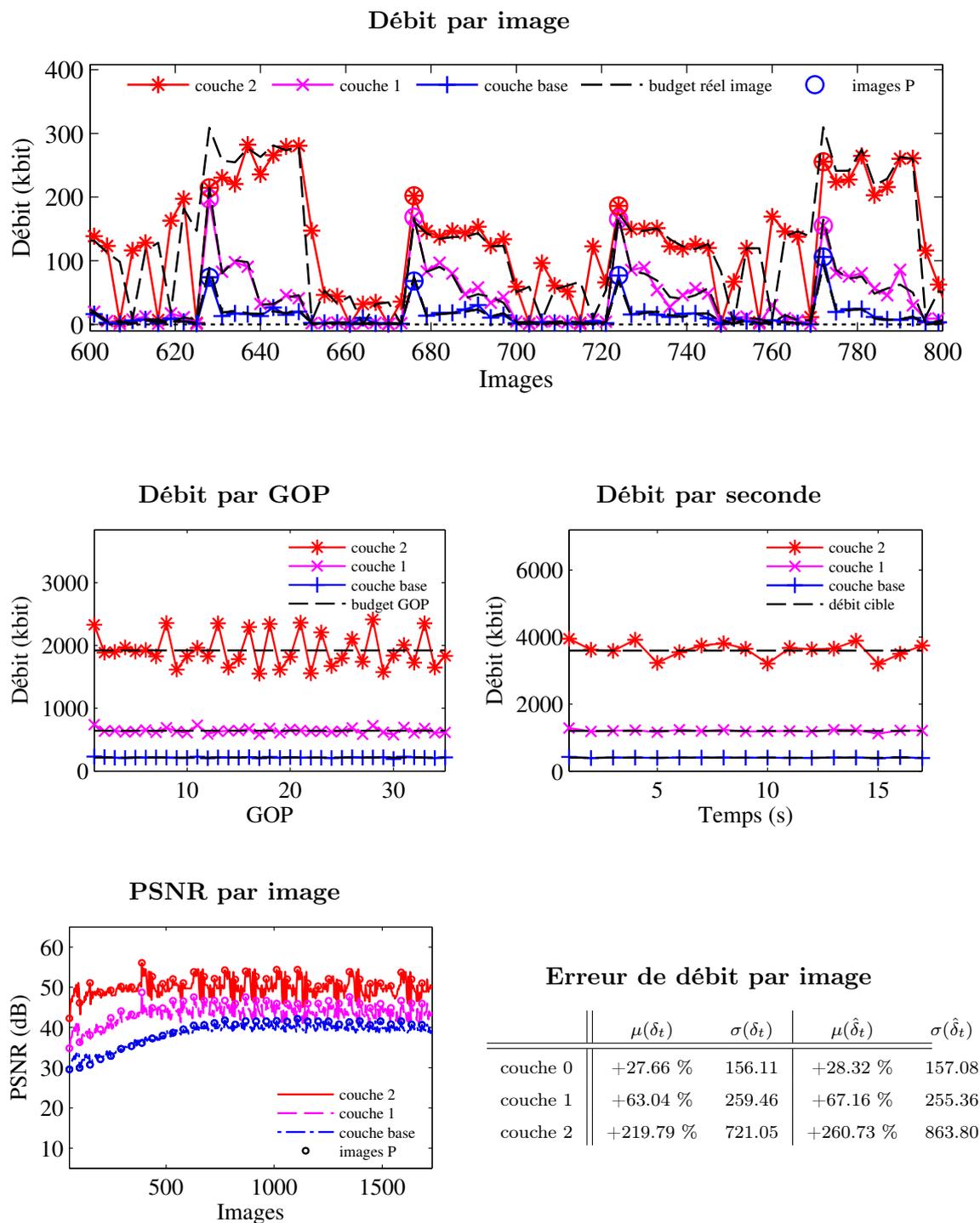
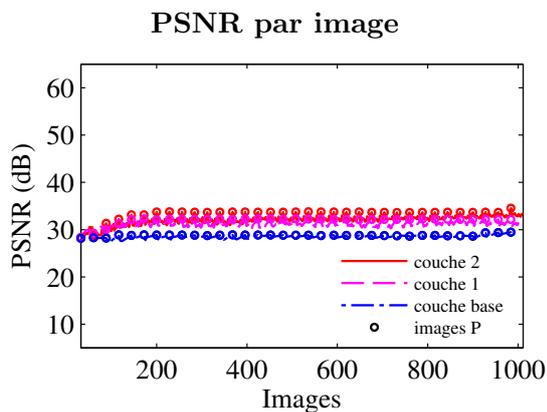
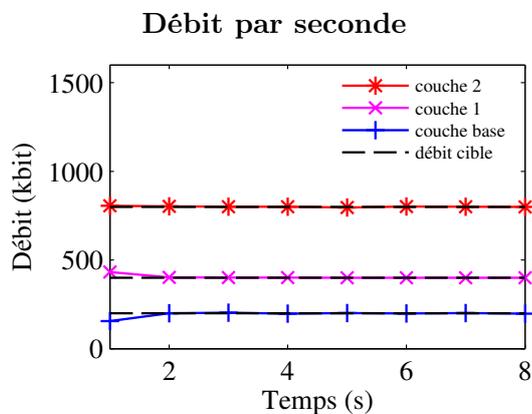
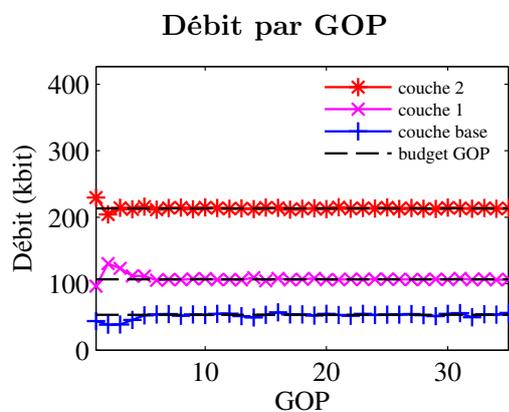
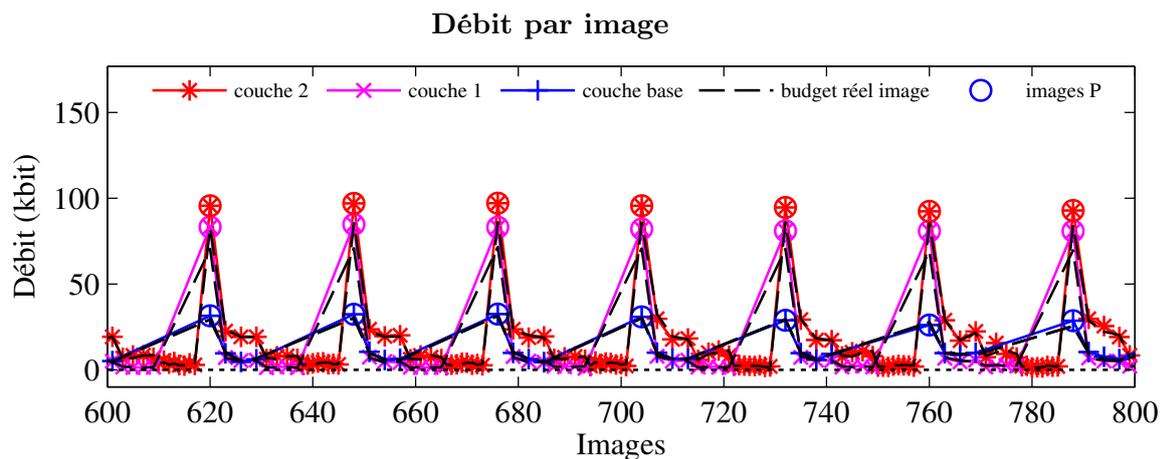


FIGURE C.7 – Performances du contrôle de débit en **une passe** pour le scénario **temporel** avec des GOP de 16 images sur la séquence **HOCKEY**.

Erreur de débit par image

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-1.45 %	12.49	+0.19 %	10.17
couche 1	+2.25 %	48.32	+5.52 %	56.73
couche 2	-1.58 %	16.90	+0.75 %	15.53

FIGURE C.8 – Performances du contrôle de débit en **une passe** pour le scénario **temporel** avec des GOP de 16 images sur la séquence **CREW**.

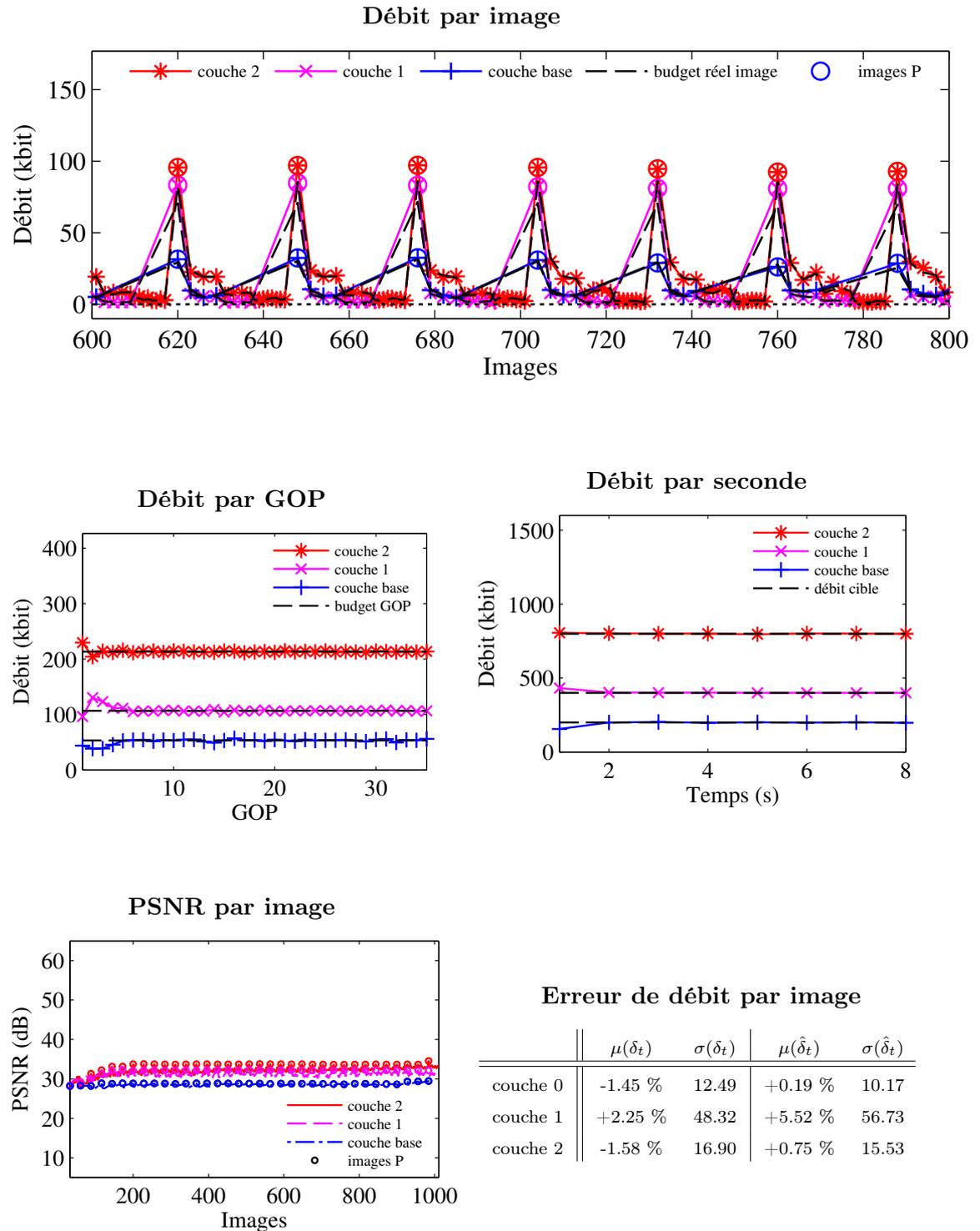
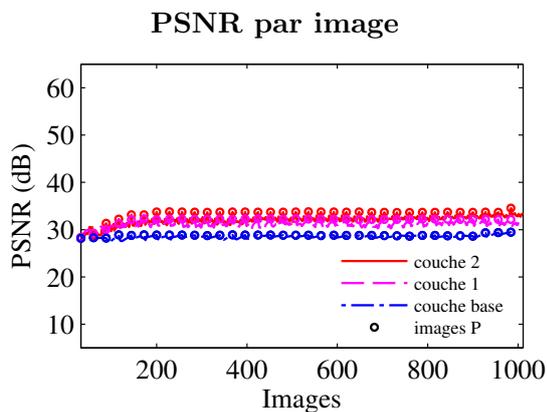
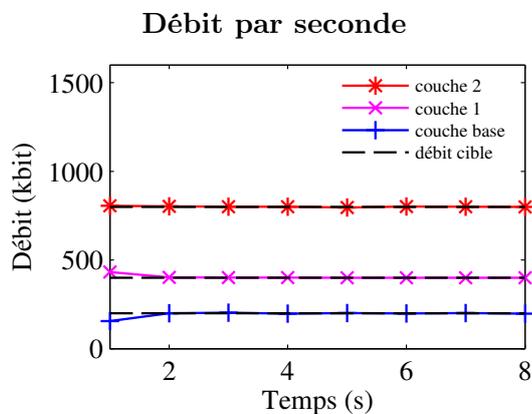
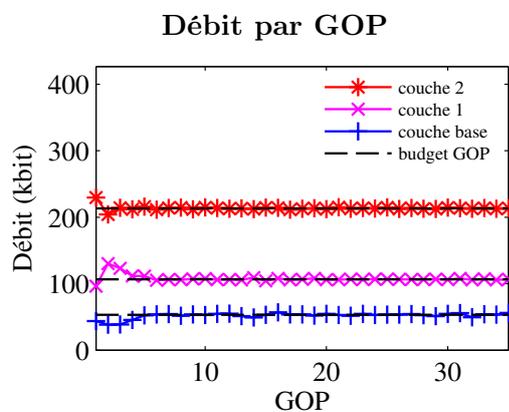
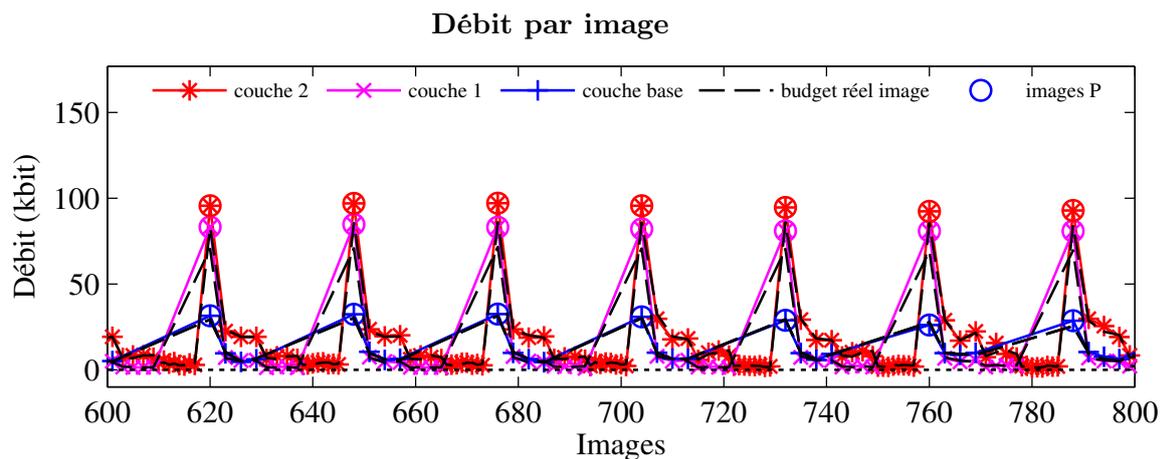


FIGURE C.9 – Performances du contrôle de débit en **une passe** pour le scénario **temporel** avec des GOP de 16 images sur la séquence **CITY**.

Erreur de débit par image

	$\mu(\delta_t)$	$\sigma(\delta_t)$	$\mu(\hat{\delta}_t)$	$\sigma(\hat{\delta}_t)$
couche 0	-1.45 %	12.49	+0.19 %	10.17
couche 1	+2.25 %	48.32	+5.52 %	56.73
couche 2	-1.58 %	16.90	+0.75 %	15.53

Annexe D

Correspondance entre QP initial et contrainte en bits par pixel

TABLE D.1 – Correspondance entre valeur de contrainte en bits par pixel et QP initial pour les trois scalabilités.**SPATIAL**

Couche 0		Couche 1		Couche 2	
bpp	QP_{INIT}	bpp	QP_{INIT}	bpp	QP_{INIT}
0.050	47	0.070	47	0.030	47
0.110	43	0.100	43	0.070	43
0.390	37	0.160	37	0.150	37
0.660	33	0.390	33	0.220	33
1.050	27	0.890	27	0.580	27
1.710	23	1.150	23	0.950	23
2.760	17	1.910	17	1.020	17
3.000	13	3.000	13	2.000	13

QUALITÉ

Couche 0		Couche 1		Couche 2	
bpp	QP_{INIT}	bpp	QP_{INIT}	bpp	QP_{INIT}
0.070	47	0.020	47	0.030	47
0.200	43	0.030	43	0.100	43
0.330	37	0.150	37	0.350	37
0.660	33	0.230	33	0.540	33
1.450	27	0.690	27	0.610	27
2.300	23	0.880	23	0.620	23
3.160	17	1.380	17	1.020	17
4.000	13	2.000	13	2.000	13

TEMPOREL

Couche 0		Couche 1		Couche 2	
bpp	QP_{INIT}	bpp	QP_{INIT}	bpp	QP_{INIT}
0.070	47	0.016	47	0.070	47
0.100	43	0.019	43	0.100	43
0.260	37	0.100	37	0.160	37
0.560	33	0.230	33	0.260	33
0.820	27	0.400	29	0.390	27
1.150	23	0.460	23	0.460	23
1.870	17	0.890	17	1.190	20
2.000	13	1.000	13	1.200	13

Bibliographie

- [1] Alfonso, Daniele. SVC low-complexity macroblock mode decision, doc. JVT-V079. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2007.
- [2] Amonou, I. and Cammas, N. and Kervadec, S. and Pateux, S. Optimized Rate-Distortion Extraction With Quality Layers in the Scalable Extension of H.264/AVC. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 17(9) :1186–1193, 2007.
- [3] Amonou, Isabelle and Pateux, Stéphane. Modification of the calculation of Scaling Factors in the JSVM, doc. JVT-O045. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2005.
- [4] Andreas Unterweger and Herbert Thoma. The influence of bitrate allocation to scalability layers on video quality in H.264 SVC. *Picture Coding Symposium (PCS)*, 2007.
- [5] Anselmo, T. and Alfonso, D. Buffer-based Constant Bit-rate Control for Scalable Video Coding. *Picture Coding Symposium (PCS)*, 2007.
- [6] Athanasios Leontaris and Alexis Michael Tourapis. Rate Control for the Joint Scalable Video Model, doc. JVT-W043. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2007.
- [7] Byung Cheol, Song and Kang Wook, Chun. A one-pass variable bit-rate video coding for storage media. *IEEE Transactions on Consumer Electronics*, 49(3) :689–692, 2003.
- [8] Ce, Zhu and Xiao, Lin et al. A novel hexagon-based search algorithm for fast block motion estimation. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 3 :1593–1596, 2001.
- [9] Chen, Hung-Wei and Yeh, Chia-Hung et al. Adaptive GOP structure determination in hierarchical B picture coding for the extension of H.264/AVC. *IEEE*

- International Conference on Communications, Circuits and Systems (ICCCAS)*, pages 697–701, 2008.
- [10] Chen, Zhenzhong and Ngan, King Ngi. Recent advances in rate control for video coding. *Image Communication Journal*, 22(1) :19–38, 2007.
- [11] Do-Kyoung Kwon and Mei-Yin Shen and Jay Kuo, C.-C. A Novel Two-Stage Rate Control Scheme for H.264. *IEEE International Conference on Multimedia and Expo (ICME)*, pages 673–676, 2006.
- [12] Do-Kyoung, Kwon and Mei-Yin, Shen and Kuo, C. C. J. A Novel Two-Stage Rate Control Scheme for H.264. *IEEE International Conference on Multimedia and Expo (ICME)*, pages 673–676, 2006.
- [13] Everett, Hugh. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research*, 11 :399–417, 1963.
- [14] Fabrice Urban. *Implantation optimisée d'estimateurs de mouvement pour la compression vidéo sur plates-formes hétérogènes multicomposants*. PhD thesis, IETR - Thomson R&D, Rennes, 2007.
- [15] Fan Yang and Qian Zhang and Wenwu Zhu and Ya-Qin Zhang. Bit allocation for scalable video streaming over mobile wireless Internet. *Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, 3 :2142–2151, 2004.
- [16] Feghali, R. and Speranza, F. and Wang, D. and Vincent, A. Video Quality Metric for Bit Rate Control via Joint Adjustment of Quantization and Frame Rate. *IEEE Transactions on Broadcasting*, 53(1) :441–446, 2007.
- [17] Gwo-Long, Li and Mei-Juan, Chen and Hung-Ju, Li and Ching-Ting, Hsu. Efficient search and mode prediction algorithms for motion estimation in H.264/AVC. *IEEE International Symposium on Circuits and Systems (ISCAS)*, 6 :5481–5484, 2005.
- [18] He, Z. and Chen, T. Linear rate control for JVT video coding. *International Conference on Information Technology : Research and Education (ITRE)*, pages 65–68, 2003.
- [19] He, Z. and Liang, Y. and Chen, L. and Ahmad, I. and Dapeng, W. Power-rate-distortion analysis for wireless video communication under energy constraints. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 15(5) :645–658, 2005.
- [20] He, Z. and Mitra, S. K. Rho-domain bit allocation and rate control for real time video coding. *IEEE International Conference on Image Processing (ICIP)*, 3 :546–549, 2001.

- [21] Heiko Schwarz and Detlev Marpe and Thomas Wiegand. Comparison of MCTF and closed-loop hierarchical B pictures, doc. JVT-P059. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2005.
- [22] Heiko Schwarz and Detlev Marpe and Thomas Wiegand. Hierarchical B pictures, doc. JVT-P014. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2005.
- [23] http://fr.wikipedia.org/wiki/Regression_lineaire. Régression linéaire.
- [24] http://ip.hhi.de/imagecom_G1/savce/downloads/. JSVM Reference Software. Version 8.6.
- [25] <http://sourceforge.net/projects/opensvcdecoder/>. Décodeur MPEG-4 SVC open source, 2006-2009.
- [26] <http://www.iso.org/iso/fr/>. Recommendation ISO/IEC JTC 1/SC 29 14496-10 : MPEG-4 Partie 10.
- [27] <http://www.iso.org/iso/fr/>. Recommendation ISO/IEC JTC 1/SC 29 14496-2 : MPEG-4 Partie 2.
- [28] <http://www.itu.int/rec/T-REC-H.264/f>. Recommendation H.264.
- [29] <http://www.jpeg.org/>. Standard de codage JPEG.
- [30] Hung-Ju, Lee and Tihao, Chiang and Ya-Qin, Zhang. Scalable rate control for MPEG-4 video. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 10(6) :878–894, 2000.
- [31] International Engineering Consortium. UMTS Protocols and Protocol Testing.
- [32] Ismaeil, I. R. and Docef, A. and Kossentini, F. and Ward, R. K. A computation-distortion optimized framework for efficient DCT-based video coding. *IEEE Transactions on Multimedia*, 3(3) :298–310, 2001.
- [33] ISO/IEC. JTC1/SC29/WG11/ N0400 - Test Model 5. Technical Report ISO/IEC JTC1/SC29/WG11/ N0400, MPEG- ISO/CEI JTC 1 SC 29/WG 11, 1993.
- [34] ISO/IEC. International Standard ISO/IEC 13818-2 - Information Technology- Generic Coding of Moving Pictures and associated audio information : Video (MPEG-2), 1996.
- [35] ISO/IEC JTC1/SC29/WG11 MPEG-2007/N9189. SVC Verification Test Plan, Version 1, 2007.
- [36] Jagmohan, A. and Ratakonda, K. MPEG-4 one-pass VBR rate control for digital storage. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 13(5) :447–452, 2003.

- [37] Jianfeng, X. and Yun, H. A novel rate control for H.264. *International Symposium on Circuits and Systems (ISCAS)*, 3 :809–12, 2004.
- [38] Jianning, Z. and Yuwen, H. and Shiqiang, Y. and Yuzhuo, Z. Performance and complexity joint optimization for H.264 video coding. *International Symposium on Circuits and Systems (ISCAS)*, 2 :888–891, 2003.
- [39] Julien Reichel and Heiko Schwarz and Mathias Wien. Joint Scalable Video Model JSVM-11, doc. JVT-X202. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2007.
- [40] Jun Xie and Liang-Tien Chia and Bu-Sung Lee. An improved distortion model for rate control of DCT-based video coding. *International Multi-Media Modelling Conference*, 2006.
- [41] Kamaci, N. and Altunbasak, Y. and Mersereau, R. M. Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 15(8) :994–1006, 2005.
- [42] Lambert, P. and De Neve, W. and De Neve, P. and Moerman, I. and Demeester, P. and Van de Walle, R. Rate-distortion performance of H.264/AVC compared to state-of-the-art video codecs. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 16(1) :134–140, 2006.
- [43] Li Liu and Xinhua Zhuang. A Novel Frame-Level Constant-Distortion Bit Allocation for Smooth H.264/AVC Video Quality. *SPIE Visual Communications and Image Processing (VCIP)*, 7257(1), 2009.
- [44] Li, Z. G. and Pan, F. and Lim, K. P. and Lin, X. and Rahardja, S. Adaptive rate control for H.264. *IEEE International Conference on Image Processing (ICIP)*, 2 :745–748, 2004.
- [45] Li, Zheng Guo and Yao, Wei and Rahardja, Susanto and Xie, Shoulie. New Framework for Encoder Optimization of Scalable Video Coding. *IEEE Workshop on Signal Processing Systems*, pages 527–532, 2007.
- [46] Lie, W.-N. and Chen.
- [47] Liu, Y. and Li, Z. G. and Soh, Y. C. A Novel Rate Control Scheme for Low Delay Video Communication of H.264/AVC Standard. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 17(1) :68–78, 2007.
- [48] Long Xu and Wen Gao and Xiangyang Ji and Debin Zhao. Rate Control for Hierarchical B-picture Coding with Scaling-factors. *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 49–52, 2007.

- [49] Lujun, Yuan and Wen, Gao and Yan, Lu. Latest arrival time leaky bucket for HRD constrained video coding. *IEEE International Conference on Multimedia and Expo (ICME)*, 2 :773–6, 2003.
- [50] Ma, Siwei and Gao, Wen and Gao, Peng and Lu, Yan. Rate control for advance video coding (AVC) standard. *International Symposium on Circuits and Systems (ISCAS)*, 2 :892–895, 2003.
- [51] Ma, Siwei and Gao, Wen and Wu, Feng and Lu, Yan. Rate control for JVT video coding scheme with HRD considerations. *IEEE International Conference on Image Processing (ICIP)*, 3 :793–6, 2003.
- [52] Maani, E. and Katsaggelos, A.K. Optimized Bit Extraction Using Distortion Estimation in the Scalable Extension of H.264/AVC. *IEEE International Symposium on Multimedia (ISM)*, pages 660–666, 2008.
- [53] Marpe, D. and Schwarz, H. and Wiegand, T. Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 13(7) :620–636, 2003.
- [54] Minqiang, Jiang and Nam, Ling. On enhancing H.264/AVC video rate control by PSNR-based frame complexity estimation. *IEEE Transactions on Consumer Electronics*, 51(1) :281–286, 2005.
- [55] Minqiang, Jiang and Xiaoquan, Yi and Nam, Ling. Frame layer bit allocation scheme for constant quality video. *IEEE International Conference on Multimedia and Expo (ICME)*, 2 :1055–1058, 2004.
- [56] Minqiang, Jiang and Xiaoquan, Yi and Nam, Ling. Improved frame-layer rate control for H.264 using MAD ratio. *International Symposium on Circuits and Systems (ISCAS)*, 3 :813–16, 2004.
- [57] Navakitkanok, P. and Aramvith, S. Improved rate control for advanced video coding (AVC) standard under low delay constraint. *International Conference on Information Technology : Coding and Computing (ITCC)*, 2() :664–668, 2004.
- [58] Overmeire, L. and Verdicchio, F. and Barbarien, J. and Schelkens, P. and Nachtergaele, L. Constant quality rate-control for video encoding based on activity segmentation. *IEEE International Conference on Image Processing (ICIP)*, 2 :1133–1136, 2004.
- [59] Ozcelebi, T. and Oguz Sunay, M. and Murat Tekalp, A. and Reha Civanlar, M. Cross-Layer Optimized Rate Adaptation and Scheduling for Multiple-User Wireless Video Streaming. *IEEE Journal on Selected Areas in Communications*, 25(4) :760–769, 2007.

- [60] Peng, Yin and Boyce, J. A new rate control scheme for H.264 video coding. *IEEE International Conference on Image Processing (ICIP)*, 1 :449–452, 2004.
- [61] Pietrowcew, A. and Buchowicz, A. and Skarbek, W. Bit-rate control algorithm based on local image complexity for video coding with ROI. *IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 582–587, 2005.
- [62] Pitrey, Y. and Babel, M. and Déforges, O. One-pass bitrate control for MPEG-4 Scalable Video Coding using rho-domain. *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2009.
- [63] Pitrey, Y. and Babel, M. and Déforges, O. One-Pass Rate Control Scheme using rho-domain for Scalable Video Coding. *Soumis à IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, Avril 2009.
- [64] Pitrey, Y. and Babel, M. and Déforges, O. and Viéron, J. Rho-domain based Rate Control Scheme for Spatial, Temporal and Quality Scalable Video Coding. *SPIE Visual Communications and Image Processing (VCIP)*.
- [65] Pitrey, Y. and Serrand, Y. and Babel, M. and Déforges, O. Rho-Domain for Low-Complexity Rate Control on MPEG-4 Scalable Video Coding. *IEEE International Symposium on Multimedia (ISM)*, pages 89–96, 2008.
- [66] Ramkishor, K. and Gupta, P. S. S. B. K. and Raghu, T. S. and Suman, K. Algorithmic optimizations for software-only MPEG-2 encoding. *IEEE Transactions on Consumer Electronics*, 50(1) :366–375, 2004.
- [67] Ribas-Corbera, J. and Shawmin, Lei. Rate control in DCT video coding for low-delay communications. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 9(1) :172–185, 1999.
- [68] Richardson, Iain E. H.264 and MPEG-4 Video Compression. John Wiley and Sons, 2003.
- [69] Roodaki, H. and Hashemi, M. R. and Fatemi, O. A Frame Layer Bit Allocation for H.264 Based on Frame Complexity. *Canadian Conference on Electrical and Computer Engineering*, pages 2025–2028, 2006.
- [70] Schwarz, H. and Marpe, D. and Wiegand, T. Overview of the Scalable H.264/MPEG-4-AVC Extension. *IEEE International Conference on Image Processing (ICIP)*, pages 161–164, 2006.
- [71] Shin, I.H. and Lee, Y.L. and Park, H.W. Rate control using linear rate-rho model for H.264. *Elsevier Signal Processing - Image Communication*, 4 :341–352, 2004.

- [72] Siwei Ma. Proposed draft description of rate control on JVT standard, doc. JVT-F086. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2002.
- [73] Siwei Ma and Zhengguo Li and Feng Wu. Adaptive Rate Control with HRD Consideration, doc. JVT-H014. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2003.
- [74] Séquences de référence de l'ITU-T. <http://www.itu.int/>.
- [75] Sullivan, G. J. and Wiegand, T. Rate-distortion optimization for video compression. *IEEE Signal Processing Magazine*, 15(6) :74–90, 1998.
- [76] Takamura, S. and Kobayashi, N. MPEG-2 one-pass variable bit rate control algorithm and its LSI implementation. *IEEE International Conference on Image Processing (ICIP)*, 1 :942–945, 2001.
- [77] Tang, Chih-Wei and Chen, Ching-Ho and Yu, Ya-Hui and Tsai, Chun-Jen. Visual sensitivity guided bit allocation for video coding. *IEEE Transactions on Multimedia*, 8(1) :11–18, 2006.
- [78] Wiegand, T. Joint Model Working Draft Number 2, doc. JVT-B118r2. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2002.
- [79] Wiegand, T. and Girod, B. Lagrange multiplier selection in hybrid video coder control. *IEEE International Conference on Image Processing (ICIP)*, 3 :542–545, 2001.
- [80] Wiegand, T. and Schwarz, H. and Joch, A. and Kossentini, F. and Sullivan, G. J. Rate-constrained coder control and comparison of video coding standards. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 13(7) :688–703, 2003.
- [81] Wiegand, T. and Sullivan, G. J. and Bjntegaard, G. and Luthra, A. Overview of the H.264/AVC video coding standard. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 13(7) :560–576, 2003.
- [82] Wien, M. and Schwarz, H. and Oelbaum, T. Performance Analysis of SVC. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 17(9) :1194–1203, 2007.
- [83] Wong, H.-M. and Au, O. and Chang, A. Fast mode decision and motion estimation, doc. JVT-N009. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2005.

- [84] Wu, Yuan and Shouxun, Lin and Yongdong, Zhang and Wen, Yuan and Haiyong, Luo. Optimum Bit Allocation and Rate Control for H.264/ AVC. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 16(6) :705–715, 2005.
- [85] Xu, L. and Gao, W. and Ji, X. and Zhao, D. and Ma, S. Rate Control for Spatial Scalable Video Coding in SVC. *Picture Coding Symposium (PCS)*, 2007.
- [86] Xu, L. and Ma, S. and Zhao, D. and Gao, W. Rate control for scalable video model. *SPIE Visual Communications and Image Processing (VCIP)*, 5960 :525–534, 2005.
- [87] Y. Liu and Z. G. Li and Y. C. Soh. Rate Control of H.264/ AVC Scalable Extension. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, 18(1) :116–121, 2008.
- [88] Yin, Ming and Wang, Hong-Yuan. A rate control scheme for H.264 video under low bandwidth channel. *Journal of Zhejiang University SCIENCE A*, pages 990–995, 2006.
- [89] Yuan Wu and Lin Shouxun and Zhang Yongdong and Luo Haiyong. Optimum Bit Allocation and Rate Control for H.264/AVC, doc. JVT-O016. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2005.
- [90] Yuanhai Zhang and Wei Huangfu and Jin Xu and Kaihui Li and Changqiao Xu. Integrated Rate Control and Buffer Management for Scalable Video Streaming. *IEEE International Conference on Multimedia and Expo (ICME)*, pages 248–251, 2007.
- [91] Z. Li and F. Pan and K. P. Lim and G. Feng and X. Lin and S. Rahardja. Adaptive Basic Unit Layer Rate Control for JVT, doc. JVT-G012. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2003.
- [92] Zhang, Dengzhi and Choo Chang, Y. Constrained variable bit rate control algorithm for MPEG-2 encoder. *SPIE Visual Communications and Image Processing (VCIP)*, pages 133–143, 2000.
- [93] Zhen Li and Tourapis, A.M. An Estimation-Theoretic Interpretation of Video Rate Distortion Optimization with Lagrangian Formulation. *Data Compression Conference (DCC)*, pages 222–231, 2008.
- [94] Zhengguo Li and Wen Gao et al. Proposed Draft of Adaptive Rate Control, doc. JVT-H017r1. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 2003.

Table des figures

1.1	Structure de l'encodeur MPEG-4 AVC/H.264.	18
1.2	Structure de GOP hiérarchique proposée dans MPEG-4 AVC/H.264. . .	23
1.3	Ordre d'encodage des images à l'intérieur d'un GOP hiérarchique.	24
2.1	Scalabilité spatiale dyadique	29
2.2	Scalabilité spatiale étendue	29
2.3	Scalabilité en qualité.	30
2.4	Scalabilité temporelle.	31
2.5	Structure de GOP hiérarchique utilisée pour la scalabilité combinée temporelle/qualité dans MPEG-4 SVC.	32
3.1	Variations du débit observées sans mécanisme de régulation (tracés obtenus en encodant un flux MPEG-4 AVC/H.264 composé exclusivement d'images P).	36
3.2	(a) relation entre le débit et le paramètre de quantification. (b) relation entre la qualité et le paramètre de quantification (tracés obtenus à partir d'une image P taille 352 * 288 codée avec MPEG-4 AVC/H.264).	45
3.3	Relation entre le débit et la qualité pour une image P de taille 352 * 288 encodée avec MPEG-4 AVC/H.264.	47
3.4	Relation entre ρ et le débit sur une image P encodée avec MPEG-4 AVC/H.264.	53
4.1	Séquences d'images utilisées pour la validation du ρ -domaine dans MPEG-4 SVC.	66
4.2	Débit en bits par pixel en fonction de ρ pour le scénario spatial	69
4.3	Débit en bits par pixel en fonction de ρ pour le scénario en qualité . . .	70

4.4	Débit en bits par pixel en fonction de ρ pour le scénario temporel (le symbole \emptyset signifie qu'il n'y a pas d'image de ce type dans la couche en question).	71
5.1	Décalage entre le modèle de débit et le débit effectif.	82
5.2	Performances du contrôle de débit pour le scénario spatial avec des GOP de 16 images sur la séquence HARBOUR	86
5.3	Performances du contrôle de débit pour le scénario spatial avec des GOP de 16 images sur la séquence HOCKEY	87
5.4	Performances du contrôle de débit pour le scénario en qualité avec des GOP de 16 images sur la séquence HARBOUR	88
5.5	Performances du contrôle de débit pour le scénario en qualité avec des GOP de 16 images sur la séquence SOCCER	89
5.6	Performances du contrôle de débit pour le scénario temporel avec des GOP de 16 images sur la séquence SOCCER	90
5.7	Performances du contrôle de débit pour le scénario temporel avec des GOP de 16 images sur la séquence HOCKEY	91
5.8	Comportement du mécanisme de contrôle de débit sur la séquence CREW avec différentes tailles de GOP.	92
5.9	Variations de la qualité des images dans les flux reconstruits avec le mécanisme de régulation de débit en deux passes.	93
6.1	Variations du PSNR observées pour la séquence HARBOUR avec la scalabilité en qualité.	96
6.2	(a) relation entre le PSNR et $1/(2^{q/6})$ pour différents types d'images dans un GOP hiérarchique de la séquence HARBOUR, avec des valeurs de QP comprises entre 28 et 36. (b) relation entre le débit et $1/(2^{q/6})$ pour la même configuration.	99
6.3	Améliorations apportées sur la qualité pour le scénario spatial sur la séquence HARBOUR	107
6.4	Améliorations apportées sur la qualité pour le scénario spatial sur la séquence SOCCER	108
6.5	Améliorations apportées sur la qualité pour le scénario qualité sur la séquence SOCCER	109
6.6	Améliorations apportées sur la qualité pour le scénario qualité sur la séquence CITY	110

6.7	Améliorations apportées sur la qualité pour le scénario temporel sur la séquence HARBOUR	111
6.8	Améliorations apportées sur la qualité pour le scénario temporel sur la séquence SOCCER	112
6.9	Influence de la prédiction inter-couches sur l'efficacité de codage des images pour la séquence HARBOUR	113
7.1	Dépendances entre les contenus d'images proches.	117
7.2	Performances du contrôle de débit en une passe pour le scénario spatial avec des GOP de 16 images sur la séquence HARBOUR	123
7.3	Performances du contrôle de débit en une passe pour le scénario spatial avec des GOP de 16 images sur la séquence SOCCER	124
7.4	Performances du contrôle de débit en une passe pour le scénario qualité avec des GOP de 16 images sur la séquence SOCCER	125
7.5	Performances du contrôle de débit en une passe pour le scénario qualité avec des GOP de 16 images sur la séquence CREW	126
7.6	Performances du contrôle de débit en une passe pour le scénario temporel avec des GOP de 16 images sur la séquence HARBOUR	127
7.7	Performances du contrôle de débit en une passe pour le scénario temporel avec des GOP de 16 images sur la séquence SOCCER	128
8.1	Variations de la qualité observées dans le scénario en qualité avec la méthode de répartition de budget présentée dans le chapitre 6.	132
8.2	Fonction d'évolution des poids d'image.	135
8.3	Valeurs de QP choisies pour les trois couches du scénario en qualité avec la séquence SOCCER	137
8.4	Relations entre les QP des différentes couches pour les scénarios spatial et en qualité.	140
A.1	Performances du contrôle de débit pour le scénario spatial avec des GOP de 16 images sur la séquence SOCCER	154
A.2	Performances du contrôle de débit pour le scénario spatial avec des GOP de 16 images sur la séquence CREW	155
A.3	Performances du contrôle de débit pour le scénario spatial avec des GOP de 16 images sur la séquence CITY	156
A.4	Performances du contrôle de débit pour le scénario en qualité avec des GOP de 16 images sur la séquence HOCKEY	157

A.5	Performances du contrôle de débit pour le scénario en qualité avec des GOP de 16 images sur la séquence CREW	158
A.6	Performances du contrôle de débit pour le scénario en qualité avec des GOP de 16 images sur la séquence CITY	159
A.7	Performances du contrôle de débit pour le scénario temporel avec des GOP de 16 images sur la séquence HARBOUR	160
A.8	Performances du contrôle de débit pour le scénario temporel avec des GOP de 16 images sur la séquence CREW	161
A.9	Performances du contrôle de débit pour le scénario temporel avec des GOP de 16 images sur la séquence CITY	162
B.1	Améliorations apportées sur la qualité pour le scénario spatial sur la séquence HOCKEY	164
B.2	Améliorations apportées sur la qualité pour le scénario spatial sur la séquence CREW	165
B.3	Améliorations apportées sur la qualité pour le scénario spatial sur la séquence CITY	166
B.4	Améliorations apportées sur la qualité pour le scénario qualité sur la séquence HARBOUR	167
B.5	Améliorations apportées sur la qualité pour le scénario qualité sur la séquence HOCKEY	168
B.6	Améliorations apportées sur la qualité pour le scénario qualité sur la séquence CREW	169
B.7	Améliorations apportées sur la qualité pour le scénario temporel sur la séquence HOCKEY	170
B.8	Améliorations apportées sur la qualité pour le scénario temporel sur la séquence CREW	171
B.9	Améliorations apportées sur la qualité pour le scénario temporel sur la séquence CITY	172
C.1	Performances du contrôle de débit en une passe pour le scénario spatial avec des GOP de 16 images sur la séquence HOCKEY	174
C.2	Performances du contrôle de débit en une passe pour le scénario spatial avec des GOP de 16 images sur la séquence CREW	175
C.3	Performances du contrôle de débit en une passe pour le scénario spatial avec des GOP de 16 images sur la séquence CITY	176

C.4	Performances du contrôle de débit en une passe pour le scénario qualité avec des GOP de 16 images sur la séquence HARBOUR	177
C.5	Performances du contrôle de débit en une passe pour le scénario qualité avec des GOP de 16 images sur la séquence SOCCER	178
C.6	Performances du contrôle de débit en une passe pour le scénario qualité avec des GOP de 16 images sur la séquence CITY	179
C.7	Performances du contrôle de débit en une passe pour le scénario temporel avec des GOP de 16 images sur la séquence HOCKEY	180
C.8	Performances du contrôle de débit en une passe pour le scénario temporel avec des GOP de 16 images sur la séquence CREW	181
C.9	Performances du contrôle de débit en une passe pour le scénario temporel avec des GOP de 16 images sur la séquence CITY	182

Abstract

This work is about designing bit rate control technics for the MPEG-4 Scalable Video Coding standard. The proposed approach benefits from a low computational complexity, so that its impact on the encoding time is very small. The quality of the encoded stream is also considered when controlling the bitrate in order to obtain optimal visual impression.

With today's heterogeneous communication technologies and video-reading devices, broadcasting video content requires large amounts of time and money to provide the user with optimal quality in any context. Scalable video coding has been developed as an answer to this need for adaptive video streams. In 2008, the MPEG-4 Scalable Video Coding (SVC) extension was finalized. To address the adaptivity issues, this new standard provides three types of scalabilities. Spatial, temporal and quality scalabilities allow a video stream to adapt the dimensions, the number of frames per second and the fidelity depending on the target's requirements.

Bit rate control is an important part of the encoding process as it allows to adapt the bitrate at the output of the encoder to the target's constraints. Unlike most of the existing approaches, our rate control technic is designed to have a low computational complexity. First, a bit rate per second constraint is specified for each video layer. Then, the available bits are dispatched among frames so that constant quality can be achieved in the decoded stream. The respect of the specified budget is finally enforced by a bit rate model, based on a simple and effective framework called ρ -domain. Using the statistics of the frame, this model allows us to choose the optimal quantization parameter while maintaining the amount of calculations very low.

Based on this simple technic, we propose two approaches. In the first approach, each image is pre-encoded to provide the bit rate model with the statistics of the frame before rate control. This approach allows us to regulate the bit rate with great accuracy, as the bit rate error is below 3% of the specified constraint on each type of scalability. In the second approach, rate control is performed using the information from the previous frame, so that no pre-encoding step is required and each frame is encoded only once. Thus, the impact on the encoding process is greatly reduced, at the cost of a small error increase. Finally, perceptual quality is considered when dispatching the bits among frames, so that quality variations are reduced and visual experience is improved.

Our bit rate control technic for MPEG-4 SVC shows great accuracy on all spatial, temporal and quality scalabilities. Its low complexity, together with quality variation control make it a valuable contribution, especially in practical applications, for which time resources are limited and the user-felt quality is important.

Résumé

Les travaux de cette thèse ont pour but de développer des stratégies de régulation de débit pour le codage vidéo scalable MPEG-4 SVC. Plusieurs approches sont proposées, en fonction des besoins en termes de précision et de la complexité désirée. La qualité du flux vidéo décodé est également prise en compte afin d'améliorer l'impression visuelle.

La multiplication des moyens de transmission et la diversification des appareils capables de lire du contenu vidéo contraignent les diffuseurs à dépenser beaucoup de temps et de moyens pour être en mesure de fournir une qualité de vidéo optimale quel que soit le contexte. La vidéo scalable a été développée en réponse à ce besoin d'adaptation des contenus vidéo aux différents contextes de diffusion. Le standard H.264/MPEG-4 SVC (Scalable Video Coding) propose trois types de scalabilité (spatiale, temporelle et en qualité), qui permettent d'adapter la résolution, le nombre d'images par seconde et la qualité du flux en fonction des besoins. Un flux vidéo unique est encodé, contenant plusieurs couches de résolution différentes codées les unes par rapport aux autres, de façon à rendre le codage de l'ensemble plus efficace.

La régulation de débit permet d'adapter le débit en sortie de l'encodeur pour respecter des contraintes liées à la transmission ou au décodage du flux vidéo. À partir d'une consigne donnée, le budget à respecter est réparti entre les différents éléments du flux. Un modèle de débit est ensuite utilisé pour anticiper le comportement du débit en fonction des paramètres d'encodage afin de respecter les contraintes imposées au flux.

À partir de cette problématique, deux approches sont proposées. La première se base sur un pré-encodage de chaque image pour fournir une base de calcul au modèle de débit. Elle permet d'obtenir une régulation très précise, avec une erreur entre le budget alloué et le débit effectif inférieure à 7% sur les trois types de scalabilité. La seconde approche utilise les informations rassemblées dans les images précédentes comme base pour le modèle de débit de l'image à encoder. Elle ne requiert pas de pré-encodage, et ne pénalise pas la complexité du processus d'encodage. En outre, la perte de performances par rapport à l'approche en deux passes est minimale et la consigne de débit par seconde est respectée de manière précise. Enfin, une méthode permettant de réduire les variations de la qualité est proposée pour améliorer l'impression visuelle ressentie par l'utilisateur.

Les résultats montrent que la méthode présentée est capable de réguler le débit avec une grande précision sur les trois types de scalabilité, tout en réduisant les variations de la qualité et en conservant une complexité de calculs très faible. Ces atouts la rendent non seulement intéressante du point de vue des performances, mais également applicable dans des contextes pratiques où les ressources en temps sont limitées.