

Efficient rate control strategies for scalable video coding Thibaud Biatek

To cite this version:

Thibaud Biatek. Efficient rate control strategies for scalable video coding. Signal and Image processing. INSA de Rennes, 2016. English. $NNT : 2016ISAR0007$. tel-01392735

HAL Id: tel-01392735 <https://theses.hal.science/tel-01392735v1>

Submitted on 4 Nov 2016

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

THESE INSA Rennes sous le sceau de l'Université Bretagne Loire pour obtenir le titre de DOCTEUR DE L'INSA RENNES Spécialité : Traitement du Signal et des Images présentée par

Thibaud Biatek

ECOLE DOCTORALE : *MATISSE* **LABORATOIRE :** *IETR*

Efficient Rate Control Strategies for Scalable Video Coding

Thèse soutenue le 25.04.2016 devant le jury composé de :

Christine Guillemot Directrice de Recherche à l'INRIA de Rennes (France) / Présidente **Marco Cagnazzo** Maitre de Conférences HDR à Telecom-ParisTech (France) / Rapporteur **Mathias Wien** Chercheur à l'Université de Aachen (Allemagne) / Rapporteur **Frédéric Dufaux** Directeur de Recherche au CNRS LTCI, Paris (France) / Examinateur **Patrick Le Callet** Professeur à Polytech Nantes (France) / Examinateur **Wassim Hamidouche** Maitre de Conférences à l'INSA de Rennes (France) / Co-encadrant **Jean-François Travers** Expert Technique Systèmes Audiovisuels à TDF (France) / Co-encadrant **Olivier Déforges** Professeur à l'INSA de Rennes (France) / Directeur de thèse

Stratégies d'Encodage pour Codeur Vidéo **Scalable**

- Efficient Rate Control Strategies for Scalable Video Coding

Thibaud Biatek

En partenariat avec

Document protégé par les droits d'auteur

Acknowledgements

First of all, I would like to thank one of my PhD supervisors Wassim Hamidouche. He has always been deeply involved, curious and critical about my work. He has always challenged me and helped me to be me more accurrate, meticulous, and ambitious with my research. I feel lucky to have been supervised by Wassim and I am convinced that I could not have achieved all this work without his valuable advice. I would also like to thank Jean-François Travers, my PhD supervisor at TDF. He was extensively helpful in bringing out the industrial relevance of my ideas and encouraged me to experiment my algorithms on pertinent use-cases, compliant with realistic broadcasting requirements. Our pleasant and interesting discussions at TDF were a rewarding experience, and I hope that our paths will cross again in the future. I also feel grateful to Mickael Raulet, who was my first contact in Rennes, and one of my supervisors during the first part of the thesis. His valuable advice helped me to publish and patent my first research results. Last but not least, this work is indebted to Olivier Déforges, my PhD surpervisor at the IETR. He has always been present to support and helped me in academic paper presentations. When I contemplate the progress made, I measure how my supervisors helped to be a better and more rigourous scientist today.

I also feel grateful for all my TDF co-workers, especially Imad Alawe, Frederic Beauchamp, Philippe Courboulay, Pape-Abdoulaye Fam and Amelie Kerbellec for their support.

I also would like to thank my jury: Marco Cagnazzo, Mathias Wien, Frédéric Dufaux, Christine Guillemot and Patrick Le Callet for accepting to evaluate my work. Their constructive feedback helped me improve the manuscript but also to point out new perspectives.

During over two years, I worked within the Institute of Research and Technology b<>com, I do wish to thank all my co-workers, particularly Cambodge Bist, Pierre Castel, Franck Chi, Nicolas Dhollande, Antonin Gilles, Victorien Lorcy, Charlène Mouton, David Pesce, Adrien Schadle and Fabrice Templon for all the time spent and for the discussions we had together wether it was in or outside work. I would also like to thank Jean-Yves Aubie, Daniele Cloatre and Xavier Ducloux for their interest in my research. All these people definitely made Bcom a really inspiring and stimulating place to work.

Finally, I would like to thank my parents Laurent and Marie-Pierre, my siblings Aurélien and Juliette and my fiancée Félicie for their invaluable support during these three years.

Résumé en français

Introduction

High Efficiency Video Coding, ou HEVC/H.265, est la dernière norme de compression vidéo dont la première version a été publiée en avril 2013. Cette nouvelle norme est le fruit d'un travail collaboratif mené par des experts de l'ITU et de l'ISO/IEC respectivement représenté par les groupes VCEG (Video Coding Experts Group) et MPEG (Moving Pictures Experts Group), formant ainsi JCT-VC (Joint Collaborative Team on Video Coding). L'objectif de HEVC est de remplacer le très populaire *Advanced Video Coding* (AVC/H.264) dont l'utilisation est très largement répandue, en permettant de réduire son débit de codage de 50% pour une qualité visuelle équivalente. Comme son prédécesseur AVC l'a permis pour l'introduction de la Haute-Définition (HD), le déploiement d'HEVC permettra l'introduction de nouveaux services plus couteux en débit tel que l'Ultra-Haute-Définition (UHD).

Actuellement, de nombreux réseaux servent de support à la transmission de contenus vidéo. Les diffuseurs historiques de la télévision numérique terrestre (TNT) perdent leurs fréquences au profit des opérateurs de téléphonie mobile. En France cette réduction du spectre hertzien s'est opéré via deux dividendes numérique au début des années 2000 puis en 2015. En plus du mobile et de la TNT, le réseau internet est devenu un support majeur de diffusion de contenus via les services over-the-top (OTT), la vidéo devrait représenter 80% du trafic en 2019. En plus de ces aspects réseaux, les supports de consommation des contenus vidéo se multiplient drastiquement. La télévision, les smartphones, tablettes, casques de réalités virtuelles et montre connectées sont les nouveaux supports et sont caractérisés par une multitude de formats différents (résolution, gamut de couleurs, dynamique, etc ...).

Cette diversité implique une lourde redondance des flux encodés à de nombreux débits, résolution, qualités, et donc un coût supplémentaire selon plusieurs aspects. Premièrement pour la diffusion OTT, un même contenu est encodé de très nombreuses fois en de très nombreuses versions et à différents débits ce qui accroit les besoins de stockage. Du point de

vue de la TNT, l'introduction de nouveaux formats multimédia est envisageable seulement en assurant une compatibilité avec le parc d'équipements existant. Cette compatibilité est très souvent assurée en opérant une diffusion dite simulcast, où toutes les versions d'un contenu sont transmises dans le canal, le récepteur décodant seulement la version qui le concerne. La compression vidéo scalable est une façon d'éviter tous ces inconvénients, et permet d'encoder un contenu en une couche de base et plusieurs couches d'améliorations de formats différents. Dans ce schéma, la compression est significativement plus efficace que le SimulCast puisque chaque couche est encodée en exploitant des informations provenant des couches inférieures.

La norme *Scalable High Efficiency Video Coding*, ou SHVC, a été introduite dans les Annexes F et H d'HEVC et publié en octobre 2014 dans la version 2 de HEVC. SHVC couvre plusieurs types de scalabilité, spatiale, en qualité, en gamut de couleur et en codec en plus du mécanisme de scalabilité temporelle déjà inclus dans la première version d'HEVC. Ces types de scalabilité peuvent être combinés entre eux et peuvent résoudre les problématiques de diffusion actuelles. Parmis ces problèmes, l'introduction de nouveaux services UHD rétrocompatibles avec le parc existant via les scalabilité HD/UHD, BT.709 vers BT.2020, SDR vers HDR. Cette technologie est déjà considérée dans plusieurs groupes de normalisation tel que l'*Advanced Television Systems Committee* (ATSC) et le consortium *Digital Video Broadcasting* (DVB). Dans la norme ATSC3.0, SHVC est la technologie retenue pour la diffusion de programme scalable (limitée à deux couches spatiales). Du côté DVB la scalabilité temporelle est déjà retenue dans la norme ETSI TS-101-154 pour la diffusion de programme UHD, et SHVC est considéré comme candidat pour la diffusion rétro-compatible des futurs formats en cours de normalisation.

Pour la diffusion de services scalables dans un environnement industriel, le contrôle de débit sur les flux générés est un enjeu majeur. Pour cette raison, l'objectif de cette thèse a été le développement d'algorithmes de régulation de débit pour SHVC. Dans un premier volet, nous allons explorer des algorithmes d'estimation du paramètre de quantification (QP) basée sur l'approche ρ-domaine afin d'atteindre un débit cible lors de l'encodage. Deuxièmement, nous développerons des algorithmes d'allocation de débit adaptatifs pour SHVC qui aurons pour but d'ajuster le ratio de débit entre les couches de façon à optimiser les performances de compression. Finalement, nous investiguerons le multiplexage statistique de programmes SHVC encodés dans le contexte d'introduction de services UHD.

Etat de l'art

Dans le Chapitre 2, les principes de compression vidéo ainsi qu'HEVC, ses outils et son extension scalable SHVC sont introduits ainsi qu'un aperçu des logiciels de référence utilisés en normalisation. HEVC, comme AVC avant lui, est basé sur un schéma de codage hybride par bloc. Les images en entrée du codeur sont découpées en blocs appelés *Coding Tree Unit* (CTU) et employant une représentation en arbre de sous-blocs *Coding Unit* (CU). Chacun de ces CU est partitionné en blocs sur lesquels la prédiction est réalisée (*Prediction Units* ou PU) et d'autres adaptés aux étapes de transformation et quantification (*Transform Units* ou TU). Une fois la meilleure prédiction réalisée (spatialement ou temporellement) sur les différents PU, la texture prédite est soustraite à l'originale pour former le résidu de prédiction. Ce résidu est ensuite transformé et quantifié selon la structure de TU choisie et les coefficients obtenus sont encodés dans le flux binaire. En plus de ces coefficients, toutes les autres informations nécessaires lors du décodage (type de prédiction, images de référence, etc ...) sont également transmises. Pour SHVC, le même coeur de codage qu'HEVC est conservé, et l'implémentation de cette technologie ne nécessite que des changements haut-niveau. Un module de traitement est placé entre les couches afin de rendre le signal reconstruit d'une couche *L* interprétable par le codeur de la couche *L*+1 comme référence lors de la prédiction. Par exemple, un sur-échantillonnage est réalisé sur le signal reconstruit pour la scalabilité spatiale alors qu'une table de correspondance assure le changement d'espace colorimétrique pour la scalabilité en gamut de couleur. De plus, les vecteurs de mouvements provenant de la couche de niveau le plus bas sont mis à l'échelle pour être utilisés dans le codage de la couche de niveau le plus haut. Ces choix architecturaux devraient permettre une large adoption industrielle compte tenu du faible coût nécessaire pour faire évoluer les implémentations HEVC en SHVC.

Le Chapitre 3 est consacré à l'état de l'art sur les algorithmes de régulations de débit employés en compression vidéo, particulièrement pour HEVC et SHVC. Généralement, les algorithmes de régulation de débits sont constitués de deux parties. La première partie consiste à allouer du débit dans la séquence afin d'atteindre finalement un débit correspondant à la consigne d'entrée. Selon la granularité souhaitée, l'allocation de débit peut se faire au niveau groupe d'images (GOP), images ou bien CTU. Dans certaines contributions, l'étape d'allocation de débit inclut la modélisation de la mémoire tampon du décodeur afin d'éviter un sur-débit ne permettant plus un décodage correct du flux. De nombreuses approches d'allocations de débit ont été explorées dans la littérature, par exemple basées sur la théorie des jeux, la classification des images ou encore l'analyse des contenus. La seconde étape de la régulation de débit consiste à estimer les paramètres d'encodages permettant d'atteindre

le débit ciblé lors de la première étape, pour une granularité donnée. L'approche classique consiste à modéliser la relation entre le débit *R* et un paramètre d'encodage contrôlable, par exemple le QP. D'autres approches existent comme le ρ-domaine qui modélise linéairement la relation entre le débit et le taux de coefficients nuls ρ dans le résidu transformé et quantifié, la relation ρ-QP étant ensuite établie aisément par une table de correspondance. Plus récemment, le modèle R-λ reposant sur l'ajustement du multiplicateur lagrangien λ utilisé dans la boucle d'optimisation débit-distorsion a été utilisé efficacement pour HEVC.

Au regard de cet état de l'art, il apparait que plusieurs pistes sont possibles pour améliorer la régulation de débit. L'approche ρ-domaine a démontré son efficacité pour l'extension scalable de AVC *Scalable Video Coding*, (SVC), mais dont le schéma de compression était différent de SHVC. Nous explorerons donc premièrement les approches ρ-domaine pour la compression vidéo HEVC/SHVC. La compression scalable repose sur l'emploi d'informations inter-couches pour réaliser un codage plus performance dans la couche de niveau supérieur, ce qui implique que le ratio de débit utilisé lors d'un codage SHVC doit avoir un certain impact sur les performances de compression. Nous investiguerons donc dans un second temps l'impact du ratio de débit sur les performances, afin de développer des algorithmes de régulation de débit adaptatifs basés sur l'estimation de ratio optimal. Enfin, le multiplexage statistique de plusieurs programmes scalable sera indispensable dans l'optique d'un déploiement de services SHVC. Dans la littérature existante, le cas de la scalabilité en grain fin (FGS) de SVC a été étudiée. Toutefois, ce type de scalabilité n'est pas compatible avec l'introduction de nouveaux services, et était plutôt utilisé pour régler aisément le débit d'un flux binaire sans utiliser de transcodage. Pour ces raisons, le multiplexage statistique sera également exploré comme dernier axe de recherche.

Contributions

Outils de régulation de débit basés ρ-domaine

Dans le Chapitre 4, le premier axe de recherche consacré au développement d'outils basé ρ-domaine est décrit. La première étape est dédiée à une étude statistique poussée sur le ρ-domain appliqué à HEVC et SHVC ainsi qu'à la répartition du débit dans les flux binaires générés par les logiciels de référence. Une des limitations de l'approche ρ-domaine pour HEVC est que ce nouveau standard, étant plus efficace, a réduit considérablement la part de résidu à compresser et a donc augmenté la part de débit consacré aux informations de prédiction. L'approche basée sur l'ajustement du multiplicateur lagrangien utilisé dans la

boucle d'optimisation débit-distorsion est certainement plus optimale au sens débit-distorsion, mais tous les codeurs ne fonctionnent pas sur ce paradigme. Intuitivement, ce postulat semble correct et les simulations menées lors de cette thèse l'ont démontré. Toutefois, la modélisation du ρ-domaine fournie dans notre étude confirme bien que cette approche peut être utilisée sans pertes de précision, que l'on considère une modélisation sur le débit généré par le résidu ou sur le débit global (résidu + informations de prédiction).

La première approche que nous avons développée est un schéma d'estimation de QP au niveau CTU qui permet de sélectionner le QP permettant d'atteindre une consigne de débit donné. Dans cette approche, l'encodeur garde en mémoire pour chaque CTU certaines informations tels que les tailles de transformés utilisées, les distributions statistiques des coefficients contenus dans le résidu, la modélisation ρ-domaine calculée à posteriori et la profondeur du CTU. Lorsqu'un CTU est encodé, sa configuration la plus probable est dérivée par rapport aux CTU voisins. Puis, le débit cible pour ce CTU combiné à la modélisation ρ-domaine permet d'estimer le ρ à atteindre. La distribution statistique des coefficients la plus probable est utilisée combinée au ρ cible pour déterminer la valeur de frontière de la zone morte (ou DZB, valeur à partir de laquelle les coefficients sont mis à zéro après la quantification) qui devra être atteinte lors de la quantification. L'équation de quantification-inverse contenue dans la norme est dont inversée pour déterminer le QP permettant d'atteindre cette valeur précise de DZB. Cet algorithme a été évalué pour HEVC et SHVC où les débits par CTU générés lors des encodages de référence à QP fixe ont été choisis comme cible. Notre algorithme a permis d'atteindre une très bonne précision de débit (<4%) sans détérioration significative de PSNR pour HEVC et SHVC. Cette méthode est particulièrement avantageuse puisqu'elle permet d'opérer une régulation de débit précise à granularité fine, avec une bonne précision et surtout en une seule passe d'encodage.

La deuxième approche développée est une méthode d'estimation déterministe du modèle ρ-domaine. Lors de notre étude statistique, nous avons constaté que la pente du ρ-domaine se localise dans un intervalle de valeur réduit pour une résolution spatiale donnée. En modélisant cette valeur de pente via une méthode des moindres carrés totaux, nous pouvons constater que la valeur de pente est linéairement liée au nombre de pixels contenus dans une image (i.e résolution). Afin de proposer une formulation analytique de cette relation, nous proposons une modélisation plus précise du ρ -domaine au niveau résidu (*Debit* = 4×*N* ² ×(1−ρ) +∆, niveau résidu avec *N* la taille du résidu). Puis nous sommons ce modèle au niveau composante (Y, U, V) puis au niveau frame afin d'obtenir une modélisation plus réaliste ($Debit = 4 \times S \times (1 - \rho) + \Delta'$, au niveau image avec *S* la surface en

nombre d'échantillons). Cette première relation, bien que précise pour les hauts débits, l'est un peu moins pour les bas débits où tous les pixels de l'image ne sont pas nécessairement compressés via un résidu. Afin de pallier à ce problème nous avons introduit le taux de résidus non-codés (ou skipés) dans l'image à notre équation afin d'avoir une estimation plus précise de la pente, définie comme:

$$
Debit = 4 \times S \times (1 - r_{skip}) \times (1 - \rho) + \Delta'
$$
 (1)

Avec *rSkip* le taux de blocs skipés dans l'image. La précision de cette approche a été mesurée sur un ensemble de données HEVC et SHVC et a montré une précision d'estimation de la pente supérieure à 90%. De plus, nous avons proposé une architecture se prêtant à l'utilisation de cette méthode et destinée aux schémas de régulation de débit ρ-domaine. Cette méthode peut être intégrée aux unités de contrôle de débit ρ-domaine existantes.

Régulation de débit adaptative pour SHVC

Dans le Chapitre 5, le second axe de recherche dédié au développement d'algorithmes de régulation de débit adaptif pour SHVC est décrit. On constate dan la littérature existante, ainsi que dans les travaux de normalisation, que le ratio de débit entre les couches d'un encodeur SHVC est fixé de façon totalement arbitraire. Cela pose problème, particulièrement lorsqu'on sait que les performances de compression dépendent de l'efficacité des prédictions inter-couches. Pour ces différentes raisons, nous avons étudiés l'impact du ratio de débit d'un schéma scalable à deux couches sur les performances de compression, le ratio étant définit comme:

$$
\tau = \frac{R_{BL}}{R_G} \tag{2}
$$

Avec *RBL* le débit de la couche de base et *R^G* le débit global (BL+EL). Afin d'évaluer les performances d'un ratio donné à l'aide des métriques de Bjøntegaard (BD-BR et BD-PSNR), nous devons avoir pour ce ratio des encodages à quatre débits globaux $R_G \in \{R_1, R_2, R_3, R_4\}$. Nous sommes donc en mesure pour chaque ratio testé $\tau \in \{0.1, 0.2, ..., 0.8, 0.9\}$ d'évaluer les performances du schéma scalable. Cette étude est réalisée sur la scalabilité spatiale, en espace colorimétrique, en dynamique ainsi que sur plusieurs combinaisons de ces scalabilités. D'après les résultats de cette étude, il apparait que le ratio de débit a un impact majeur sur les performances de compression et que l'ajustement du ratio optimal permet de réduire significativement le surcout introduit par la scalabilité par rapport à l'encodage simple-couche du niveau le plus haut. Cette étude préliminaire a mis en avant le gain atteignable en utilisant une approche de compression à ratio variable sous contrainte d'un débit global constant. Nous avons donc investigué deux approches, premièrement l'optimisation de l'encodage dans un intervalle de ratio fixe et deuxièmement l'ajustement optimal de ratio sous contraintes de débit et qualité dans chacune des couches.

Le premier algorithme développé a pour but d'estimer dynamiquement lors de l'encodage le meilleur ratio possible dans un intervalle Φ_G fixe et constant (par exemple $\Phi_G = [0.2, 0.4]$). Afin de nous intégrer parfaitement à l'architecture de régulation existante dans le logiciel de référence, nous avons substitué l'allocation de débit au niveau GOP par une allocation jointe des deux couches. Dans une première étape, l'algorithme alloue un débit global à répartir entre les deux couches pour le GOP courant, en adéquation avec la consigne de débit global. Puis, le ratio optimal à appliquer au GOP est déterminé en résolvant le problème d'optimisation suivant:

$$
\tau_{opt} = \underset{\tau \in \Phi_G}{\arg \max} \ G(\tau) \tag{3}
$$

Ou τ_{opt} est le ratio optimal, et *G*(τ) est une fonction qui évalue les performances de compression pour un ratio donné. Dans le cas de la mesure de performance d'un schéma scalable comparativement à l'encodage simple de la couche de niveau le plus haut, l'optimisation des performances est équivalente à l'optimisation de la qualité dans la couche d'amélioration. La fonction $G(\tau)$ est donc approximée dans cette approche par la fonction reliant τ le ratio mesuré à la distorsion atteinte dans la couche d'amélioration. Une modélisation linéaire de cette fonction est simple et efficace quand il s'agit d'un intervalle constant. Ce modèle peut être estimé de plusieurs façons lors de l'encodage, par exemple en maintenant une table de correspondance distorsion-ratio sur laquelle l'estimation est réalisée. Ce premier algorithme a été évalué pour le cas de la scalabilité spatiale HD/UHD dans l'intervalle [0.26,0.43] sur un ensemble de dix séquences. En moyenne, cette approche a permis de réduire significativement le surcoût introduit par SHVC, le faisant passer de 20% à 16%.

Afin d'améliorer l'algorithme précédemment développé et de le rendre plus compatible avec des cas d'usages industriels, nous lui avons adjoint des contraintes de qualité et de débit par couche, toujours sous consigne d'un débit global constant. On peut légitimement imaginer dans le cas d'une diffusion hybride où les deux couches sont transmises dans des canaux séparés qu'une qualité minimale est souhaitable sur la couche de base et que la couche d'amélioration doit tenir dans un canal de capacité fixe. Dans ce cas on aurait alors une qualité minimale imposée dans la couche de base et un intervalle de débit à respecter dans la couche d'amélioration. Plus généralement, la recherche du ratio optimal dans cette seconde approche revient à la résolution du même problème d'optimisation, mais ou l'intervalle de

recherche est calculé de la façon suivante:

$$
\Phi_G = \begin{cases} \Phi_{BL} , si \Phi_{BL} \cap \Phi_{EL} = \emptyset \\ \Phi_{BL} \cap \Phi_{EL} , sinon \end{cases}
$$
 (4)

Ou Φ_{BL} et Φ_{EL} sont respectivement les intervalles de ratio permettant le respect des consignes de débits et qualités dans les deux couches de base et d'amélioration. On notera que dans le cas où toutes les contraintes ne peuvent être respecté, on peut favoriser une couche ou l'autre (dans ce cas précis on favorise la couche de base). Cette approche nécessite la mise en oeuvre d'un modèle débit-distorsion dans chaque couche afin de calculer les sous-intervalles Φ*BL* et Φ*EL*. Cette approche a été évaluée dans plusieurs cas d'usages de diffusion hybride, pour la scalabilité HD-BT.709/UHD-BT.2020, HD-SDR/HD-HDR et UHD-SDR/UHD-HDR avec plusieurs jeux de contraintes adaptés à ces cas. En moyenne, cette approche a permis d'optimiser les performances de compression en apportant un gain BD-BR de -7.51% pour l'HD/UHD et de -8.30% pour le SDR/HDR. Ces algorithmes de régulation de débit adaptatifs sont avantageux puisqu'ils sont entièrement configurables et se prêtent donc à de multiples cas d'usages où ils favorisent une utilisation optimale de SHVC en termes de performances et supportent le déploiement de nouveaux services grâce à l'utilisation du standard SHVC.

Multiplexage statistique de programmes SHVC

L'optimisation des débits de serait pas complète sans une investigation complémentaire sur la technique de multiplexage statistique, dont les premiers travaux sont décrits dans le Chapitre 6. Pour commencer, nous nous plaçons dans le cas ou *N^P* programmes scalables à deux couches doivent être envoyés dans deux canaux différents destinés aux récepteurs fixes et mobiles. Le premier canal de capacité *R^M* est destinée principalement aux récepteurs mobiles, les couches de bases sont multiplexées pour être transmises aux récepteurs mobiles et fixe (service de base). Le second canal de capacité R_{TNT} est dédié au réseau de diffusion traditionnels TNT, où les couches d'améliorations sont multiplexées à destination des récepteurs fixes (en complément de la couche de base). Cette approche permet une mutualisation des réseaux et évite la diffusion SimulCast sur la couche TNT.

Notre première approche consiste à fixer tout d'abord les débits globaux à attribuer au programmes, puis a fixer le ratio des différents programmes de la façon la plus optimale possible. Considérons τ*^p* et *R^p* comme étant respectivement le ratio et le débit global attribué au *p ieme* programme. Notre algorithme d'optimisation doit respecter la contrainte de débit global et trouver la meilleure combinaison de ratios pour générer un flux compatible avec la capacité du réseau mobile:

$$
\sum_{p=1}^{N_p} R_p \le R_M + R_{TNT} \tag{5}
$$

$$
\sum_{p=1}^{N_p} \tau_p \times R_p \le R_M \tag{6}
$$

Le multiplexage statistique doit également permettre une qualité homogène des programmes et limiter les variations de qualité dans chaque programme. Ces aspects sont d'autres paramètres que notre algorithme doit prendre en compte lors du multiplexage.

Pour mesurer les performances du multiplexage statistique de façon simple et rapide, un algorithme reposant sur l'utilisation d'une base de données est proposé. La base de données contient différents flux scalable encodés à différents débit globaux et différents ratios. Chaque flux est découpé en morceaux indépendamment décodables auxquels sont attachés des informations sur le débit et la qualité des deux couches. L'algorithme réalise un assemblage de chunks optimisés, tout en respectant les contraintes précédemment citées, formant ainsi un flux binaire multiplexé favorisant les gains en compression et compatible avec une diffusion sur un réseau fixe/mobile. Ce type d'algorithme pourrait être notamment utilisé dans le cadre de la diffusion terrestre à partir de serveurs OTT. L'algorithme a été implémenté dans cette optique et a montré des performances intéressantes, autant d'un point de vue performance où le surcout introduit par la scalabilité est très faible que d'un point de vue qualité de service en minimisant les variations de façon très satisfaisante.

Conclusion

Dans cette thèse, nous avons exploré et proposé plusieurs méthodes originales relatives à la régulation et à l'allocation optimale de débit, appliquées à la compression vidéo scalable SHVC. Les algorithmes proposés dans le Chapitre 4 peuvent être intégrés pour améliorer la régulation de débit d'encodeurs existants. Dans le Chapitre 5, un algorithme de régulation de débit adaptatif pour SHVC est proposé pour faciliter l'introduction de services UHD scalables. Enfin, le Chapitre 6 introduit une méthode de multiplexage statistique basé sur l'optimisation du ratio de débit entre couches, destiné au déploiement de services hybrides en réception fixe/mobile sur des réseaux de diffusion hybrides. Ces approches originales ont montré des performances et des gains significatifs avec plusieurs perspectives d'amélioration. Enfin, ces travaux de recherches ont été valorisés par des publications lors de conférences internationale, dépots de brevets et contribution en normalisation.

Les deux outils développés dans le Chapitre 4 pourraient être fusionnés. L'estimation déterministe de la pente du ρ-domaine pourrait être adapté au niveau CTU et intégrée dans l'algorithme d'estimation de QP pour améliorer sa précision. Deuxièmement, des outils intercouches pourraient être investigués où des paramètres estimés dans la couche de rehaussement pourraient l'être en s'aidant des paramètres de la couche de base. L'implémentation actuelle dans le logiciel de référence duplique les algorithmes pour chaque couche, et l'introduction d'outils inter-couche devrais intuitivement améliorer les performances.

Le travail exposé dans le Chapitre 5 pourrais également être étendu à d'autres aspects comme la granularité qui pourrait être plus fine et testée au niveau image plutôt qu'au niveau GOP, dans l'objectif d'améliorer les performances débit-distorsion. Par ailleurs, notre algorithme a été testé sur des contenu HDR, mais uniquement évalué par des métriques objectives qui sont sujetes à controverse au sein de MPEG. Une évaluation subjective via un protocole normalisé permettrait d'avoir une autre appréciation des performances de notre méthode. Enfin, l'étude d'autres types de scalabilités pourraient être inclus à ces travaux tels que la scalabilité en qualité ou bien en codec (couche de base AVC et couche d'amélioration HEVC).

Le travail initié dans le Chapitre 6 a ouvert de nombreuses pistes pour la poursuite de ces travaux. L'implémentation actuelle est basée sur l'utilisation d'une base de données et est nécessite donc un encodage offline. Une première piste à explorer pourrait être le développement d'outils et de modèles plus appropriés à l'encodage de contenus en temps-réel. Parmi les autres pistes, d'autres modèles mathématiques de résolutions de problèmes d'optimisation pourraient être investigués puisque notre modèle relève d'un problème multi-variable joint entre le choix des débits globaux et des ratios sous contraintes de qualité.

Pour conclure, cette thèse nous a permis de proposer de nouveaux outils innovants de contrôle de débit pour les schémas de compression scalables. Ces outils nous ont permis de démontrer que l'introduction du format de codage SHVC pour de nouveaux services scalable est tout à fait réaliste, tant sur le plan de sperformances que de la complexité et avec un surcoût maitrisé.

Table of contents

IV Conclusion 139

V Appendix 147

List of figures

List of tables

A.1 Dataset used for the study on the bitrate ratio impact over SHVC performance159

Acronyms

- HDTV High-Definition Television
- HFR High Frame Rate
- HLS High Level Syntax
- HM HEVC Test Model
- HVS Human Visual System
- IDR Instantaneous Decoding Refresh
- ILR Inter-Layer Reference
- IRAP Intra Random Access Picture
- ITU International Telecommunication Union
- ISO International Organization for Standardization
- **IEC** International Electrotechnical Commision
- JCT-VC Joint Collaborative Team on Video Coding
- JM Joint Model (AVC Reference Software)
- JTC Joint Technical Committee
- JVT Joint Video Team
- KTA Key Technical Area
- LD Low-Delay
- LDB Low-Delay B
- LDP Low-Delay P
- LUT1D 1D Look-Up-Table
- LUT3D 3D Look-Up-Table
- MAD Mean Absolute Difference
- MPEG Moving Picture Experts Group
- MSE Mean Square Error

SDR Standard Dynamic Range

- SDTV Standard-Definition Television
- SE Syntax Element
- SEI Supplemental Enhancement Information
- SHM SHVC Test Model
- SHVC Scalable High Efficiency Video Coding
- SL Single-Layer
- SMX Statistical Multiplexing
- **SNR** Signal to Noise Ratio
- SPS Sequence Parameter Set
- SSE Sum of Square Error
- SSIM Structural Similarity Index
- StatMux Statistical Multiplexing
- SVC Scalable Video Coding
- SW Smoothing Window
- TB Transform Block
- Telco Telephone Company
- TMuC Test Model under Consideration
- TMVP Temporal Motion Vector Predictor
- TSB Transform Sub-Block
- TU Transform Unit
- UHDTV Ultra-High-Definition Telivision
- UHF Ultra High Frequency
- VBR Variable Bitrate
- VBV Virtual Buffer Verifier

Part I

Introduction
Chapter 1

Introduction

1.1 Preamble

High Efficiency Video Coding (HEVC) [1, 2]– is the latest video coding standard which has been jointly standardized by the International Telecommunication Union (ITU); the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), under the Joint Collaborative Team on Video Coding (JCT-VC). This team is composed of experts from the Video Coding Experts Group (VCEG) and the Moving Picture Experts Group (MPEG) which represent the ITU and the ISO/IEC, respectively. Released in February 2013, HEVC was designed to replace the successful Advanced Video Coding (AVC [3, 4]) standard, with the target of reducing the bitrate by 50% for the same visual quality.

The JCT-VC also developed optional extensions on the top of HEVC, tackling specific applications and services such as scalable, multi-view, range-extension, 3D and screen content coding extensions. Among these extensions, the scalable one –called Scalable High Efficiency Video Coding (SHVC)– enables layered representation, which can simultaneously address multiple targets while enabling backward compatibility. SHVC extension supports temporal, spatial, quality, bitdepth and color gamut scalability. Moreover, scalable coding enables substantial gains compared to separate encoding of each layer –called simulcast encoding–, by exploiting additional correlation between layers while encoding. These new standards have been designed to efficiently help the industry to switch from High Definition (HDTV) to Ultra High Definition (UHDTV) content delivery. Indeed, the UHDTV market has grown over the last few years under the pressure of display manufacturers that offer new challenges to broadcasters and content providers. This work falls within this context and aims at addressing the challenges encountered by broadcasters in this new area.

1.2 Context and stakes

During the last decade, three main trends can be observed. First of all, the spectrum allocated to Digital Terrestrial Television (DTT) has been continuously reduced in favor of Telco services, with the help of analog switch-off and mobile telecommunications technology development. In France, the second digital dividend was ratified in January 2015 and involved the transfer of 700 MHz ultra high frequency (UHF) channel to the up-coming 5G Telco services. Second, the multiplicity of networks such as terrestrial, satellite or broadband and the appearance of numerous content providers involve a boom in video content broadcast over these networks. The major internet actors predict that 80% of the internet traffic in the world will be video in 2019 [5]. Third, the advent of mobile devices such as smartphones, tablets and smart-watches in addition to the progress in display technology, involve a multiplication of format and a variety of users that can be addressed and who are potential video consumers.

All these elements imply new challenges and stakes for content providers, broadcasters and even device manufacturers. The huge bandwidth requirement over all the networks forces these actors to redesign the way of delivering contents to avoid potential network congestion. In addition, many users consuming the same content on a wide variety of devices have to be addressed which involves many encodings of the same content at different bitrates, resolutions, frame-rate, color-gamut and even codecs. The fast evolution in this area and the coming up of new formats such as UHDTV or new efficient codecs such as HEVC force the broadcasters to address these new markets while continuing to insure services to already present legacy receivers. These reasons imply a strong constraint on backward-compatibility among solutions of different generations which is a real challenge for broadcasters.

1.3 Objectives and motivations

The aim of this thesis is to propose new solutions to answer these challenges. Indeed, SHVC is a really promising solution to backward compatibility issues encountered by broadcasters or content providers. Several standardization bodies such as the Advanced Television Systems Committee (ATSC) or the Digital Video Broadcasting consortium (DVB [6]) consider SHVC as a candidate for solving compatibility issues brought by new formats introduction such as UHDTV, High-Frame-Rate (HFR), High-Dynamic-Range (HDR) or even Wide-Color-Gamut (WCG) [7]. In addition, scalability may save substantial bandwidth and thus could be an answer to the content-delivery issues such as network congestion, storage and low delay.

The work of this thesis will focus on designing new efficient rate control algorithms for both HEVC standard and its scalable extension. First, this work will explore and validate the concept of ρ −*domain*, introduced for rate control in the previous generation of video coding standard but has not been explored yet in SHVC. The ρ -domain will be investigated with the objective of designing accurate and low-complex innovative rate control algorithm in the context of HEVC et SHVC. Second, adaptive rate-control algorithm will be explored through studying the impact of bitrate ratio between layers over SHVC performance. Furthermore, an innovative adaptive rate-control algorithm will be proposed based on the preliminary study. Finally, the statistical multiplexing of scalable programs will be investigated followed by a first proposed approach.

1.4 Outline

Chapter 2 is dedicated to the state-of-the art on video coding, and especially the introduction of the related standards. Section 2.1 firstly clarifies notions around video signal such as Color-Gamut and Dynamic-Range but also introduces the standards which specify Standard-Definition Television (SDTV), HDTV and UHDTV. In addition, video quality assessment and methods for video coding performance evaluation are introduced. After a brief history on video coding history, the new tools introduced in HEVC are described in Section 2.2 as well as the related performance. Finally, SHVC will be introduced in Section 2.3.

Chapter 3 provides a complete state-of-the-art on rate-control schemes for video coding. First, the rate control concepts and stakes are introduced as well as an overview of rate-control architecture. Second, the existing bitrate allocation approaches are described and discussed. Third, the parameter estimation approaches, which enable to reach a targeted bitrate are also explored and discussed. Finally, this chapter will be concluded by an analysis which highlights the interesting fields of investigation and motivations of this work.

Chapter 4 develops our first contribution, dedicated to ρ -domain based rate-control scheme for HEVC and SHVC. This Chapter starts with an exhaustive statistical study which validates the concept of ρ -domain for both HEVC and SHVC. Then, an original algorithm for quantization parameter estimation is provided in Section 4.3. Then, a deterministic ρ domain formulation avoiding the use of look-up-table is provided in Section 4.4. Eventually, this work is concluded and discussed in Section 4.5.

Chapter 5 describes our second contribution on adaptive rate control algorithm in SHVC. A statistical study on the impact of bitrate ratio between layers over performance in SHVC is firstly provided in Section 5.2. Then, a first original approach is described in Section 5.3 which proposes an algorithm which dynamically adjusts the bitrate among layers in accordance with a fixed ratio interval. A more complete version of the algorithm is then described in Section 5.4 with additional constraints and dynamic ratio interval. The results of these algorithms are provided and analysed in Section 5.5 for different scalability configurations including spatial, color gamut and dynamic range.

Chapter 6 introduces the statistical multiplexing for scalable video coding and provides a first approach based on pre-encoding. This first algorithm has been developed based on the observations made during the statistical study on the bitrate ratio impact. Indeed, performing a statistical multiplexing of scalable programs does not only involve to allocate the bitrate among different programs, but also to adjust the bitrate ratio in these programs to use SHVC in the most optimal way. In Section 6.1, the statistical multiplexing problem is presented and the notations defined. Then, our pre-encoding based approach is described and the performance are evaluated in 6.2. Finally, Section 6.3 concludes this first approach and draw some prospectives for future work.

Chapter 7 concludes this document. The initial targets are firstly reminded and the direction taken are discussed in Section 7.1. Then, a summary of achieved targets is provided in Section 7.2 and the encountered issues are debated. The future work prospectives and potential improvements are finally addressed in Section 7.3.

Appendix A and B provide exhaustive statistical studies and additional results obtained in our contributions, respectively, while the appendix C lists the publications, standardization contributions and patent applications published during this thesis.

References

- [1] ITU-T, "Recommendation H.265: High Efficiency Video Coding."
- [2] ISO/IEC, "International Standard 23008-2: High Efficiency Video Coding."
- [3] ITU-T, "Recommendation H.264: Advanced Video Coding."
- [4] ISO/IEC, "International Standard 14496-10: Advanced Video Coding."
- [5] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2014-2019."
- [6] DVB, "DVB Fact Sheet: Introduction to the DVB Project," May 2014.
- [7] W. Husak, "ISO/IEC JTC1/SC29/WG11-M37372: ATSC Liaison on SHVC," Oct. 2015.

Part II

State-of-the-art

Chapter 2

Single-layer and scalable video coding

2.1 Video signal

The video signal is defined as a sequence of pictures usually called frames. Each frame can be characterized by several properties described in standardized specifications defined in the next sections.

2.1.1 Framerate and resolution

The video is recorded at a certain temporal frequency, measured in frames per second (FPS) unit. Figure 2.1 shows an example of the video sequence characterized by its framerate *f* .

Fig. 2.1 Representation of a video sequence with framerate *f*

Each frame might be represented by a matrix of pixels (in 2 dimensions). To properly represent colors, a pixel is composed of three channels (components) related to red, blue and green components, in RGB color space (see Section 2.1.3). Due to human visual system (HVS) response, and its high sensitivity to the luminance, and for compression purposes, it is more suitable to represent a pixel through the YUV representation, with Y the luminance and UV the chrominance components. Mathematical tools enabling conversion from RGB to YUV representations are standardized, for example in the ITU-R BT.709 [1] which is the

Y v Ľ Y Y $\widehat{\left\langle \text{V}\right\rangle }$	v (Y v X Y Y u U	⊛ Ŷ Y Y) Y v D				
Y Y V Y. Y. Y. \circledR	v Œ Y Y Y v .u ___	D Y. Y Y Y Y				
Ÿ Y ∿ E Y Y. $\widehat{\mathsf{W}}$	___ Y (Y) Y Y) Y 'v υ	$^\circledR$ Y Y) Y v Y y u				
Y Y V $\frac{1}{2}$ Œ Y	$\overline{\mathfrak{D}}$ œ Y Y Œ Y O U \mathbf{v}	$(\widehat{\mathsf{Y}})$ Y. (Y) (Y) Y Y.				
4:4:4	4:2:2	4:2:0				

Fig. 2.2 Example of 444, 422 and 420 color sampling on a frame of 4x6 pixels

specification of parameters values for the HDTV standards for production and international programme exchange. Equation 2.1 describes conversion in conformity with this standard.

$$
\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.2126 & 0.7152 & 0.0722 \\ -0.1146 & -0.3854 & 0.500 \\ 0.4809 & -0.4542 & -0.0458 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix}
$$
(2.1)

2.1.2 Color-sampling and bitdepth

As mentioned, the HVS is more sensitive to luminance component than chrominance ones. Thus, U and V matrices can be down-sampled with a minimum degradation with respect to the HVS and using YUV representation enables to lighten the signal by regrouping subsampled U and V components of collocated pixels. This approach is called color sampling, and three representations are commonly used in the video compression field: 444, 422 and 420 sampling. In 444 sampling, each pixel has its own YUV channels. In 422 sampling, each horizontal pair of pixels shares the same U and V channels. In 420 sampling, each frame is split into 2x2 pixel blocks sharing the same U and V channels. These sampling representations are illustrated in Figure 2.3. Numerically, each color component is coded on a number of bits specified by the bitdepth value. Then, a pixel using a 444 color sampling with a bitdepth of 8 bits is represented over 24 bits while the same pixel using a 422 color sampling only costs 16 bits and 12 bits in 420 color sampling. All these aspects can be combined with another tool inherited from old TV called interlaced video. An interlaced video consists in representing a succession of even and odd frames. The even frames are only composed of even rows, while odd frames are composed of odd rows. The interlaced video is the opposite of progressive video, where all rows are represented in one frame. With this technique, the cost of an interlaced frame is halved compared to a progressive one. This enables to significantly reduce the cost of a sequence and it perfectly suits to analog television scanning patterns. Then, a received interlaced frame has to be processed to extract the two half-frames in it. In addition, other post-processing operations to interpolate the data

between rows can be achieved in order to have a decent progressive frames to be displayed. Using interlaced video is not problem-free since annoying visual artefacts may be created. Hence, using interlaced video means choosing lower bitrate at the cost of visual quality.

2.1.3 Color gamut

Color gamut is defined as the limitation on the range of colors that can be represented in a particular framework [2]. This framework can be HVS, a camera, or a display. In the television industry, colorimetry was standardized based on the limitation of display devices. SDTV color space is defined in ITU-R BT.601 [3] and HDTV is defined ITU-R BT.709 [1]. With the advent of UHDTV and new display technologies (e.g. OLED and quantum dot) the required coverage for color goes beyond the BT.709 specification. The UHDTV color space is specified in the ITU-R BT.2020 [4] standard and has a larger color gamut than BT.709 and hence, referred to as Wide color Gamut or WCG. WCG is a significant feature of the UHDTV signal as it requires rendering a richer set of colors which brings the user closer to reality. Figure 2.3 shows HDTV and UHDTV color gamut in the CIE 1931 chromacity diagram [5].

2.1.4 Dynamic range

Dynamic range of a scene is defined as the ratio of the maximum light intensity to the minimum light intensity [7]. In terms of camera acquisition, cameras measure the dynamic range of a scene in terms of f-stops, which defines the range of light intensity in power of 2. In terms of display, we consider the contrast ratio to associate the full on/off luminance ratio of a television display. Essentially, the higher the contrast ratio of a display is, the better the dynamic range is. Table 2.1 gives a brief understanding of the relationship between f-stop, contrast and dynamic range. Current television displays provide Standard Dynamic Range (SDR), which supports a limited range of brightness especially with respect to the HVS. These displays provide a range of light intensity of around 0.1−100 *cd*/*m* 2 (or "nits"). This is in huge contrast with the HVS which can perceive the night sky of 0.005 nits and can also adapt to different colors and textures on a 10000 nits sunny day. Currently, 1000 nits Extended Dynamic Range (EDR) displays are available in the market and brighter HDR displays are expected in the future to bring the user closer to reality.

Table 2.1 HDR characteristics [8]

F-Stops	Contrast Ratio	Dynamic Range
10	$2^{10} = 1024:1$	Standard Dynamic Range
16	$2^{16} = 65536:1$	Extended Dynamic Range
20	$2^{20} = 1048576:1$	High Dynamic Range

		SDTV				HDTV	UHDTV		
Aspect-Ratio			4/3 & 16/9			16/9	16/9		
Resolution	864×625 858×525				1920×1080	3840×2160	7680×4320		
Scan mode	Interlace					Interlace or Progressive	Progressive		
Frame Frequency	50 60/1.001				60, 60/1.001, 50, 30,	120, 60, 60/1.001, 50, 30,			
						30/1.001, 25, 24, 24/1.001	30/1.001, 25, 24, 24/1.001		
Primaries	X		X		X		X		
red	0.630	0.340	0.640	0.330	0.640	0.330	0.708	0.292	
blue	0.310	0.595	0.290	0.600	0.300	0.600	0.170	0.797	
green	0.155	0.070	0.150	0.060	0.150	0.060	0.131	0.046	
white	0.3127	0.3290	0.3127	0.3290	0.3127	0.3290	0.3127	0.3290	

Table 2.2 Comparison of SDTV, HDTV and UHDTV standards

2.1.5 The SDTV, HDTV and UHDTV standards

All the sequences handled through our work consider signals characterized by the previously described parameters. In addition to the video sequences used during standardization process, our algorithms have to be evaluated over sequences compliant with industrial standards. In order to create a synergy between all industrial actors involved in the market and the consumers, signal formats have been standardized by the ITU. The SDTV was first introduced in 1995 in version 5 of Recommendation ITU-T BT.601-5. The HDTV was also standardized

Fig. 2.3 Representation of Rec.709 (HDTV) and BT.2020 (UHDTV) color spaces in the CIE 1931 diagram [6]

in 1995 as version 2 of Recommendation ITU-T BT.709. Eventually, the UHDTV was recently released in 2012 as the version 1 of Recommendation ITU-T BT.2020. In Table 2.2, the major differences between these standards are shown. We can notice that the transition between SDTV and HDTV mainly involves extension of spatial resolution and additional framerate, but with identical color primaries. However, we can observe that the transition between HDTV and UHDTV brings a better resolution but also a wider color gamut with extended color primaries, as mentioned in Section 2.1.3.

2.1.6 UHD introduction phases

The DVB Project [9] is an international consortium providing technical standards for broadcast services [10]. This includes the commercial requirements for UHD broadcast. It is intended that UHD be deployed in phases to accommodate the availability of the HEVC decoders and the advancements in display and transport technology. UHD is defined in two profiles in terms of resolution, namely UHD-1 (3840 \times 2160) and UHD-2 (7680 \times 4320). The diagram of UHD-1 Phase 1, UHD-1 Phase 2 and UHD-2 is shown in Figure 2.4.

Fig. 2.4 DVB phased UHD deployment [10]

This phased approach gives room for technology to develop and cater to the needs of the UHD signal. Furthermore, broadcasting using a scalable codec makes sense especially considering the importance of backward compatibility between the phases. This DVB use-case is discussed in detail in Section 5.5.3. It must be noted that we have not mentioned high frame rate which is also a key feature of the UHD signal. HFR has not been treated in this work but it is a definite candidate to be considered in rate control as future work.

2.1.7 Video quality assessment

The quality of a compressed video signal can be evaluated in different ways. First of all, it can be assessed via standardized subjective methods such as described in Recommendations ITU-R BT.500-13 [11], ITU-T P.910 [12] or ITU-R BT.1788 [13]. However, these methods require to gather people and are time-consuming. On the other hand, quality can be evaluated through objective metrics. The mean square error (MSE), is the most basic way of evaluating the difference between compressed and uncompressed (original) signals and is defined by the following equation:

$$
MSE = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (S_C(i, j) - S_0(i, j))^2,
$$
\n(2.2)

where M and N are respectively the spatial dimensions of the 2D video signal, S_0 the original signal (pixel's value) and S_C the de-compressed (distorted) value of the video signal. The widely used peak signal-to-noise ratio (PSNR) is derived from the MSE combined with the maximum reachable value of pixels. The PSNR is computed in *dB* and is defined in the following equation, where MAX_I is the maximum possible value of the pixel:

$$
PSNR = 10 \times log_{10} \left(\frac{MAX_I^2}{MSE} \right). \tag{2.3}
$$

Since pictures are composed of channels, PSNR is usually separately computed for Y, U and V components (in YUV representation). The global YUV-PSNR [14] value is widely computed as the weighted mean of these channels, as described in the following equation:

$$
PSNR_{YUV} = \frac{6 \times PSNR_Y + PSNR_U + PSNR_V}{8}.
$$
\n(2.4)

In addition to these simple methods, more advanced and complex metrics exist such as the structural similarity index (SSIM) [15] and the video quality metric (VQM) [16].

2.1.8 Performance evaluation of a video coding technology

In order to evaluate the performance of a particular coding technology, objective and subjective methods can also be used. The widely used method is called Bjøntegaard metric [17]. This metric consists in measuring the average difference between two Rate-Distortion (R-D) curves interpolated through four bitrate points. This method enables to compute the bitrate gain (BD-BR) for an equivalent level of distortion, or the PSNR gain (BD-PSNR) for an equivalent bitrate. The method firstly consists in interpolating rate-distortion functions defined as a log-based third order polynomial:

$$
D \triangleq f_1(R) = a_1 + b_1 * log(R) + c_1 * log(R)^2 + d_1 * log(R)^3,
$$
\n(2.5)

$$
log(R) \triangleq f_2(D) = a_2 + b_2 * D + c_2 * D^2 + d_2 * D^3,
$$
\n(2.6)

where *D* and *R* represent the distortion (in PSNR) and the rate of the video, respectively. Then, the difference between both R-D curves is computed, as illustrated in Figure 2.5. This step enables to estimate the average gain in bitrate or PSNR between two coding tools.

Fig. 2.5 Illustration of computation of average PSNR between two R-D curves

2.2 Foreword: video compression history

Video coding consists in reducing the amount of data used to represent a digital video signal as described in Section 2.1. The coding may be lossy or lossless depending on the industrial application and on a visual-degradation (quality)/compression-ratio (bitrate) trade-off. A

number of paradigms and tools have been tested and used in this field which led to build many video coding standards over the years, as we can observe in Figure 2.6. The video coding standardization enables to unify the coding tools and standardize the encoder output (ie. video bitstream) to ensure that this bitstream could be correctly decoded by any compliant decoder. Thus, a video standard ensures interoperability among encoders and decoders over the world. In 1984, the International Telegraph and Telephone Consultative Committee (CCITT) published the first digital video compression standard known as H.120 [18]. The

Fig. 2.6 History of video coding technology and standards

first version of H.120 included techniques such as scalar quantization, variable-length coding or differential pulse-code modulation (DPCM) for instance. In the second version, published in 1988, motion compensation and background prediction tool were added. In 1992, the CCITT has been merged with the ITU in a new branch dedicated to telecommunication area and called ITU-T. Under this new name, the ITU-T group ratified the successor to H.120 called Recommendation H.261 [19], first member of the H.26x family. This first standard was developed inside of the VCEG and is considered to be the first widely used video coding technology, which defines the basis of modern video compression. The majority of techniques and concepts used in recent standards such as macroblock-based motion compensation, Discrete Cosine Transform (DCT), zig-zag scanning, run-length and variable-length coding had been introduced in H.261. In parallel with VCEG standardization activity, a new standardization group called MPEG was established in 1988 as a collaborative group between the ISO and the IEC. Technically, the MPEG group operates in the Joint ISO/IEC Technical Committee (JTC) 1 on the Information Technology as the Working Group (WG) 11 of Subcommittee 29 (SC). At this time, the mandate of MPEG was to develop standards for the representation of moving pictures and/or audio. In 1993, the ISO/IEC ratified the first standard developed by MPEG called MPEG-1, or formally ISO/IEC 11172-2 [20]. This standard inherited many features from H.261 but also added new ones such as bi-directional motion prediction, half-pixel motion estimation, slice-structured coding or quantization weighting matrices. In order to create a synergy between the MPEG and VCEG activities, a first collaborative work has been initiated to jointly develop the Recommendation ITU-T H.262 [21] and ISO/IEC 13818-2 [22], released in 1994. Inherited from both H.261 and MPEG-1, the H.262/MPEG-2 standard mainly added the support of interlaced pictures and introduced the scalability concept in video coding (SNR and spatial). This standard encountered success and massive adoption through DVD, broadcast and broadband industry. This joint collaboration did not prevent these two groups from developing other standards on their part. In 1995, the ITU-T ratified the Recommendation ITU-T H.263 [23] which introduced some of the well known P and B frames, the arithmetic entropy coding and an increased motion vector range. In the same way, MPEG independently developed the MPEG-4 Visual [24]–or ISO/IEC 14496-2– which introduced new coding features such as segmented coding of shapes, 10b/12b sampling and specific tools for synthetic and semi-synthetic content. Both H.263 and MPEG-4 Visual brought significant gains, but cannot knock down MPEG-2, especially in broadcasting field. In 2001, the Joint Video Team (JVT) was officially created, regrouping experts from MPEG and VCEG with the mandate of finalizing a more efficient standard, successor to MPEG-2, and initiated in 1998 by VCEG: H.264 –or Advanced Video Coding (AVC)–. The Recommendation H.264 [25] or ISO/IEC 14496-10 [26], developed by the JVT, was

ratified and published in 2003 and consists in a significant improvement of MPEG-2, with additional tools such as in-loop filter for example. This standard encountered a huge success across the industry and has been widely adopted, especially in telephone companies (Telcos), broadcast and digital storage area especially for HD video signal. In 2010, the JCT-VC was created with the mandate of reducing by 50% the bitrate compared to H.264/AVC. In 2013, HEVC was finalized and published as Recommendation H.265 [27] or ISO/IEC 23008-2 [28]. The HEVC standard is much more efficient than AVC, insuring at least 50% of bandwidth reduction [29, 30] at similar subjective quality. In addition to MPEG and VCEG standardization efforts, several other technologies exist. In 1995, a private company called On2 Technology launched the development of TrueMotion S, a proprietary video codec addressed to 3D-rendered video. Several enhanced versions were released, TrueMotion RT in 1996 and TrueMotion 2 in 1997. In 2000, On2 Technology launched the VP3 codec suitable for natural scene coding. The VP3 was turned into free-software which enables the Xiph organization to design its first open-source coding technology Theora [31] in 2001. Several generations of VP codec were released with VP4 in 2001, VP5 in 2002 and VP7 in 2005. In 2008, the VP8 [32] was released by On2 Technology which claimed better performance than the AVC standard [33]. In February 2010, On2 Technology was acquired by Google which made the source-code free within the WebM project [34], under BSD-License agreement. In 2011, the development of VP9 [35] was launched with the objectives of overcoming the performance of both VP8 and HEVC. In 2006, Microsoft standardized its proprietary coding technology called VC-1 [36], a serious alternative to H.264/AVC which was supported in Blu-Ray Discs and in the dead HD-DVD. Eventually, Xiph organization released their latest codec named Daala [37] in 2013 which is still supposed to be an alternative to VP9 and H.265/HEVC.

In addition to these 2D single layer (SL) technologies, scalable coding also evolved during the last decades. Unlike the classic SL approaches, scalable coding enables to produce a layered version of a video signal, each layer representing a different and enhanced representation of the base layer signal. This enhancement may be represented, for instance, by an improved spatial or temporal resolution but also by a better fidelity or a wider color gamut. Indeed, the MPEG-1 standard already included mechanism for performing temporal scalability. In the same way, the H.262/MPEG-2 standard also included tools for spatial, temporal and fidelity scalable coding. In October 2003, the MPEG group issued a Call for Proposal (CfP) on scalable video coding. This CfP results in fourteen proposals, submitted in October 2004. Among these proposals, twelve were wavelet-based while two were extensions to H.264/AVC (DCT transform based). The AVC-based solution proposed by the Heinrich Hertz Institute

(HHI) was selected by MPEG as a basis for SVC standardization. The SVC extension was approved in July 2007 as the Annex G of H.264/AVC. In the same way, an HEVC scalable extension SHVC was developed and approved in October 2014 as the Annex H of HEVC.

2.3 HEVC standard

2.3.1 Standardization timeline

The development of HEVC was achieved through several steps [38], as described in Figure 2.7. In 2004, VCEG initiated coordinated investigations in a software platform inherited from the AVC Joint-Model (JM) reference software called Key Technology Area (KTA). In April 2005, an official group was created to maintain the software. The KTA was used to integrate and test many promising technologies. From 2006 to 2008, MPEG organized a rotation of workshops dedicated to the investigation on potential technologies that may help to improve the coding efficiency of the future video coding standard. In 2009, MPEG launched a Call for Evidence (CfE) on future video coding in order to find evidence that a the creation of a future video coding standard is relevant. Judging that a new standardization cycle can be started, the JCT-VC group was created in 2010, and issued a Call for Proposal (CfP) on future video coding standard. Twenty seven proposals were received and evaluated through objective and subjective evaluations. After evaluating these responses, the project name was defined as High Efficiency Video Coding (HEVC) and a basis called Test Model under Consideration (TMuC) was created. The TMuC combined technologies proposed by seven proposals, and was the basis of HEVC development. From 2010 to 2013, HEVC was developed and technologies were added until overtaking the 50% coding gain with respect to AVC. Finally, the first version of HEVC was ratified in January 2013 and the Rec. ITU-T H.265 and its twin ISO/IEC 23008-2:2013 was released in April 2013.

Fig. 2.7 Timeline of HEVC development, from KTA to standardization

2.3.2 HEVC profiles and extensions

The first version of HEVC was released in February 2013 and came with three profiles the Main, Main10 and Main Still Picture. In addition, HEVC comes with tiers and levels specifying constraints on the bitstream such as maximum bitrate, luma sample rate or maxium number of tiles or slices. The version 2 of HEVC was published in 2015 and includes rangeextension, scalable and multiview extensions. These extensions came with 24 new profiles: 21 range-extensions, two scalable and one multiview profiles. More details on the scalable extension SHVC will be provided in section 2.4. Eventually, additional 3D Main profiles were added in HEVC version 3, few months after the second version.

Fig. 2.8 Schematic block diagram of any HEVC encoder ([39])

2.3.3 Overview of coding tools

The HEVC standard is based on the same hybrid scheme used in previous video coding standards [39], illustrated in Figure 2.8. The input video picture is first split into blocks called coding tree unit (CTU) which are basically processed sequentially in raster scan order. In HEVC, the CTU is represented according to a flexible quadree decomposition in coding units (CU). Each CU leaf is then the basis for prediction and quantization/transformation stages which are respectively performed on sub-partitions called prediction unit (PU) and transform unit (TU). Each of these units are accompanied by their Luma and Chroma blocks referring to texture information, for instance the Y, U and V coding tree block (CTB), coding block (CB), prediction block (PB) or transform block (TB). For each PU, two kinds of prediction may be performed. The first one is called intra prediction and consists in predicting the textures in the current PU according to already coded textures in the same picture. The second one is named inter prediction and consists in predicting the texture in the current PU according to one –or the combination of two– texture(s) located in the already coded frames, which can be temporally located in the past or the future. The information required to perform the prediction such as direction of intra prediction, spatial offset between texture and its temporally collocated references (also called motion vector, MV), index of pictures used in predictions and any other essential information will be entropy coded and transmitted throughs the final bitstream. Once the textures in a CU are predicted, and all information related to prediction stored, the predicted textures are subtracted to uncompressed ones to build the residual. The obtained residual is firstly transformed by a DCT or even a Discrete Sine Transform (DST) in order to decorrelate the coefficients and concentrate all the energy in few ones, and set the rest of coefficients close to zero. These transformed residuals are then scaled and quantized to only keep the most important coefficients related to low frequency and set the other ones related to high frequency to zero. The transformed and quantized residuals are then efficiently described through several syntax elements (SE) which are entropy coded. In HEVC, all the SE related to the video coding layer (VCL) are entropy coded using a context-adaptive binary arithmetic coding (CABAC). The other SE refer to the high level syntax (HLS) and are encoded using fixed or variable-length codes. In order to have reference textures for future predictions, the encoder includes a decoding loop which performs inverse quantization and transformation on the residuals. The decompressed residuals are added to predicted textures, and two filters including deblocking filter (DBF) and sample adaptive offset (SAO) may be applied to remove and reduce blocking effects and ringing artefacts. Eventually, all information related to prediction, quantization, transform, filtering, decoded picture buffer and all other information required in the decoder are formatted according to HEVC specification to produce a compliant HEVC bitstream.

Parallel processing

HEVC was designed with a particular attention to complexity to enable real time and low delay processing of high resolution video. Therefore, HEVC introduces several parallel tools [40] which enable both parallel encoding and decoding. The independent slice concept (which already exists in AVC) consists in forming CTU aggregates, parsed in raster-scan order. A frame can be split into several independent slices, independently decodable, as illustrated in Figure 2.9b. As shown in Figure 2.9c, the tile concept splits the picture into rectangular groups of CTUs, called tiles. The independent slices and tiles break the CABAC

Fig. 2.9 Illustration of HEVC parallel-friendly tools

and the intra prediction dependencies and thus can be used for parallel encoding and decoding. However, intra prediction limitation and resetting the CABAC probabilities decrease the coding performance in terms of rate distortion, especially for large number of tiles/slices per frame. Moreover, the in-loop filters cannot be performed in parallel at the tile/slice edges without additional control mechanism. The Wavefront Parallel Processing (WPP) solution splits the frame into CTU rows enabling parallel encoding/decoding in wavefront pipeline. In the WPP mode, the CABAC context is initialized at the start of each CTU row. The overhead caused by this initialization is limited since the CABAC context at each CTU row is initialized by the CABAC context state at the second CTU of the previous CTU row. Figure 2.9d illustrates the principle of the WPP. All those tools make possible various multi-threaded implementations of HEVC, at both encoder [41] and decoder [42] sides, which enables to leverage current multi-core/many-core processors and chipsets available in the market.

Partitioning structure

To fit with larger resolution coming with UHDTV and 4K, a new generation codec such as HEVC have to propose a flexible block structure which enables to match with both small and large spatial variations of every video signal. This way, HEVC provides a very flexible and efficient quadtree decomposition previously introduced based on CTU splitting. Each CTU in a picture is described according to a quadtree representation [43]. The quadtree is represented through different kinds of blocks : the CTU, Coding Units (CU), Prediction Unit (PU) and Transform Units (TU). These units have an associated textures blocks related to Y, U and V textures, respectively called Coding Tree Blocks (CTB), Coding Block (CB), Prediction Block (PB) and Transform Block (TB). In the YUV color-component point of view, a CTU is composed of a luma CTB and two chroma CTB, a CU is composed of a luma CB and two chroma CB, and so on. Each CTU can be recursively represented with CU leaves. Then, the CU can be split again into four square CUs in the underlying depth, and so on until reaching the tree depth. In Figure 2.10, we can observe a non-split CTU in the left representation and its associated quadtree. In the right example, a CTU with more

Fig. 2.10 Illustration of CTU quadtree partitioning

complex quadtree is represented. The CU are processed following a descending raster-scan order, as described in the right example. The CU representation and its associated descriptive syntax allow to specify how the pixel areas are processed inside the CTU. Each CU can then be described through other partitioning schemes related to prediction and transform modes. In a similar way as macroblocks in AVC, a CU in HEVC can be categorized as a skipped, inter-coded or intra-coded according to the prediction type used for the block. For each prediction type, a set of available partitioning schemes is defined, as described in Figure 2.11.

Fig. 2.11 Available PU partitioning in HEVC

If the CU is coded in skip mode, then it cannot be split and is only represented by a 2Nx2N partition of same dimensions. If the CU is intra-coded, it can be split into NxN partition which is composed of four new 2Nx2N sub-partition. These 2Nx2N blocks can be split again with 8x8 as minimal reachable dimension. In the same way, eight kinds of partitioning are available for inter-coded CU, the classical 2Nx2N and NxN, two symmetrical 2NxN and Nx2N modes and four asymmetrical modes 2NxnU, 2NxnD, nLx2N and nRx2N. In addition to this prediction-related partitioning, a transform-tree structure called residual quadtree of TU (RQT) is used to describe the splitting related to transforms. In the same way as CU partitioning work for CTU, the TU partitioning enables to describe how the transform is applied to residual areas within the CU. Thus, a CU can be split into four square partitions, and so on until reaching the 8x8 maximal luma dimension. All those CTU, CU, PU, TU and related blocks are described through syntax elements according to the standard specification.

Intra coding

Three ways of achieving intra-coding exist in HEVC and may be applied in a PU, the DC mode, the planar mode and the angular mode [44]. In Figure 2.12, the reference samples,

Fig. 2.12 Intra prediction modes in HEVC

the available modes and the related notations are provided. The available reference samples for an intra-predicted $N \times N$ luma block, denoted $R_{x,y}$, are the 2*N* vertical and horizontal rows started from the up-left corner sample $R_{0,0}$, as illustrated in Figure 2.12a. In DC mode, the samples are predicted with the average value of available collocated reference samples, as represented in Figure 2.12b. The planar mode aims at preserving the continuities across boundaries and is computed as the mean of a horizontal and vertical predictions, as illustrated in Figure 2.12c and defined in the following equations:

$$
P_{x,y}^V = (N - y) \times R_{x,0} + y \times R_{0,N+1}
$$
\n(2.7a)

$$
P_{x,y}^{H} = (N - x) \times R_{0,y} + x \times R_{N+1,0}
$$
\n(2.7b)

$$
P_{x,y} = (P_{x,y}^V + P_{x,y}^H + N) \gg (log_2(N) + 1)
$$
\n(2.7c)

Finally, the angular mode enables to predict the samples according to a set of 33 angles illustrated in Figure 2.12d. The pixels are predicted as the value of interpolated reference sample resulting from the angular projection. This value can be computed as following, where *d* is the projection displacement resulting from angular prediction.

$$
P_{x,y} = ((32 - w_y) \times R_{i,0} + w_y \times R_{i+1,0} + 16) >> 5
$$
 (2.8a)

$$
c_y = (y \times d) \gg 5 \tag{2.8b}
$$

$$
w_y = (y \times d) \& 31 \tag{2.8c}
$$

$$
i = x + c_y \tag{2.8d}
$$

This improved angular mode compared to AVC enables a significant objective quality improvement since this angle set is more efficient to track boundaries and thus to perform better predictions. In addition to the visual aspect, this set of HEVC intra tools provides an average bitrate reduction of 22% compared to AVC in all intra mode with a maximum gain of 36% [44].

Index i		-3	-2	-1	$\mathbf 0$	$+1$	$+2$	$+3$	$+4$
Luma	h filter[i]	-1	4	-11	40	40	-11	4	
	qfilter $[i]$		4	-10	58	17	-5		N/A
Chroma	filter $1[i]$	N/A	N/A	-2	58	10	-2	N/A	N/A
	filter $2[i]$	N/A	N/A	-4	54	16	-4	N/A	N/A
	filter $3[i]$	N/A	N/A	-6	46	28	-4	N/A	N/A
	filter $4[i]$	N/A	N/A	-4	36	36	-4	N/A	N/A

Table 2.3 Filter Coefficients for Chroma and Luma Fractional Sample Interpolation

Inter coding

As for AVC standard, the inter-picture prediction in HEVC consists in using samples areas localized according to the MV as a reference. Indeed, the related picture is identified thanks to the reference picture index while the pixel location is determined with the shifting involved by the MV vertical and horizontal components. If the computed shifting values are integer, then the location of reference pixels directly refer to existing samples. If the shifting values are fractional, then the reference picture has to be re-sampled according to a eight/seven-tap filter for luma samples and four-tap filters for chroma samples, which enables a quarter-pixel accuracy. In Figure 2.13, we can observe the positions of integer and fractional reference samples in HEVC inter-picture prediction, the filter coefficients are recorded in Table 2.13. The fractional samples vertically and horizontally aligned with the integer positions are firstly computed, according to the following Equations, using 7-tap and 8-tap filters. In these

$A_{-1,-1}$		$A_{0,-1}$	$\mathbf{a}_{\text{o},\text{-}1}$	$\mathbf{b}_{0,-1}$	$\mathbf{c}_{0,-1}$	$A_{1,-1}$		$A_{2,-1}$
$A_{-1,0}$		$A_{0,0}$	$\mathbf{a}_{0,0}$	$\mathsf{b}_{0,0}$	$\mathbf{C}_{0,0}$	$A_{1,0}$		$A_{2,0}$
$d_{-1,0}$		$d_{0,0}$	$\mathbf{e}_{0,0}$	$\boldsymbol{f}_{0,0}$	$\mathbf{g}_{\text{0,0}}$	$\mathsf{d}_{1,0}$		$\mathsf{d}_{2,0}$
$h_{-1,0}$		$\mathsf{h}_{\mathtt{0,0}}$	$i_{0,0}$	$j_{0,0}$	$\mathbf{k}_{0,0}$	$\mathsf{h}_{1,0}$		$\mathsf{h}_{2,0}$
$n_{-1,0}$		$\mathsf{n}_{\scriptscriptstyle 0,0}$	$\mathsf{p}_{\scriptscriptstyle 0,0}$	$\mathbf{q}_{0,0}$	$r_{0,0}$	$\mathsf{n}_{1,0}$		$n_{2,0}$
$A_{-1,1}$		$A_{0,1}$	$\mathbf{a}_{0,1}$	$\mathbf{b}_{0,1}$	$\mathbf{C}_{0,1}$	$A_{1,1}$		$A_{2,1}$
$A_{-1,2}$		$\mathsf{A}_{0,2}$	$\mathbf{a}_{0,2}$	$\mathsf{b}_{0,2}$	$ \mathbf{c}_{0,2} $	$A_{1,2}$		$A_{2,2}$

Fig. 2.13 Sample positions for luma interpolation in HEVC [39]

Equations, *B* represents the bitdepth of the samples processed.

$$
a_{0,j} = \left(\sum_{i=-3..3} A_{i,j} \times qfilter[i] \right) \gg (B-8)
$$
 (2.9a)

$$
b_{0,j} = \left(\sum_{i=-3..4} A_{i,j} \times hfilter[i] \right) \gg (B-8)
$$
 (2.9b)

$$
c_{0,j} = \left(\sum_{i=-2..4} A_{i,j} \times qfilter[1-i] \right) \gg (B-8)
$$
 (2.9c)

$$
d_{0,0} = \left(\sum_{i=-3..3} A_{0,j} \times qfilter[j] \right) \gg (B-8)
$$
 (2.9d)

$$
h_{0,0} = \left(\sum_{i=-3..4} A_{0,j} \times hfilter[j] \right) \gg (B-8)
$$
 (2.9e)

$$
n_{0,0} = \left(\sum_{i=-2..4} A_{0,j} \times qfilter[1-j] \right) \gg (B-8)
$$
 (2.9f)

Once these vertical and horizontal fractional references are computed, the remaining samples can be computed, according to the following equations.

$$
e_{0,0} = \left(\sum_{v=-3..3} a_{0,v} \times qfilter[v] \right) \gg 6 \tag{2.10a}
$$

$$
f_{0,0} = \left(\sum_{v=-3..3} b_{0,v} \times qfilter[v] \right) >> 6 \tag{2.10b}
$$

$$
g_{0,0} = \left(\sum_{v=-3..3} c_{0,v} \times qfilter[v]\right) >> 6 \tag{2.10c}
$$

$$
i_{0,0} = \left(\sum_{v=-3..4} a_{0,v} \times hfilter[v]\right) >> 6 \tag{2.10d}
$$

$$
j_{0,0} = \left(\sum_{v=-3..4} b_{0,v} \times hfilter[v]\right) >> 6
$$
 (2.10e)

$$
k_{0,j} = \left(\sum_{\nu=-3..4} c_{0,\nu} \times hfilter[\nu] \right) >> 6 \tag{2.10f}
$$

$$
p_{0,0} = \left(\sum_{v=-2..4} a_{0,v} \times qfilter[1-v]\right) >> 6\tag{2.10g}
$$

$$
q_{0,0} = \left(\sum_{v=-2..4} b_{0,v} \times qfilter[1-v]\right) >> 6 \tag{2.10h}
$$

$$
r_{0,0} = \left(\sum_{v=-2..4} c_{0,v} \times qfilter[1-v]\right) >> 6\tag{2.10i}
$$

In the same way as luma samples, the chroma samples are also interpolated using smaller 4-taps filers. The filter1, filter2, filter3 and filter4 are respectively used for computing the 1/8th, 2/8th, 3/8th and 4/8th fractional samples of chroma components. The 5/8th, 6/8th and 7/8th positions are computed in a symmetrical way.

Fig. 2.14 Illustration of AMVP candidates and merge list construction in HEVC

In HEVC, the motion vector can be predicted in a more efficient way than AVC which only provides one motion vector predictor for differential MV encoding. A mechanism called Advanced Motion Vector Prediction (AMVP) is then introduced and enables competition between several candidates, the index of the selected candidate being transmitted. In this scheme the candidates, illustrated in 2.14a, are derived as follows. Four spatially collocated neighbours A_0 , A_1 , B_0 , B_1 and B_2 are considered, plus one temporally collocated candidate C_1 or C_0 as temporal motion vector predictor (TMVP). If the AMVP candidate list contains less than two candidates then the list is completed with zero MV. In addition to this classical approach for quarter-pel motion compensated predictions, a new tool called Merge-Mode is introduced in HEVC [45]. The merge mode aims at reducing the redundancies introduced by the transmission of prediction information in the quadtree. In the same way as AMVP, blockmerging algorithm consists in building a list of merging candidate among the neighbours for the selected PU. If the merge mode is used, then the current PU will share and copy the motion information of the pointed out candidate. The initial candidates are firstly checked and added to the list, by analyzing the spatial and temporal neighbourhood, including a redundancy check. Then, if the list contains fewer candidates than 5, some additional candidates such as combined and zero-motion candidates are added until reaching the appropriate number of candidates (i.e 5).

(a) Diagonal scan pattern (b) Horizontal scan pattern (c) Vertical scan pattern

Fig. 2.15 Illustration of available scanning patterns in HEVC

Transform and quantization

The residual information in HEVC is represented according to the RQT which gives a flexible way of coding the residual information. Several transforms derived from the DCT are considered to decorellate the residual signal, for the 32x32, 16x16, 8x8 and 4x4 TB. In addition to these DCT-based transforms, a 4x4 DST transform is applied to 4x4 luma intra-predicted residuals. Once transformed, the TB are scaled and quantized in a similar way

as AVC, with a 0 to 51 range of quantization parameters (QP) and an optional quantization scaling matrices. HEVC introduces new paradigms in signaling the residual information [46], especially with residual scanning, significance map, coefficient level and sign coding. In AVC, the zigzag was the only way to scan a 2D residual array to turn it into a 1D treatable vector. In HEVC, a TB is scanned per 4x4 sub-block, each of these sub-blocks being scanned per coefficient. In Figure 2.15, the different way of scanning a 8x8 TB in HEVC is provided. As we can see, three kinds of scan pattern are available. The diagonal scan, represented in Figure 2.15a can always be used in HEVC and is not mode-dependent. However, two additional vertical and horizontal scanning modes, respectively represented in Figures 2.15c and 2.15b, can be used for scanning intra-predicted residuals in 4x4 and 8x8 TB. For each TB, the residuals are scanned per set of 16 coefficients (related to each 4x4 sub-block) called Transform Sub-Block (TSB). For each of the scanned coefficients, the SE described in Table 2.4 may be used, whether the coefficient is zero or non zero, has an absolute value greater than 1 or 2 and is positive or negative.

Table 2.4 Syntax-Elements used in HEVC for residual coding

SЕ	Role
significant_coeff_flag	Indicate if the coefficient is zero/nonzero
coeff_abs_level_greater1_flag	Indicate if the absolute value is greather than 1
coeff_abs_level_greater2_flag	Indicate if the absolute value is greather than 2
coeff_sign_flag	Indicate the sign 0 is positive and 1 is negative
coeff_abs_level_remaining	Indicate the remaining value of the absolute level,
	compared to the previous one

In-loop filters

In order to improve the quality of reconstructed reference picture, and thus to improve the coding efficiency, HEVC includes two in-loop filters: the DBF [47] and SAO [48] filter. The DBF, already present in AVC, has been well optimized in a computational sense and is more parallel friendly for HEVC. The SAO is a new feature of HEVC, which can be activated together with the DBF. The SAO algorithm classifies the reconstructed samples in two ways: band offset (BO) and edge offset (EO) samples. In the case of EO classification, the reconstructed samples are compared with the neighbouring samples, and a class is selected according to edge behavior on the block boundary as illustrated in Figure 2.16. The estimated offset it then applied to reconstructed samples to remove the artefacts. In the same way, the BO classified blocks are only evaluated according to the band samples, and an offset may also be applied. The SAO parameters are transmitted through a set of SE encoded by CABAC at the CTU level.

Fig. 2.16 Illustration of EO classification in SAO

Syntax element coding

In HEVC, several coding techniques are used to binarize the transmitted SE. This binarization is achieved according to several parsing process including various coding techniques such as entropy, fixed-length or variable-length coding. The information related to CU, PU, TU and SAO are coded thanks to CABAC. All the high level syntax (HLS) signaling is coded by a fixed-length or variable-length coding.

Fig. 2.17 Three main functions in the CABAC

CABAC in HEVC

The CABAC engine in HEVC consists of three main functions: binarization, context modeling and arithmetic coding [49]. These three main functions of the CABAC are illustrated in Figure 2.17. First, the binarization step converts syntax elements to binary symbols (bin). Second, the context modeling updates the probabilities of bins, and finally the arithmetic coding compresses the bins into bits according to the estimated probabilities. Five binarization methods are used in HEVC namely Unary (U), Truncated Unary (TU), Fixed Length (FL), Truncated Rice code with an adaptive context *p* (TRp) and k*th*-order Exp-Golomb (EGk) codes. The U code represents an unsigned integer *Y* with a binstring of length $Y + 1$ composed

of *Y* 1-bins followed by one 0-bin. The TU code is defined with the largest possible value of the syntax element *cMax* ($0 \le Y \le cMax$). When the syntax element value $Y \le cMax$, the TU is equivalent to U code, otherwise *Y* is represented by a binstring of *cMax* 1-bins. The FL code represents a syntax element *Y* with its binary representation of length $\lceil \log_2(cMax + 1) \rceil$. The TRp code is a concatenation of a quotient $q = \lfloor Y/2^p \rfloor$ and a remainder $r = Y - q2^p$. The quotient *q* is first represented by TU code as a prefix concatenated with a suffix *r* represented by the FL code of length *p*. The EGk code is also a concatenation of prefix and suffix. The prefix part of the EGk code is the U representation of $l(Y) = \lfloor \log_2(\frac{Y}{2^i}) \rfloor$ $\frac{Y}{2^k} + 1$]. The suffix part is the FL code of $Y + 2^k(1 - 2^{l(Y)})$ with $cMax = k + l(Y)$.

The arithmetic coder can be performed either by an estimated probability of a syntax element (context coded) or by considering equal probability of 0.5 (bypass coded).

2.3.4 Reference software

The HEVC Reference Software is the result of a long collaborative work based on tools integration, test and validation. The current implementation, called Test Model 16 (HM16) [50], enables several encoding configurations. The All-Intra (AI) only considers intra-prediction

Fig. 2.18 Illustration of the HEVC coding configurations

modes and avoid temporal references pictures through encoding, which leads to a sequence of independently decodable pictures, or Instantaneous Decoding Refresh (IDR) picture, as described in Figure 2.18a. The Low-Delay P and B (LDP and LDB) configuration is suitable for Low-Delay (LD) scenarios, the encoding begin with an IDR picture followed by B or P pictures encoded in temporal order, as illustrated in Figure 2.18c and 2.18b. The Random-Access (RA) configuration, suitable for broadcast application, enables open Group-of-Picture (GOP) composed of P or B frame, with possible access points clearing coding dependencies, as illustrated in Figure 2.18d. In addition, HEVC enables a coding structure where GOP are composed of hierarchically decodable temporal layers, illustrated as L0, L1, L2 and L3 in the Figure 2.18d. This enables to achieve temporal scalability through this structure with the possibility to drop the highest temporal layer L3, which is not used as reference during encoding process. It must be noted that the term GOP is also used in LD configurations to identify a group of pictures started from an Intra frame. The current implementation achieves high quality encoding thanks to an exhaustive evaluation among all the available modes through a Rate-Distortion Optimization process (RDO [51]). The RDO process consists in comparing the Rate-Distortion (RD) cost of several coding modes and to select the mode which minimizes this cost. The RD cost *J* is computed as the Lagrangian function which combines the bitrate *R* with a distortion measure *D* for a particular Lagrangian multiplier λ :

$$
J = D + \lambda \times R \tag{2.11}
$$

In the current HM implementation, the encoding process starts at the larger CU size (ie 64x64) and checks all the available coding modes, then the mode bringing the smallest RD cost is selected. After this first step, the CU is split into four sub-CU which are also tested. If the sum of sub-CU RD cost is smaller than the cost of the larger CU, then the split is selected and the process continues for each sub-CU until reaching the minimum authorised CU size. The Sum of Square Error (SSE), the Sum of Absolute Difference (SAD) and the Hadamard transformed SAD (SATD) are considered as distortion measures in HEVC, and described in the following Equations:

$$
D(i, j) = BlockA(i, j) - BlockB(i, j)
$$
\n(2.12)

$$
SSE = \sum_{i,j} D(i,j)^2 \tag{2.13}
$$

$$
SAD = \sum_{i,j} |D(i,j)| \tag{2.14}
$$

$$
SATD = \frac{1}{2} \sum_{i,j}^{\infty} |T(D)(i,j)|
$$
\n(2.15)

A different cost function can be applied depending on the encoding stage. The cost functions described in the Equations 2.16a and 2.16b are used for prediction parameter decision while the cost function in Equation 2.16c is used for mode decision. Two specific Lagrangian parameters λ_{pred} and λ_{mode} are used depending on the estimated function, and are both calculated according to the selected QP value. A weighting factor *wchroma*, also computed with the QP value, is applied to the SSE of chroma samples. Eventually, the cost of evaluated mode and prediction parameter in terms of bits are respectively called *Bmode* and *Bpred*.

$$
J_{Pred, SAD} = SAD + \lambda_{Pred} \times B_{Pred}
$$
 (2.16a)

$$
J_{Pred, SATD} = SATD + \lambda_{Pred} \times B_{Pred}
$$
 (2.16b)

$$
J_{Mode} = (SSE_{luma} + w_{chroma} \times SSE_{chroma}) + \lambda_{mode} \times B_{mode}
$$
 (2.16c)

Through the whole encoding process, each coding mode, and underlying coding parameters are evaluated based on their RD costs and selected so as to form the bitstream which provides the best trade-off between distortion and bitrate.

2.3.5 Performance and complexity

The initial goal of HEVC was to halve the bitrate compared to AVC for the same visual quality. Several objective and subjective performance comparisons have been achieved during standardization process but also in scientific literature. In [29], HEVC is compared to AVC on the usual JCT-VC dataset and showed that HEVC provides an average bitrate saving of 50%, according to subjective tests [30]. HEVC has also been tested on UHD content in [52] and showed an increased bitrate saving of 60% in average.

From complexity point of view, HEVC has also been analyzed. In [53], an HEVC complexity and implementation analysis is provided. It appears that, regarding Rate-Distortion complexity, HEVC in Main-Profile is 3.2x, 1.2x, 1.5x and 1.3x more complex than AVC High-Profile for AI, RA, LDB and LDP configurations, respectively [53]. On the other hand, and for the same profile, the HEVC decoding is respectively 2x, 1.6x, 1.5x and 1.4x more complex than an AVC for AI, RA, LDB and LDP configurations [53].

2.4 HEVC scalable extension

2.4.1 Standardization timeline

As mentioned previously, SHVC was standardized as an extension of HEVC. In Figure 2.19, the SHVC standardization timeline is provided. In July 2012, JCT-VC issued a Joint CfP on scalable video coding extension for HEVC [54]. Four months later, JCT-VC received twenty responses to the Joint CfP showing a significant achievable gain. These significant improvements motivated JCT-VC to start the SHVC standardization process. Two years later, the extension was approved as the Annex H and was published in October 2014 in HEVC version 2 with two profiles Scalable Main and Scalable Main 10.

Fig. 2.19 Timeline of SHVC, from Joint CfP to release

2.4.2 Overview of scalability tools

Types of scalability

SHVC [55] enables to perform spatial, color gamut, bitdepth, SNR and Codec scalability. We can observe in Figure 2.20 some illustrations of available types of scalability. In Figure 2.20a, the spatial scalability is described with 3-layer bitstreams composed of SDTV, HDTV and UHDTV services. In Figure 2.20b, the color-gamut scalability is also provided in a 3-layer scheme with a BT.709, DCI.P3 and BT.2020 gamuts. The case of codec scalability is provided in Figure 2.20c where the BL is encoded in AVC while the enhancement layer is compliant with HEVC. Finally, the PSNR scalability is addressed in Figure 2.20d where the bitstream is composed of several layers corresponding to different levels of quality. In addition, SHVC provides a bitdepth scalability where the different layers can use a different bitdepth, i.e a 8 bits BL with a 10 bits EL. Finally, a combination of the previously cited scalability can be achieved, for instance a AVC-HDTV BL with a supplemental HEVC-UHDTV EL.

Fig. 2.20 Illustration of differents types of scalability available in SHVC

Coding scheme

Regarding to the failure in SVC industrial adoption, caused by its late release and its extra complexity compared to AVC, the SHVC development was focused on simplicity. This way, an SHVC encoder is mainly composed of several HEVC encoders capable of extracting and handling inter-layer information such as MV and textures as Inter-Layer References (ILR). Thus, the modifications needed to turn HEVC encoders into an SHVC encoder are restricted to slice header level and above, the block level encoding core remaining unchanged. The extracted information is processed so as to be handled by upper layer, during interlayer processing which includes texture resampling, color-mapping and MV mapping. In Figure 2.21, an illustration of SHVC encoder is provided. We can observe that both HDTV and UHDTV signals are placed into encoder input and are respectively encoded by BL and EL HEVC encoders.

Fig. 2.21 Illustration of SHVC encoder in spatial scalability case

Inter-layer texture resampling

In the case of spatial scalability, an upsampling operation is performed on the base-layer reconstructed picture [56]. For luma samples, a 16-phases 8-tap filter is specified and applied during resampling process, while a 16-phases 4-tap filters is applied on chroma samples. These filters are described in Table 2.5 and are backward compatible with existing filters used during motion compensation interpolation, which results in 6 new additional filters. The up-samling filters enables an arbitrary ratio (from 1 to 2) between the resolutions of the two layers. In addition to this upsampling aspect, cropping can be specified for inter-layer reference before and after the resampling process through offset parameters signalled in the HLS.

Phase p	Luma $f_L[p, i]$								Chroma $f_L[p, k]$			
	$i=0$	$i =$	$i=2$	$i=3$	$i = 4$	$i=5$	$i=6$	$\overline{7}$ $i =$	$k=0$	$k=1$	$k=2$	$k=3$
Ω	Ω	Ω	Ω	64	Ω	Ω	Ω	θ	θ	64	Ω	Ω
$\mathbf{1}$	θ	1	-3	63	$\overline{4}$	-2	1	θ	-2	62	$\overline{4}$	$\mathbf{0}$
2	-1	2	-5	62	8	-3	1	Ω	-2	58	10	-2
3	-1	3	-8	60	13	-4	1	Ω	-4	56	14	-2
$\overline{4}$	-1	$\overline{4}$	-10	58	17	-5	1	Ω	-4	54	16	-2
5	-1	$\overline{4}$	-11	52	26	-8	3	-1	-6	52	20	-2
6	-1	3	-9	47	31	-10	$\overline{4}$	-1	-6	46	28	-4
$\overline{7}$	-1	$\overline{4}$	-11	45	34	-10	$\overline{4}$	-1	-4	42	30	-4
8	-1	4	-11	40	40	-11	4	-1	-4	36	36	-4
9	-1	4	-10	34	45	-11	4	-1	-4	30	42	-4
10	-1	$\overline{4}$	-10	31	47	-9	3	-1	-4	28	46	-6
11	-1	3	-8	26	52	-11	4	-1	-2	20	52	-6
12	Ω		-5	17	58	-10	4	-1	-2	16	54	-4
13	θ		-4	13	60	-8	3	-1	-2	14	56	-4
14	θ		-3	8	62	-5	2	-1	-2	10	58	-2
15	$\mathbf{0}$		-2	4	63	-3		$\mathbf{0}$	Ω	$\overline{4}$	62	-2

Table 2.5 Interpolation filter coefficients for luma and chroma samples

Inter-layer motion vectors mapping

To improve coding efficiency, inter-layer motion parameters can be used in SHVC by using the inter-layer reference picture as co-located reference in TMVP. Since 16x16 blocks are used during TMVP, this size is used during derivation of the motion field in the inter-layer reference picture. Let (xPCtr,yPCtr) be the position of the central sample of the current 16x16 block in resampled picture. In order to find the related motion information in the lower layer, the location (xPCtr,yPCtr) is scaled to the dimension of the lower layer according to the specified spatial scalability scaling factor and optional cropping informations. The obtained scaled position, denoted (xRef,yRef) is then rounded to fit with lower-layer 16x16

blocks alignment. The motion information located at the rounded position (xRL,yRL) is considered. Finally, the MVs of the BL are up-scaled by the scalability ratio to match with the resolution of the EL.

Inter-layer color mapping

If a BL and the EL are represented in different color-spaces, a color-mapping is performed during inter-layer processing [57]. As used in post-production environment, the colorgamut scalability (CGS) is based on the use of 3D Look-Up-Table (LUT3D) which links tri-chomatic location in the first space to its related position in a second space.

Fig. 2.22 Illustration of a LUT3D [57]

In SHVC, the LUT3D is represented by a 3D cuboid with granularity in each dimension. The partitioning of Y component is limited to 8 regions while both U and V components are limited to 2 regions. For each cuboid partition, the conversion from a space to the other one is performed under a matrix form, described in the following equation:

$$
\begin{bmatrix}\nY_{EL_Space} \\
U_{EL_Space} \\
V_{EL_Space}\n\end{bmatrix} = \begin{bmatrix}\na_Y & b_Y & c_Y \\
a_U & b_U & c_U \\
a_V & b_V & c_V\n\end{bmatrix} \times \begin{bmatrix}\nY_{BL_Space} \\
U_{BL_Space} \\
V_{BL_Space}\n\end{bmatrix}
$$
\n(2.17)

Thus, each partition in the cuboid has 12 matrix coefficients, which are transmitted through the bitstream in HLS signaling. In addition, the cuboid is uniformly partitioned in Y axis while the U/V partitions are non-uniform. Hence, the partition thresholds along U and V axis have to be transmitted and are also signalled in HLS.

2.4.3 Reference software

The SHVC Test Model (SHM) reference software is based on the same architecture as the HM software. For each specified layer, the SHM software instantiates a specific encoder derived from the HM. These encoders are capable of extracting and sending the inter layers information to the inter-layer processing unit, but also to receive and handle information which comes from the underlying layer as reference for encoding. In addition to these modifications, an inter-layer processing units is added to resample and map the information between layers. The upsampling filters previously described and the MV mapping process are implemented as well as the color mapping LUT3D. To enable efficiency in color-mapping, the SHM software performs an optimal LUT3D coefficient selection for each inter-layer reference. For each cuboid partition, the coefficients are selected after an optimization process which consists in minimizing the distortion between original and mapped samples. The optimization problem is described in the following Equation:

$$
\begin{bmatrix} a_Y & b_Y & c_Y \ a_U & b_U & c_U \ a_V & b_V & c_V \end{bmatrix}^{OPT} = \arg\min \{ \sum (Y_{src} - Y_{dst})^2 + \sum (U_{src} - U_{dst})^2 + \sum (V_{src} - V_{dst})^2 \}
$$
\n(2.18)

2.4.4 Performance and complexity

As for HEVC, SHVC performance has been checked during standardization process but has also been published recently [55], for the SHM7.0 reference software. It appears that SHVC can respectively provide 16.5%, 27%, 21% and 18.8% of bitrate reduction compared to simulcast for 2x-spatial, 1.5x-spatial, SNR, color-gamut and mixed Spatial-2x/CGS scalability in RA coding configuration. Compared to its predecessor SVC and according to the type of scalability, SHVC provides from 50% to 60% of bitrate reduction [55].

From complexity point of view, SHVC has been investigated at the decoder side [58]. It appears that SHVC introduces from 40% to 71% of additional complexity compared to simulcast configuration for SNR and spatial scalability. This additional complexity is significantly introduced by inter-layer prediction, for instance this inter-layer stage occupies up to 20% of the whole encoding time for spatial scalability while it represents only 7% for SNR scalability [58].

References

- [1] ITU-R, "Recommendation BT.709-5: Parameter Values for the HDTV Standards for Production and International Programme Exchange."
- [2] E.-J. Giorganni, T.-E. Madden, and M. Kriss, *Digital Color Management: Encoding Solutions*. Wiley, 2 ed., 2008.
- [3] ITU-R, "Recommendation BT.601-7: Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-Screen 16:9 Aspect Ratios."
- [4] ITU-R, "Recommendation BT.2020-1: Parameters Values of Ultra-High Definition Television Systems for Production and International Programme Exchange."
- [5] T. Smith and J. Guild, "The C.I.E. Colorimetric Standards and their Use," *Transactions of the Optical Society*, vol. 33, no. 3, p. 73, 1931.
- [6] "Wikipedia: White-point: [https://en.wikipedia.org/wiki/White_point.](https://en.wikipedia.org/wiki/White_point)"
- [7] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*. Natick, MA, USA: AK Peters (CRC Press), 2011.
- [8] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann (Elsevier), 2 ed.
- [9] DVB, "DVB Fact Sheet: Introduction to the DVB Project," May 2014.
- [10] ETSI, "TS-101-154: Digital Video Broadcasting (DVB), Specification for the use of Video and Audio Coding in Broadcasting Applications based on MPEG-2 Transport Stream."
- [11] ITU-R, "Recommendation BT.500-13: Methodology for the Subjective Assessment of the Quality of Television Pictures."
- [12] ITU-T, "Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications."
- [13] ITU-R, "Recommendation BT.1788: Methodology for the Subjective Assessment of Video Quality in Multimedia Applications."
- [14] J.-R. Ohm, G.-J. Sullivan, H. Schwarz, T.-K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards–Including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1669–1684, Dec. 2015.
- [15] Z. Wang, A.-C. Bonvik, H.-R. Sheikh, and E.-P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, Apr. 2004.
- [16] F. Xiao, "DCT-based Video Quality Evaluation." Final Project EE392J, Winter 2000.
- [17] G. Bjøntegaard, "VCEG-M33: Calculation of Average PSNR Differences Between RD-Curves," Apr. 2001.
- [18] CCITT, "Recommendation H.120: Codecs for Videoconferencing Using Primary Digital Group Transmission."
- [19] ITU-T, "Recommendation H.261: Video Codec for Audiovisual Services at $p \times 64$ kbits."
- [20] ISO/IEC, "International Standard 11172-2: Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbit/s – Part 2: Video."
- [21] ITU-T, "Recommendation H.262: Transmission of Non-Telephone Signals."
- [22] ISO/IEC, "International Standard 13818-2: Generic Coding of Moving Pictures and Associated Audio Information – Part2: Video."
- [23] ITU-T, "Recommendation H.263: Video Coding for Low Bit Rate Communication."
- [24] ISO/IEC, "International Standard 14496-2: Coding of Audio-Visual Objects Part 2: Visual."
- [25] ITU-T, "Recommendation H.264: Advanced Video Coding."
- [26] ISO/IEC, "International Standard 14496-10: Advanced Video Coding."
- [27] ITU-T, "Recommendation H.265: High Efficiency Video Coding."
- [28] ISO/IEC, "International Standard 23008-2: High Efficiency Video Coding."
- [29] J.-R. Ohm, G.-J. Sullivan, H. Schwarz, T.-K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards–Including High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1669–1684, Dec. 2012.
- [30] T.-K. Tan, R. Weerakkody, M. Mrak, and N. Ramzam, "Video Quality Evaluation Methodology and Verification Testing of HEVC Compression Performance," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 26, pp. 76–90, Jan. 2016.
- [31] "Xiph.Org Foundation: Theora Specification [http://www.theora.org/doc/Theora.pdf,](http://www.theora.org/doc/Theora.pdf)" 2011.
- [32] J. Bankoski, P. Wilkins, and Y. Xu, "Technical Overview of VP8, an Open Source Video Codec for the Web," *IEEE International Conference on Multimedia and Expo (ICME)*, 2011.
- [33] C. Feller, J. Wuenschmann, T. Roll, and A. Rothermel, "The VP8 Video Codec -Overview and Comparison to H.264/AVC," *IEEE International Conference on Consumer Electronic - Berlin (ICCE-Berlin)*, 2011.
- [34] "Webm-project: [http://www.webmproject.org/.](http://www.webmproject.org/)"
- [35] D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, and R. Bultje, "The Latest Open-Source Video Codec VP9 – An Overview and Preliminary Results," *Picture Coding Symposium (PCS)*, 2013.
- [36] H. Kalva and J.-B. Lee, "The VC-1 Video Coding Standard," *IEEE Multimedia*, vol. 14, pp. 80–91, Oct. 2007.
- [37] "Xiph.org foundation: [https://www.xiph.org/daala/.](https://www.xiph.org/daala/)"
- [38] T. Wiegand, J.-R. Ohm, G.-J. Sullivan, W.-J. Han, R. Joshi, T.-K. Tan, and K. Ugur, "Special Section on the Joint Call for Proposals on High Efficiency Video Coding (HEVC) Standardization," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 20, pp. 1661–1666, Dec. 2010.
- [39] G.-J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1649–1668, Dec. 2012.
- [40] C.-C. Chi, M. Alvarez-Mesa, B. Juurlink, G. Clare, F. Henry, S. Pateux, and T. Schierl, "Parallel Scalability and Efficiency of HEVC Parallelization Approaches," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1827–1838, Dec. 2012.
- [41] T.-K. Heng, W. Asano, T. Itoh, A. Tanizawa, J. Yamaguchi, T. Matsuo, and T. Kodama, "A Highly Parallelized H.265/HEVC Real-Time UHD Software Encoder," *IEEE International Conference on Image Processing (ICIP)*, Oct. 2014.
- [42] A.-M. Kotra, M. Raulet, and O. Deforges, "Efficient Parallelization of Different HEVC Decoding Stages," *Data Compression Conference (DCC)*, 2013.
- [43] I.-K. Kim, J. Min, T. Lee, W.-J. Han, and J. Park, "Block Partitioning Structure in the HEVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1697–706, Dec. 2012.
- [44] J. Lainema, F. Bossen, W.-J. Han, J. Min, and K. Ugur, "Intra Coding of the HEVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1792–1801, Dec. 2012.
- [45] P. Helle, S. Oudin, B. Bross, D. Marpe, M.-O. Bici, K. Ugur, J. Jung, G. Clare, and T. Wiegand, "Block Merging for Quadtree-Based Partitioning in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1720– 1731, Dec. 2012.
- [46] J. Sole, R. Joshi, N. Nguyen, T. Ji, M. Karczewicz, G. Clare, F. Henry, and A. Duenas, "Transform Coefficient Coding in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1765–1777, Dec. 2012.
- [47] A. Norkin, G. Bjøntegaard, A. Fuldseth, M. Narroschke, K.-A. Ikeda, M. Zhou, and G. Van der Auwera, "HEVC Deblocking Filter," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1746–1754, Dec. 2012.
- [48] C.-M. Fu, E. Alshina, A. Alshin, Y.-W. Huang, C.-Y. Chen, C.-Y. Tsai, C.-W. Hsu, S.-M. Lei, J.-H. Park, and W.-J. Han, "Sample Adaptive Offset in the HEVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1755–1764, Dec. 2012.
- [49] V. Sze and M. Budagavi, "High Throughput CABAC Entropy Coding in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1778–1791, Oct. 2012.
- [50] C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G. Sullivan, "JCTVC-T1002: High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Improved Encoder Description Update 2," Feb. 2015.
- [51] G.-J. Sullivan and T. Wiegand, "Rate-Distorsion Optimization fo Video Compression," *IEEE Signal Processing Magazine*, pp. 1755–1764, Nov. 1998.
- [52] R.-W. Marta Mrak, V. Baroncini, J.-R. Ohm, T.-K. Tan, and G.-J. Sullivan, "Verification Testing of HEVC Compression Performance for UHD Video," *IEEE GlobalSIP*, 2014.
- [53] J. Vanne, M. Viitanen, T.-D. Hämäläinen, and A. Hallapuro, "Comparative Rate-Distorsion-Complexity Analysis of HEVC and AVC Video Codecs," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1885–1898, Dec. 2012.
- [54] ISO/IEC JTC 1/SC 29/WG 11 and ITU-T SG 16 WP 3, "N12957: Joint Call for Proposal on Scalable Video Coding Extensions of High Efficiency Video Coding (HEVC)," July 2012.
- [55] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, Dec. 2015.
- [56] E. Alshina, A. Alshin, Y. Cho, J.-H. Park, W. Pu, J. Chen, X. Li, V. Seregin, and M. Karczewicz, "Inter-Layer Filtering for Scalable Extension of HEVC," *Picture Coding Symposium (PCS)*, 2013.
- [57] P. Bordes, p. Andrivon, and R. Zakizadeh, "Color Gamut Scalable Video Coding for SHVC," *Picture Coding Symposium (PCS)*, 2013.
- [58] W. Hamidouche, M. Raulet, and O. Deforges, "4K Real-Time and Parallel Software Video Decoder for Multilayer HEVC Extensions," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 26, no. 1, pp. 169–180, 2015.

Chapter 3

Rate control for video coding

3.1 Introduction

Rate-control (RC) and adaptive rate control (ARC) are major stakes in video applications. In broadcasting environment, programs are often jointly encoded into a fixed-bandwidth channel by taking advantages of statistical-multiplexing techniques. For storage applications such as DVD or Blu-Ray, a fixed available volume has to be filled by video content. In over-the-top (OTT) content delivery, contents are encoded and stored in a multiple versions of different qualities and bitrates so as to address users with different bandwidth requirements. For all these applications, the video encoder needs to be equipped with a bitrate controller which achieves the targeted bitrate. According to the application, several rate control techniques can be used based on constant bitrate (CBR) or variable bitrate (VBR) control. In addition, an RC algorithm can perform one or multiple encoding passes to reach the targeted bitrate. In the following subsections, the major concepts used in rate-control will be introduced before a deeper discussion in Sections 3.2, 3.3 and 3.4.

3.1.1 Stakes

Rate control algorithms are mainly composed of three steps which can be represented at the top of an existing encoder, as illustrated in Figure 3.1. In this configuration, the video encoder compresses the video signal into a bitstream subject to some bitrate constraints. These constraints can be for example a targeted bitrate, a constraint on a virtual buffer occupancy, an authorized deviation or a selected granularity. The first and most important step consists in considering all the constraints and bitrate instructions to allocate the available number of bits according to the selected granularity (GOP, frame, CTU). To optimize the allocation, the picture itself and its global or local complexities can be analysed so as to

allocate more bits to the most complex and hard-to-encode regions. Once this step is achieved, the relations between the targeted bitrate and some controllable encoding parameters (QP, λ) have to be estimated. Various models and techniques can be used according to the type of content, frames, CTU and even to the video-signal itself. With the estimated models, adjustable encoding parameters such as QP or Lagrangian multiplier λ are selected in the most optimal way, so as to reach the targeted bitrate. Since encoding behavior is hard to predict and may change all along the encoding, a feedback is usually sent to the previously mentioned steps, so as to refine the models and fix the bitrate mismatches. This feedback can be composed of information about achieved bitrates and distortions, coefficient distributions, statistics about predictions, etc.

Fig. 3.1 Rate-control steps

3.1.2 Single and multi-pass rate control

The three previously described steps can be used in single or multi pass encoding according to the tolerance level of the system, and are described in Figure 3.2. In application such as video-conferencing where the encoding delay has to be minimized, single-pass approaches are only considered. For off-line encoding, multi-passes encoding can be used since there are no delay constraints. One the one hand, we can notice that a single-pass encoding consists in achieving the best possible encoding and to compensate the bitrate mismatch thereafter. On the other hand, multi-pass encoding enables to produce a perfectly tuned encoding respecting all the bitrate instructions.

Fig. 3.2 Single/Multi-Pass Strategies

3.1.3 Constant and adaptive rate control

Some applications require a joint-encoding of several video contents. In broadcasting environment, several programs have to fit into a fixed-bandwidth channel. In this case, the complexity of each program is evaluated which enables to weight the program allocated bitrate in a dynamic way. This is particularly interesting and useful when a consistent quality is needed among several programs jointly encoded. The most complex programs, which require more bitrate, are favoured contrary to less complex contents. For live-streaming applications, congestion on the networks may force the encoder to lighten the targeted bitrate also in a dynamic way. These applications imply that targeted bitrates vary during the encoding process, which rely on VBR encoding. On the contrary, CBR approaches are also used for instance in storage applications where the support size will not evolve during encoding process.

3.2 Bitrate allocation approaches

For a SL encoder, rate control is often a matter of reaching a targeted bitrate R_T for a given number of frames N_f with an associated framerate f_r , subject to CBR or VBR allocation. Various contributions have been published, and can be classified according to the wished level of granularity. In this Section, the state of the art on bitrate allocation is provided for GOP, frame, region-of-interest (ROI) and block level. We also introduce the statistical multiplexing (StatMux) that consists in a pool of programs efficiently and jointly encoded. In addition, some RC algorithms often consider virtual buffer verifier (VBV) to avoid overflow or underflow at the decoder side are presented.

3.2.1 GOP-level

There are several methods to handle GOP-level bitrate allocation, the most simple approaches do not considers the VBV constraints. For each GOP in the sequence, the number of frames in the i^{th} GOP is noted $N_{GOP}(i)$. First, the number of bits to reach for the i^{th} GOP in the sequence is simply computed as:

$$
T_{GOP}(i) = \frac{R_r(k)}{f_r} \times N_{GOP}(i)
$$
\n(3.1)

where $R_r(k)$ refers to the updated targeted bitrate for the rest of sequence from the k^{th} frame (typically the first one in the current GOP), which is initially equal to R_T and can be thereafter computed for instance as:

$$
R_r(k) = \frac{\frac{R_T \times N_f}{f_r} - B_{Write}}{f_r} \times (N_f - N_{Code})
$$
\n(3.2)

where *NCoded* and *BWrite* refer to the number of already coded frames and the number of consumed bits until the *k th* frame, respectively. Then, to avoid abrupt variation of bitrate among the successive GOP, the targeted bitrate is often smoothed as follows:

$$
T'_{GOP}(i) = \alpha \times T_{GOP}(i) + (1 - \alpha) \times B_{GOP}(i - 1)
$$
\n(3.3)

where α is the smoothing factor (between 0 and 1, 0.5 typically), and $B_{GOP}(i-1)$ is the number of bits allocated for in the previous (*i*−1) *th* GOP. This approach, applied for HEVC in [1–3], is enhanced in [4]. In [4], the smoothing process described in Equation 3.3 is improved by introducing a dynamic adaptation, in the case of RA where intra periods are separated by several GOP. This adaptation exploits the weighting factors w_i^j l_l of each frame

of type *t*, with $t \in \{I, P, B\}$, and associated to a temporal level *l*. For a given frame in *GOP*(*i*−1), this weighting factor is computed as the ratio of bits produced by this frame to the number of bits produced by the remaining P-frames. This more advanced smoothing is described in the following Equation:

$$
T'_{GOP}(i) = \alpha \times \frac{W_{GOP}(i)}{W_{Remaining GOP}} \times T_{GOP} + (1 - \alpha) \times \frac{W_{GOP}(i)}{W_{GOP}(i-1)} \times B_{GOP}(i-1)
$$
(3.4)

where $W_{GOP}(i)$ refers to the sum of w_i^i $\frac{d}{dt}$ in the *i*th GOP and *W_{RemainingGOP* the sum of w_i^t} l ^t in the remaining GOP before the next intra period. For HEVC, several approaches considering VBV also exist and enable to avoid video buffer underflow or overflow at the decoder side. For this issue, the GOP-level bitrate is generally updated for each *j th* frame in the GOP which leads to two-parameters GOP budget $T_{GOP}(i, j)$. In [5], the most simple method is described. If we consider $V_i(j)$ the state (i.e. the number of bits in the buffer) of the video buffer at the *j*th frame of the *i*th GOP, the GOP-level bitrate allocation is performed:

$$
T_{GOP}(i,j) = \begin{cases} \left(\frac{R_r}{f_r} - \beta \times \frac{V_i(1)}{f_r}\right) \times N_{GOP} & , j = 1\\ T_{GOP}(i,j-1) - b_i(j-1) & , j = 2...N_{GOP} \end{cases}
$$
(3.5)

where β is a convergence speed parameter and $b_i(j-1)$ the number of bits consumed in the $(j-1)$ th frame of the *i*th GOP. In [6], an alternative VBV management is proposed, which considers the bitrate variation through the GOP encoding, but which does not offer converge speed adjustment. This alternative approach is defined as follows:

$$
T_{GOP}(i,j) = \begin{cases} \frac{R_T}{f_r} \times N_{GOP} - V_i(j) & , j = 1\\ T_{GOP}(i,j+1) + \frac{R_r(j) - R_r(j-1)}{f_r} \times (N_{GOP} - j + 1) - b_i(j-1) & , j = 2...N_{GOP} \end{cases}
$$
(3.6)

In [7], a GOP-level VBV is considered and propose to allocate the budget of the *j th* GOP according to its distortion indicator compared to the remaining GOP in the sequence:

$$
T_{GOP}(i,j) = \frac{SATD(i,j)}{\sum_{k \in \{RemainingGOP\}} SATD(i,k)} \times V_i(j)
$$
(3.7)

Eventually, several approaches such as [8] and [9] are not directly using VBV but consider a smoothing window (SW) to manage buffer occupancy and picture quality [10]. In this approach, the GOP budget is computed in the following way:

$$
T_{GOP}(i) = \frac{R_T \times (N_{Coded} + L_{SW}) - B_{Write} \times f_r}{L_{SW} \times f_r} \times N_{GOP}(i)
$$
(3.8)

where L_{SW} refers to the length of the SW. These methods are only the first steps of bitrateallocation which often encapsulate frame, ROI and block level allocation, described in the following Sections.

3.2.2 Frame-level

The frame-level bitrate allocation may vary, according to the coding configuration. For all-intra video coding, an approach has been developed in [11] based on a simple bitrate target updating considering VBV:

$$
T_{Pic}(j) = \frac{R_T}{f_r} + \left(\frac{V_{Size}}{2} - V(j)\right)
$$
\n(3.9)

Where V_{Size} is the buffer capacity in bits and $V(j)$ the buffer occupancy at the jth frame. This straightforward approach has also been used for LD in [12]. In [1–3] the GOP-budget is weighted according to the picture position in the hierarchical GOP. This approach enables a basic and simple bitrate allocation while considering the importance of each picture:

$$
T_{Pic}(i,j) = \alpha_{i,j} \times T_{GOP}(i,j)
$$
\n(3.10)

where $\alpha_{i,j}$ represents the weighting factor which may vary according to the temporal level of the frame in the GOP. This method is slightly enhanced in [4], considering the previously described *w t* l_i for computing $\alpha_{i,j}$. Other approaches have been investigated with VBV consideration. The simplest way of achieving frame-level bitrate-allocation is considered in [5, 13], for LD applications, where a weighted mean is performed on GOP-level allocated bitrate and buffer occupancy:

$$
T_{Pic}(i,j) = \beta \times \left(\frac{T_{GOP}(i,j)}{N_{Left}}\right) + (1-\beta) \times \left(\frac{R_T}{f_r} + \gamma \times (L-V_i(j))\right)
$$
(3.11)

where γ is a constant and β balances the distribution between GOP and buffer allocated bitrate (for example $\beta = 0.75$ and $\gamma = 0.2$). *N*_{Left} represents the number of non-coded pictures in the GOP and *L* the targeted buffer level. In a similar way, this method is developed in [14, 15], with on-the-fly T_{GOP} computation based on per-frame weighting factors (i.e. [4]). This approach is enhanced in [16] where the GOP-level allocation is avoided, and where coding-complexity and header size predictions are performed. First, the targeted bitrate for

each frame is computed according to its type:

$$
T_{Pic}(i,j) = \begin{cases} T'_{Pic}(i,j) & , if I picture \\ \beta \times T'_{Pic}(i,j) + (1-\beta) \times \left(\frac{R_T}{f_r} + \gamma \times (\widetilde{V}_i(j+1) - V_i(j))\right) & , otherwise \end{cases}
$$
(3.12)

with

$$
T'_{Pic}(i,j) = \frac{C_{i,j}}{\sum_{p \in GOP} C_{i,p}} \times \left(\frac{R_T \times N_f}{f_r} - B_{Write} - H(i,j)\right) + h_{i,j}
$$
(3.13)

where $V_i(j+1)$, $C_{i,j}$, $H_{i,j}$ and $h_{i,j}$ are respectively the prediction of buffer occupancy at the $(j+1)$ th frame, the complexity estimation of the *j*th frame of the *i*th GOP, the headers cost prediction for the rest of the GOP and the header cost prediction for the *j th* picture. It must be noted that this method considers separate models for parameters estimation according to the picture position in the hierarchical GOP. In [17], another hierarchical bitrate-allocation is proposed for GOP containing two temporal layers on the top of IDR pictures, in this approach each layer has a specific allocation scheme described as follows:

$$
T_{Pic}(i,j) = \begin{cases} T_I = T_{GOP}(i,j) \times \alpha & , if I picture \\ T_{LO} = \frac{T_{GOP}(i,j) \times (1-\alpha) \times \beta}{N_0} & , if Level 0 \\ T_{L1} = \frac{T_{GOP}(i,j) - T_I - T_{LO} \times N_0}{N_{Left}}, otherwise \end{cases}
$$
(3.14)

where α and β are constants and N_0 and N_{Left} respectively the number of remaing picture belonging to layer 0 and the number of remaining pictures in the GOP. Eventually, [9] provides a simple approach which considers predefined weighting factors varying according to the targeted Bits-per-Pixel (bpp) range, defined as follows:

$$
T_{Pic}(i,j) = \frac{T_{GOP}(i,j) - Coded_{GOP}(i)}{\sum_{NotCode dPictures} \omega_{Pic}} \times \omega_{CurrPic}
$$
\n(3.15)

where *Coded_{GOP(i)}* is the number of already written bits in the GOP. The weighting factors are described in Table 3.1.

Position	bpp							
	[0, 0.05]	[0.05, 0.1]	[0.1, 0.2]	$ 0.2, +\infty $				
Level 0								
Level 1								
Level 2								
Level 3								

Table 3.1 Example of ω*Pic* weighting factors values for 3-Layers RA hierarchical GOP [9]

In addition to these frame-based approaches, several algorithms have been proposed for ROI rate-control, especially for HEVC in [18, 19]. This method consists in enhancing particular regions of the picture with higher bitrate than no ROI regions, in the following way:

$$
T_{\overline{ROI}}(i,j) = \frac{P_{\overline{ROI}}(i,j)}{M \times (1 + P_{ROI}(i,j) \times (K-1))} \times T_{Pic}(i,j)
$$
\n(3.16)

$$
T_{ROI}(i, j) = T_{Pic}(i, j) - T_{\overline{ROI}}(i, j)
$$
\n
$$
(3.17)
$$

where *TROI* and *TROI* are respectively the targeted number of bits in the ROI and non-ROI regions of the jth frame of the ith GOP. *P_{ROI}* and *P_{ROI}* are the area of ROI and non-ROI regions (for example 0.2). *K* is defined as the ratio between ROI and non-ROI regions defined as $R_{ROI} = K \times R_{\overline{ROI}}$, and *M* the total number of pixels in the *j*th frame.

3.2.3 Block-level

The most basic approach for sharing bitrate among every block in the picture consists in performing an equal bitrate distribution. In [17], this approach is applied for HEVC where the picture-level allocated bitrate is equally shared among all blocks and update for each *b th* block through the encoding process:

$$
T_B(i, j, b) = \frac{T_{Pic}(i, j) - H(i, j) - Coded_{Pic}(i, j, b)}{N_B - N_{CB}}
$$
(3.18)

where $H(i, j)$ is the number of bits already consumed in header signaling in the jth picture of the *i*th GOP, *Coded_{Pic}*(*i*, *j*,*b*) is number of already consumed bit until the *b*th block. *N_B* and *N_{CB}* refer to the number of blocks and to the already coded blocks in the picture, respectively. This approach is enhanced in several contributions, where a particular weighting factor is applied to each block according to a quality or complexity criterion. The bit allocation of the bth block is then computed in the following way, considering the weight of the current block compared to the remaining ones:

$$
T_B(i,j,b) = \frac{(T_{Pic}(i,j) - H(i,j) - Coded_{Pic}(i,j,b)) \times w_b}{\sum_{k \in NotCodedblocks} w_k}
$$
(3.19)

In [9, 20] and [14, 21], a content dependent weighting factor evaluation is considered, based on the predicted mean absolution difference (MAD) of each CTU:

$$
w_b = \begin{cases} \quad MAD_b^2 & , \text{ in } [9, 20] \\ \quad MAD_b & , \text{ in } [14, 21] \end{cases} \tag{3.20}
$$

An alternative content dependent approach based on predicted-SSIM value is proposed in [22], where the weighting factor is computed as follows:

$$
w_b = (1 - SSIM_b)^2
$$
 (3.21)

In a similar way, a pre-encoding based method is described in [23] for HEVC. In this method, a first encoding is performed on the content with only 16x16 CU, the corresponding bitrate are then stored as $R_{16x16}(i, j, b)$. These stored values are then used in the second encoding pass, with the following weighting factors:

$$
w_b = R_{16 \times 16}(i, j, b) \tag{3.22}
$$

A quite different approach considering SW and gradient-based content descriptor is proposed in [24] for HEVC intra coding. First, the gradient per pixel (GPP) is computed on each frame:

$$
GPP = \frac{1}{H \times W} \sum_{k=0}^{H-1} \sum_{l=0}^{W-1} \left(|I_{k,l} - I_{k+1,l}| + |I_{k,l} - I_{k,l+1}| \right) \tag{3.23}
$$

Where $I_{k,l}$ is the luminance pixel value at the position (k,l) . Then, the initial budget of each block is estimated according to its gradient activity:

$$
T_B(j,b) = \frac{GPP_b}{\sum_{k \in All Blocks} GPP_k} \times T_{Pic}(j)
$$
\n(3.24)

The final CTU budget is computed as follows:

$$
T'_{B}(j,b) = \left[T_{Pic}(j) - Coded_{Pic}(j,b) + \frac{\sum_{k=1}^{b-1} (T_{B}(j,k) - Coded_{B}(j,k)))}{L_{SW}} \right] \times w_{b}
$$
 (3.25)

with

$$
w_b = \frac{GPP_b}{\sum_{k \in NotCodedBlocks} GPP_k}
$$
(3.26)

where $Codedg(j,k)$ represents the number of coded bits in the k^{th} block of the j^{th} picture. In the same way as GOP or frame level bitrate allocation, algorithms considering VBV are also proposed for block-level allocation. As for picture-level allocation, these approaches mix frame-budget and buffer consideration, as follows:

$$
T_{Block}(i, j, b) = \beta \times (T_{Pic}(i, j) - Coded_{Pic}(i, j, b)) \times \frac{w_b}{\sum_{k=b}^{N_B} w_k} + (1 - \beta) \times T_{Buff}(i, j, b)
$$
\n(3.27)

with

$$
T_{Buff}(i,j,b) = T_{Pic}(i,j) \times \frac{w_b}{\sum_{k=1}^{N_B} w_k} - \frac{V_i(j,b)}{N_B - N_{CB}}
$$
(3.28)

and

$$
w_b = \begin{cases} N_{\text{pixels}}(b) & \text{, in [6]}\\ x_b \times \lambda^{y_b} & \text{, in [5]} \end{cases} \tag{3.29}
$$

where $N_{\text{pixels}}(b)$ refers to the number of pixel in the *bth* block. x_b and y_b are linked with the Rate- λ modeling of each CTU through the encoding process ($R = x \times \lambda^y$). This Rate- λ modeling is also used in [25] where the targeted bitrate is estimated according to the following optimization problem:

$$
D_{Pic}(i,j) = \min_{\{T_{Block}(i,j,b)\}_{b=1}^{N_B}} \sum_{b=1}^{N_B} d_{Block}(i,j,b) \quad s.t. \sum_{b=1}^{N_B} T_{Block}(i,j,b) \le T_{Pic}(i,j) \tag{3.30}
$$

where $D_{Pic}(i, j)$ and $d_{Block}(i, j, b)$ are the distortion in the jth picture of the ith GOP and the distortion in the *b th* block, respectively. This expression can be turned into an unconstrained optimization problem:

$$
\min_{\{T_{Block}(i,j,b)\}_{b=1}^{N_B}} \sum_{b=1}^{N_B} \left((d_{Block} \circ T_{Block})(i,j,b) + \lambda \times T_{Block}(i,j,b) \right) \tag{3.31}
$$

This equation can be solved by setting the derivative to zero which leads to the following solution:

$$
\lambda = \frac{\partial d_{Block}(i,j,b)}{\partial T_{Block}(i,j,b)} \text{ and } \sum_{b=1}^{N_B} T_{Block}(i,j,b) = T_{Pic}(i,j) \quad , b = 0, 1, 2...N_B \tag{3.32}
$$

Combining this solution with an Hyperbolic model for R-D function $d_{Block}(i, j, b) = c_b \times c_b$ $T_{Block}(i, j, b)^{-k_b}$ leads to the following solution formulation:

$$
\sum_{b=1}^{N_B} T_{Block}(i, j, b) = \sum_{b=1}^{N_B} \left(\frac{c_b \times k_b}{\lambda}\right)^{-\frac{1}{k_b+1}} = T_{Pic}(i, j)
$$
(3.33)

Thus, the solution is found thanks to a recursive Taylor expansion which enables to achieve the optimal bitrate allocation at block-level. Finally, a game-theory (GT) approach has been

investigated in [26] for MPEG-4. This approach consists in considering each block in a given picture as a player which is a competitor for sharing the available frame-level budget. In this game, each player plays a particular strategy which refers to the amount of bits he will ask, in addition a utility function is defined for each one representing its preference in terms of visual quality (or distortion). This way, $\langle U, d \rangle$ can be defined as the game configuration, where *U* represents the set of achievable utilities and $d = (d_1, d_2, ..., d_{N_B})$ the desirable quality of the game regrouping the distortion of each player with $u_b > d_b$. According to the Nash's GT [27], there is a particular solution called Nash bargaining solution (NBS) which guarantees a fair budget allocation, this NBS is used to allocate a bit budget for each block in the picture.

3.2.4 Scalable-based enhanced approaches

For scalable video coding, the existing correlation between layers offers prospective for enhanced RC algorithms. The existing contributions are mostly dedicated to enhanced encoding parameters estimation based on inter-layer indicators, but several bitrate allocation algorithms exist. The most simple method consists in considering each scalable layer as a single-layer stream and to allocate the bitrate budget independently from other layers. This simple approach is implemented for instance in [28], [29] and [30]. Another original contribution based on GT and especially NBS is treated in [31]. In a similar way as [26], this approach considers frame instead of block as a player in the game and still with distortion as utility function, in addition inter-layer based R-D models are used. Finally, a multi-buffer based framework is proposed in [32] for SVC where per-layer buffer and targeted bitrates are considered.

Fig. 3.3 Illustration of CBR versus VBR-StatMux approaches

3.2.5 Joint allocation: the statistical multiplexing case

In broadcasting environment, programs are often jointly encoded using StatMux techniques which exploit statistical properties of the content to reduce the needed bandwidth without

quality deterioration, as illustrated in Figure 3.3. This is strongly different from traditional approaches where programs are separately encoded and send in separate channels. The StatMux approaches for SL have been widely treated in literature, for instance in [33, 34]. The general approach consists in computing a complexity indicator for each program and then to allocate more bis to the most complex program, and vice-versa. For scalable video coding, several algorithms have been proposed. In [35], a StatMux algorithm for joint programs encoding using MPEG-4 fine granularity scalability (FGS) is proposed. For the *v th* video program, the *j th* GOP base-layer bitrate is computed as:

$$
R_{BL_GOP}(v,j) = R_{BL} \times \left[\alpha \times \frac{SA_{Pic}(v,j)}{\sum_{k=1}^{N_{Prog}} SA_{Pic}(k,j)} + (1-\alpha) \times \frac{MA_{Pic}(v,j)}{\sum_{k=1}^{N_{Prog}} MA_{Pic}(k,j)} \right]
$$
(3.34)

where $SA_{Pic}(v, j)$ and $MA_{Pic}(v, j)$ are the spatial and motion activities for the jth GOP in the vth program, respectively. *R_{BL}* represents the BL dedicated bitrate and α the proportion of bitrate concentrated in the intra frame in the BL GOP. Once the BL bitrates are computed, the FGS layers can be truncated to fit into the channel bandwidth:

$$
\sum_{\nu=1}^{N_{Prog}} R_{\nu,j} \le R_{Channel} \tag{3.35}
$$

Since FGS enables a fine refining in the EL, the global truncation bitrate of each program (BL+EL) will be estimated so as to minimize the distortion difference between programs:

$$
\min_{R_{\nu,j}} Var = \frac{1}{N_{Prog}} \times \sum_{\nu=1}^{N_{Prog}} (MSE_{\nu,j} - \overline{MSE_j})^2 \quad s.t. \quad \sum_{\nu=1}^{N_{Prog}} R_{\nu,j} \le R_{Channel} \tag{3.36}
$$

Once the bitratres are set, the EL bitstreams are truncated to form the optimal multiplexed bitstream containing all the programs. This approach has been also explored in [36] but considers the PSNR instead of MSE and also in [37] where other complexity metrics are considered. In [38, 39], an enhanced method is proposed based on improved complexity indicators, a temporal indicator derived from SSIM called temporal complexity (TCX) and a gradient-based spatial complexity indicator (SCX) are used:

$$
R_{BL_GOP}(v, j) = R_{BL} \times \frac{TCX_{v,j} + SCX_{v,j}}{\sum_{k=1}^{N_{Prog}} (TCX_{k,j} + SCX_{k,j})}
$$
(3.37)

In addition, a smoothing factor is introduced at GOP-level allocation to avoid abrupt bitrate fluctuations. In [40], a different approach is proposed where BL programs and the EL are jointly encoded. In a similar way, the BL bitrates are computed thanks to a complexity indicator based on MSE, for the *m th* frame:

$$
R_{BL_GOP}(v,m) = R_{BL} \times \left(\frac{C_{v,m}^{BL}}{\sum_{k=1}^{N_{Prog}} C_{k,m}^{BL}}\right)
$$
(3.38)

where $C_{v,m}^{BL}$ is the BL complexity indicator of the m^{th} frame of the v^{th} program. In a similar way, the EL bitrate is computed based on complexity indicators:

$$
R_{EL_GOP}(v,m) = R_{EL} \times \left(\frac{C_{v,m}^{EL} + \widetilde{C}_{v,m}^{BL}}{\sum_{k=1}^{N_{Prog}} (C_{k,m}^{EL} + \widetilde{C}_{k,m}^{BL})}\right)
$$
(3.39)

where $C_{v,m}^{EL}$ is the EL complexity indicator of the m^{th} frame of the v^{th} program. $\tilde{C}_{v,m}^{BL}$ is an adapted BL complexity indicator which enables an even video quality among EL.

3.3 Encoding parameters selection

Once the bit budget is allocated, the following step consists in estimating the encoding parameters which enable to reach the targeted bitrate. Historically, tweaking the QP was the first way of reaching a desirable bitrate, which leads to model functions linking bitrate and quantization step (or R-Q). This model can be inaccurate or hard to model. Thus, alternative models such as ρ -domain has been explored (R- ρ). This approach consists in passing through an intermediate and more simple linear model to avoid the modeling of R-Q relationship. This model links the bitrate *R* with the number of non-zeros residual coefficients in a linear way and is thus easier to build. These approaches were efficient but might be not adapted to the most recent standards where the header bitrate has grown significantly and where the residual bitrate is reduced, due to better and more efficient predictions. In addition, the emerging of RDO-based encoding led to consider models based on Lagrangian multiplier adjustment which has a crucial role during encoding $(R-\lambda)$. All these approaches have been tackled in HEVC and are discussed in the following sub-Sections, as well as enhanced versions explored for scalable schemes.

3.3.1 R-*Q* approaches

The R−*Q* approach consists in modeling the bitrate as a function of the QP. This can be achieved by various ways, investigated in the literature, and thus enables to reach a targeted bitrate value through QP value adjustment. The simplest method is incremental, as proposed in [11] for frame-level rate control in AVC. In this approach the accumulated bitrates are compared to the initial allocated bitrate and the QP is adjusted, for instance as follows:

$$
QP = \begin{cases} QP - 1, & if \Delta_{Br} > \gamma \\ QP + 1, & if \Delta_{Br} < -\gamma \end{cases}
$$
 (3.40)

where ∆*Br* is the measured difference between reached and targeted bitrate in the picture, and γ is the selected threshold . Another simple way of estimating the QP related to a targeted bitrate is used in [12] and [41], where the frame complexity *CPic* is considered through SATD metric:

$$
R_{Pic} = \frac{\alpha \times C_{Pic}}{QP}
$$
 (3.41)

where α is a model parameter. A more complex quadratic modeling has been introduced in [26] for block-level rate control in MPEG-4 video coding and considers the standard deviation of prediction errors in the model:

$$
\frac{R_{Block}(b)}{m_b^{\alpha}} = \frac{K}{QP_b^2}
$$
\n(3.42)

where m_b is the standard deviation of prediction errors in the b^{th} block, *K* a model parameter, α a constant and QP_b the quantization step of the block. A similar and more accurate approach has been introduced in [21], also at block level:

$$
R_{Block}(b) = MAD(b) \times \frac{K_1}{QP_b} + MAD(b) \times \frac{K_2}{QP_b^2}
$$
\n(3.43)

where K_1 and K_2 are the model parameters and MAD_b refer to the predicted MAD value of the *b th* block. Very similar approaches have also been investigated in [2] and [14]. In [15], this approach has been enhanced for HEVC for both frame or CTU level and also considered the number of pixels:

$$
\frac{R_{Block}(b)}{N_{Pixels}(b)} = MAD(b) \times \frac{K_1}{QP_b} + MAD(b) \times \frac{K_2}{QP_b^2}
$$
\n(3.44)

where *N*_{Pixels} is the number of pixels in the block. It must be noted that blocks can be replaced by a frame in Equation 3.44. A method based on Cauchy function –firstly introduced in [42]– also exists and is introduced for HEVC in [16], first for the lower picture in the hierarchical GOP structure:

$$
R_{Pic} = a \times Q P_{Pic}^{-\alpha} \tag{3.45}
$$

where (a, α) are the model parameters. For the upper levels in the GOP, the selected QP is clipped and adapted according to the importance of the picture and other buffer considerations. A different approach considering log-exp modeling is proposed in [20] for CTU-level RC, where the R-Q model is formed as follows:

$$
QP_b = \eta \times ln(R_{Block}(b)) + \gamma \tag{3.46}
$$

where (η, γ) are the model parameters. This approach has been enhanced in [13] where the R-Q model consider low, medium and highly complex CTU:

$$
R_{Block}(b) = \alpha_L N_L \times e^{-(1-f)QP_b \eta_L} + \alpha_M N_M \times e^{-(1-f)QP_b \eta_M} + \alpha_H N_H \times e^{-(1-f)QP_b \eta_H}
$$
(3.47)

where $(\alpha_L, \alpha_M, \alpha_H)$, (η_L, η_M, η_H) and (N_L, N_M, N_H) are respectively the (α, η) model parameters and the block sizes, for low, medium and highly complex CTU, and where *f* is a rounding offset. Finally, other approaches consist in first estimating picture-level QP and then locally improve specific CTU, with a fixed ∆QP for the most complex ones. This method is exploited for instance in [17].

3.3.2 $R-\rho$ approaches

The ρ-domain, firstly introduced in [43], is supposed to avoid the complexity of the R−*Q* modeling. It links the number of non-zero coefficients in the transformed/quantized residual to the bitrate, in a linear way which can be easily computed and updated during the encoding process. Generally, the ρ -domain is modelled as described in the following Equation:

$$
R(\rho) = \theta \times (1 - \rho) \tag{3.48}
$$

where θ is the ρ -domain slope. Then, the QP is generally estimated based on a 1D lookup-table (LUT1D) which simply and efficiently maps the targeted ρ to its related QP. This approach has been exploited in [10, 44]. For HEVC, the ρ -domain was investigated in [4] where the *OP* computation is not based on a one-to-one mapping, but on the Cauchy approach discussed in the last Section:

$$
(1 - \rho) = c \times QP^{-\alpha}
$$
\n(3.49)

where *c* and *a* are model parameters. In [1] and [3], an alternative frame-level quadratic approach is explored based on Laplacian distribution model for residual coefficients. First, the ρ -domain is described according to the number of non-zero coefficients in the i^{th} picture:

$$
R_{Pic}(i) = \theta_i \times N_{NonZero}(i)
$$
\n(3.50)

where θ_i and $N_{NonZero}(i)$ are a model parameter and the number of non-zero coefficients in the *i th* picture, respectively. Then, the Laplacian distribution which models the residual coefficients for each depth j in the picture enables to express $N_{NonZero}(i)$ in the following way:

$$
N_{NonZero}(i,j) = N(i,j) \times e^{-\delta_i \mu_j q_i}
$$
\n(3.51)

where $N(i, j)$ is the whole number of pixels in depth *j*, μ_j is the model parameter of the jth depth, δ_i a model parameter and q_i the related quantization step for the frame i. This Equation can be transformed thanks to Taylor expansion in a quadratic form:

$$
N_{NonZero}(i) = N(i) \times (1 + a_i \times q_i^2 + b_i \times q_i)
$$
\n(3.52)

where $N(i)$ is the number of pixels into the ith picture (i.e its size), and a_i and b_i computed as follows :

$$
a_i = \sum_{j=0}^{N_{Depth}} (\delta_i \mu_j)^2 \quad , \quad b_i = -\sum_{j=0}^{N_{Depth}} \delta_i \mu_j \tag{3.53}
$$

where N_{Depth} is the maximum depth in the quadtree (i.e equal to 3 for HEVC). The ρ -domain approaches were mainly used to be integrated into low-complexity encoders because they only require a linear interpolation plus a 1D look-up-table for QP determination.

3.3.3 R-λ approaches

The R- λ based approach has been explored during HEVC standardization process [9, 45] and is composed of two steps. The Lagrangian multiplier λ used during RDO process is first fixed and the QP derivation is performed thereafter. This approach is based on the assessment that λ impacts all parts of the bitstream while a ρ -domain approach is supposed to be efficient for residual information only since it is based on QP tweaking. The initial approach is provided in [45] and is based on the accurate modeling between the bitrate and the Lagrangian multiplier:

$$
\lambda = \alpha \times (bpp)^{\beta} \tag{3.54}
$$

where α and β are the model parameters and *bpp* the bitrate in bit-per-pixel unit. The parameters are updated according to the following equations:

$$
\alpha_{new} = \alpha_{old} + \delta_{\alpha} \times \left(ln(\lambda_{real}) - ln(\lambda_{comp}) \right) \times \alpha_{old}
$$
 (3.55)

$$
\beta_{new} = \beta_{old} + \delta_{\beta} \times \left(ln(\lambda_{real}) - ln(\lambda_{comp}) \right) \times ln(bpp_{real})
$$
\n(3.56)

where δ_{α} and δ_{β} are two constants set to 0.1 and 0.005, respectively and λ_{real} and λ_{comp} are the achieved and reached λ values. Then, thanks to QP determination by λ value in HEVC [46], the QP can be directly computed:

$$
QP = 4.2005 \times ln(\lambda) + 13.7122 \tag{3.57}
$$

This approach is combined with λ and QP clipping so as to avoid abrupt quality and bitrate variations through the encoding. This $R-\lambda$ approach has been adopted in several contributions, for low-delay applications in [5] and for ROI in [18]. The QP determination has been improved for a particular dataset in [22]. In [25], the λ approach is joint to the CTU-level bit allocation where a recursive λ estimation is proposed to find the optimal bit allocation. Eventually, this method is enhanced in [24] where the GPP gradient introduced in Equation 3.23 is considered in the $R-\lambda$ model:

$$
\frac{bpp}{GPP} = \alpha_1 \times \lambda^{\beta_1} \tag{3.58}
$$

where α_1 and β_1 are the model parameters.

3.3.4 Scalable-based enhanced approaches

For scalable video coding, a way of improving encoding parameters estimation would be to consider inter-layer information for more accurate modeling. In the GT-based solution [31], the R-D function of a given spatial layer *l* is used as utility function is determined based on lower layers:

$$
D_l = (\mu_l \times Q_{l-1}) \times (R_l + c_l)^{-1}
$$
\n(3.59)

where µ*^l* and *c^l* are the model parameters for layer *l* and *Ql*−¹ the quantization step used in the reference lower layer. Other enhanced approaches have been proposed for R-Q model, for example in [47] where the predicted MAD value used in the R-Q function is derived from underlying layers, as follows:

$$
M\hat{A}D_{EL}(j) = a_1 \times MAD_{EL}(j-1) + a_2 + a_3 \times (MAD_{BL}(j) - M\hat{A}D_{BL}(j))
$$
 (3.60)

where (a_1,a_2,a_3) are model parameters. $MAD_{BL}(j)$ and $M\hat{A}D_{BL}(j)$ are respectively the actual and predicted value of BL MAD. $\hat{MAD}_{EL}(j)$ and $\hat{MAD}_{EL}(j-1)$ are the predicted and actual

values of the j^{th} and $(j-1)^{th}$ frames, respectively. In a similar way, the QP estimation of EL in [48] is based on BL computed QP, in a VBR context:

$$
QP_{EL} = QP_{BL} \times (1 + \gamma \times \log_{10} \frac{\Delta_{EL}}{R_{BL}})
$$
\n(3.61)

where *γ* is a model parameter and Δ _{*EL*} the bitrate relocation ordered by bandwidth fluctuation. For the R-λ approach, the HEVC-based algorithm has been transposed into the SHVC reference software [49] but without additional inter-layer tools. In [50], the λ parameter estimation in the EL is helped by BL information, as follows:

$$
\lambda = \frac{H_{EL} \times W_{EL}}{H_{BL} \times W_{BL}} \times 2^{(QP_{BL} - QP_{EL})/6}
$$
\n(3.62)

where (H_{BL}, H_{EL}) and (W_{BL}, W_{EL}) are the height and width of BL and EL pictures, respectively. *QPBL* and *QPEL* are the initialized QP in BL and EL.

3.4 Rate control in HEVC and SHVC reference softwares

Several RC algorithms have been successively considered during the SHM reference software development. The Unified Rate Quantization (URQ) method [15], introduced for HEVC in the HM7.1, has been firstly introduced in the first scalable test model SHM1.0. Then, more efficient R-λ approach [9, 45] has been considered in both HM9.1 and SHM1.2. In this software, the $R-\lambda$ mechanism is duplicated for each layer, associated with its specific targeted bitrate. This approach is composed of two mandatory bitrate allocation steps, at group of picture (GOP) and picture levels, and one optional step at Coding Tree Unit (CTU). First, the targeted number of bits for the upcoming GOP is computed according to the initial targeted bitrate and the number of already consumed bits (Equation 3.8). Then, a second picture-level bit allocation is performed for each picture in the GOP, according to its type and position within the GOP (Equation 3.15). In the same way, a block or CTU-level budget can also be allocated (Equation 3.19). Once the bitrate is allocated, the associated λ value is computed and applied to the upcoming RDO process (Equation 3.54). Eventually, the QP is computed (Equation 3.57) thanks to QP determination by λ value in HEVC [46].

3.5 Conclusion

According to the existing approaches, we can draw some prospectives for bitrate-control improvements in HEVC and SHVC. First we have seen in Section 3.3.4 that scalable

approaches based on ρ -domain are still unexplored for SHVC, and that there are promising results for HEVC [1, 3, 4]. In addition, existing work on SVC suggests that using ρ -domain for scalable video coding can significantly reduce complexity compared to classic multi-pass approaches [28]. Second, we can conclude from Section 3.2.4 that solutions for bitrate allocation between layers in a scalable scheme have not been explored yet. Indeed, strategies for bitrate allocation in separate layers are considered, but do not exploit correlation between layers. In addition, we can notice in Section 3.2.5 that efficient approaches for joint-encoding of several scalable programs exist [38–40], but never consider the joint encoding of several layers inside a program. These reasons clearly motivate and justify the need for better and more efficient rate-control strategies for scalable video coding, and especially for the latest SHVC standard. This way, this works aims at exploring the following tracks:

- rate-control strategies based on ρ -domain for SHVC,
- adaptive bitrate allocation between layers for SHVC,
- statistical multiplexing for SHVC coded programs.

Exploring these paths should lead to a more accurate and less complex rate control for SHVC, but also to improve the bitrate allocation between layers which should bring a better coding efficiency. These three main contributions of this thesis will be investigated in more details in Chapters 4, 5 and 6, respectively.

References

- [1] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Quadratic ρ-Domain Based Rate Control Algorithm for HEVC," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.
- [2] S. Wang, S. Ma, L. Zang, S. Wang, D. Zhao, and W. Gao, "Multi Layer Based Rate Control Algorithm for HEVC," *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013.
- [3] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP Based Rate Control for High Efficiency Video Coding," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, Dec. 2013.
- [4] Y. Y.-J., H. Kim, S.-H. Jung, D. Jun, Y. Kim, J. Choi, and S.-J. Ko, "A New Rate Control Method for Hierarchical Video Coding in HEVC," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2012.
- [5] Z. Yang, L. Song, Z. Luo, and X. Wang, "Low Delay Rate Control for HEVC," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2014.
- [6] H. Choi, J. Yoo, J. Nam, D. Sim, and I.-V. Bajić, "Pixel-Wise Unified Rate-Quantization Model for Multi-Level Rate Control," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 1112–1123, Dec. 2013.
- [7] L. Sun, O.-C. Au, C. Zhao, and F.-H. Huang, "Rate Distorsion Modeling and Adaptive Rate Control Scheme for High Efficiency Video Coding (HEVC)," *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2014.
- [8] M. Wang, K.-N. Ngan, and H. Li, "Efficient Frame-Content Based Intra Frame Rate Control for High Efficiency Video Coding," *IEEE Signal Processing Letters*, vol. 22, pp. 896–900, July 2015.
- [9] B. Li, H. Li, and L. Li, "λ-Domain Rate Control Algorithm for High Efficiency Video Coding," *IEEE Transactions on Image Processing*, vol. 23, pp. 3841–3854, Sept. 2014.
- [10] L. Xu, D. Zhao, L. Deng, S. Kwong, and W. Gao, "Window-Level Rate Control for Smooth Picture Quality and Smooth Buffer Occupancy," *IEEE Transactions on Image Processing*, vol. 20, pp. 723–734, Mar. 2011.
- [11] Y.-G. Lee and B.-C. Song, "An Intra-Frame Rate Control Algorithm for Ultralow Delay H.264/Advanced Video Coding (AVC)," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 19, pp. 747–752, May 2009.
- [12] J. Si, S. Ma, X. Zhang, and W. Gao, "Adaptive Rate Control for High Efficiency Video Coding," *Visual Communications and Image Processing (VCIP)*, 2012.
- [13] B. Lee, M. Kim, and T.-Q. NGuyen, "A Frame-Level Rate Control Scheme Based on Texture and Nontexture Rate Models for High Efficiency Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 24, pp. 465– 479, Mar. 2014.
- [14] D.-I. Park, H. Choi, J.-S. Kim, J.-S. Choi, and J.-G. Kim, "LCU-Level Rate Control for Hierarchical Prediction Structure of HEVC," 2013.
- [15] H. Choi, J. Nam, J. Yoo, D. Sim, and I. V. Bajic, "JCTVC-H0213: Rate Control based on Unified RQ Model for HEVC," Feb. 2012.
- [16] S. Sanz-Rodríguez and T. Schierl, "A Rate Control Algorithm for HEVC with Hierarchical GOP Structures," 2013.
- [17] J. Si, S. Ma, and W. Gao, "Efficient Bit Allocation and CTU Level Rate Control for High Efficiency Video Coding," 2013.
- [18] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Region-of-Interest-Based Rate Control Scheme for High-Efficiency Video Coding," *APSIPA Transactions on Signal and Information Processing*, vol. 3, no. 16, 2014.
- [19] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Region-of-Interest-Based Rate" Control Scheme for High-Efficiency Video Coding," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.
- [20] X. Liang, Q. Wang, Y. Zhou, B. Luo, and A. Men, "A Novel R-Q Model Based Rate Control Scheme in HEVC," *Visual Communications and Image Processing (VCIP)*, 2013.
- [21] M. Naccari and F. Pereira, "Quadratic Rate Control in the Emerging HEVC Standard," *Picture Coding Symposium (PCS)*, 2012.
- [22] H. Zhao, W. Xie, Y. Zhang, L. Yu, and A. Men, "An SSIM-Motivated LCU-Level Rate Control Algorithm for HEVC," *Picture Coding Symposium (PCS)*, 2013.
- [23] J. Wen, M. Fang, M. Tang, and K. Wu, "R-λ Model Based Improved Rate Control for HEVC with Pre-Encoding," *Data Compression Conference (DCC)*, 2015.
- [24] M. Wang, K.-N. Ngan, and H. Li, "An Efficient Frame-Content Based Intra Frame Rate Control for High Efficiency Video Coding," *IEEE Signal Processing Letters*, vol. 22, pp. 896–900, July 2015.
- [25] S. Li, M. Xu, and Z. Wang, "A Novel Method on Optimal Bit Allocation at LCU Level for Rate Control in HEVC," *IEEE International Conference on Multimedia and Expo (ICME)*, 2015.
- [26] I. Ahmad and J. Luo, "On Using Game Theory to Optimize the Rate Control in Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 16, pp. 209–219, Feb. 2008.
- [27] J.-F. Nash, "The Bargaining Problem," *Econometrica*, vol. 18, no. 2, pp. 155–162, 1950.
- [28] Y. Pitrey, M. Babel, and O. Deforges, "One-Pass Bitrate Control for MPEG-4 Scalable Video Coding Using ρ-Domain," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2009.
- [29] R. Atta, R.-F. Abdel-Kader, and A. Abd-AlRahem, "Single-Pass Dependent Bit Allocation for Spatial Scalability Coding of H.264/AVC," *European Signal Processing Conference (EUSIPCO)*, 2014.
- [30] L. Li, B. Li, and J. Zhang, "JCTVC-M0037: Rate Control by R-Lambda Model for SHVC," Apr. 2012.
- [31] X. Wang, S. Kwong, L. Xu, and Y. Zhang, "Generalized Nash Bargaining Solution to Rate Control Optimization for Spatial Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 23, pp. 4010–4021, Sept. 2014.
- [32] S. Sanz-Rodríguez and F. Díaz-de María, "In-Layer Multibuffer Framework for Rate-Controlled Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1199–1212, Sept. 2012.
- [33] Z. He and D.-O. Wu, "Linear Rate Control and Optimum Statistical Multiplexing for H.264 Video Broadcast," *IEEE Transactions on Multimedia*, vol. 10, pp. 1237–1249, Nov. 2006.
- [34] N. Changuel, B. Sayadi, and M. Kieffer, "Predictive Encoder and Buffer Control for Statistical Multiplexing of Multimedia Contents," *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 401–416, 2012.
- [35] Y. Wang, L.-P. Chau, and K.-H. Yap, "Joint Rate Allocation for Multiprogram Video Coding Using FGS," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 20, pp. 829–837, June 2010.
- [36] Y. Wang, L.-P. Chau, and K.-H. Yap, "Bit Allocation for Scalable Video Coding of Multiple Video Programs," *IEEE International Conference on Image Processing (ICIP)*, 2010.
- [37] M. Jacobs, S. Tondeur, T. Paridaens, J. Barbarien, R. Van de Walle, and P. Schelkens, "Statistical Multiplexing Using SVC," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2008.
- [38] W. Yao, L.-P. Chau, and S. Rahardja, "Joint Rate Allocation for Statistical Multiplexing in Video Broadcast Applications," *IEEE Transactions on Broadcasting*, vol. 58, pp. 417– 427, Sept. 2012.
- [39] W. Yao, L.-P. Chau, and S. Rahardja, "Joint Rate Allocation for Statistical Multiplexing of SVC," *IEEE International Conference on Image Processing (ICIP)*, 2012.
- [40] J. Jeong, Y.-H. Hung, and Y. Choe, "Statistical Multiplexing using Scalable Video Coding for Layered Multicast," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2009.
- [41] J. Si, S. Ma, S. Wang, and W. Gao, "Laplace Distribution Based CTU Level Rate Control for HEVC," *Visual Communications and Image Processing (VCIP)*, 2013.
- [42] N. Kamaci, Y. Altunbasak, and R.-M. Mersereau, "Frame Bit Allocation for the H.264/AVC Video Coder Via Cauchy-Density-Based Rate and Distorsion Models," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 15, pp. 994–1006, Aug. 2005.
- [43] Y.-K. Kim, Z. He, and S.-K. Mitra, "A Novel Linear Source Model and a Unified Rate Control Algorithm for H.263 / MPEG-2 / MPEG-4," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2001.
- [44] Z. He and S.-K. Mitra, "Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-Domain Source Modeling," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 12, pp. 840–849, Oct. 2002.
- [45] B. Li, H. Li, L. Li, and J. Zhang, "JCTVC-K0103: Rate Control by R-Lambda Model for HEVC," Oct. 2012.
- [46] L. Li, Z. Dong, H. Li, and J. Xu, "JCTVC-I0426: QP Determination by Lambda Value," Apr. 2012.
- [47] X. Lu and G. Martin, "An Improved Rate Control Algorithm for SVC with Optimised MAD Prediction," *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2014.
- [48] H. Lee, Y. Lee, D. Lee, J. Lee, and H. Shin, "Implementing Rate Allocation and Control for Real-Time H.264/SVC Encoding," *IEEE International Conference on Consumer Electronic (ICCE)*, 2010.
- [49] L. Li, B. Li, and H. Li, "JCTVC-M0037: Rate Control by R-Lambda Model for SHVC," Oct. 2012.
- [50] L. Li and H. Li, "λ-Domain based Optimal Bit Allocation for Scalable High Efficiency Video Coding," *IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015.

Part III

Contributions

Chapter 4

ρ-domain based rate control tools for HEVC and SHVC

4.1 Preamble

As mentioned in Section 3.3.2, the ρ-domain has been introduced in [1] for rate control algorithm and consists in linearly linking the bitrate with the rate of zero coefficients (noted ρ and referring to the number of zero coefficients over the number of all coefficients) in the transformed and quantized residuals, as described in the following Equation:

$$
R = \theta \times (1 - \rho) \tag{4.1}
$$

where R , θ and ρ are the bitrate, the model parameter and the ratio of zero coefficients in the transformed and quantized residuals, respectively. In [2], the authors throw doubts on ρ -domain approach, caused by the increase of no-residual syntax elements in HEVC with more efficient intra and inter predictions would lead to a lower amount of residual and then to a less accurate modeling. However, several contributions such as in [3], [4] or [5] show that the linear relation remains valid for HEVC. In this chapter, we provide a statistical analysis which investigates the ρ -domain and especially its validation for both HEVC and its scalable extension SHVC. We study the bitrate distribution in the bitstream (i.e. which amount is dedicated to residuals, MV, HLS, etc...) to be able to model the ρ -domain with both global and residual bitrate. Thereafter, we will explore alternative way for estimating the QP using ρ -domain for both HEVC and SHVC without using a LUT1D. Finally, we will explore new tool that enables to estimate the ρ -domain parameters in a deterministic way, so as to avoid pre-coding process. The algorithms developed in this section aims at being integrated in the existing ρ -domain based rate-control schemes, especially for HEVC and SHVC encoders.
4.2 Statistical study

4.2.1 Experimental procedure and dataset

For HEVC, the statistical study is performed on the dataset described in the Table 4.1, including the sequences considered in the CTC [6]. The encoding are carried out by the HM10.0 [7] reference software in random-access configuration which enables hierarchical GOP structure composed by four dependent temporal levels as illustrated in Figure 4.1a.

Fig. 4.1 Illustration of a CTC SHVC and HEVC GOP in RA configuration

For SHVC, the statistical study is performed on the dataset provided in Table 4.2. In this case, we consider the 2x and 1.5x spatial scalability which provides scalability from 960x540p and 1280x720p to 1920x1080p (HDTV), respectively. The encodings are carried out by the SHM2.0 reference software encoder in random access configuration with two layers. The same statistical study is performed on HEVC and in both SHVC layers. For SHVC, the statistics are only provided for the EL layer since the BL is encoded in the same way as in HEVC. In this Section, a representative subset of the whole results is provided, additional results can be found in Appendix A.1

Data Set	Name	Description	
Class A	PeopleOnStreet	2560x1600p308b	
	Traffic	SDR - BT.709	
	Kimono	1920x1080p248b	
	ParkScene	SDR - BT.709	
Class B	Cactus	1920x1080p508b	
	BasketballDrive	SDR - BT.709	
		1920x1080p608b	
	BOTerrace	SDR - BT.709	
		832x480p60 8b	
	BQMall	SDR - BT.709	
	BasketballDrillText		
Class _C	BasketballDrill	832x480p508b	
	PartyScene	SDR - BT.709	
	RaceHorses	832x480p30 8b	
	Keiba	SDR - BT.709	
	BasketballPass		
	BlowingBubbles	416x240p508b SDR - BT.709	
Class _D	BQSquare		
	Keiba	416x240p308b	
	RaceHorses	SDR - BT.709	
	FourPeople		
Class _E	Johnny	1280x720p60 8b SDR - BT.709	
	KristenAndSarah		

Table 4.1 Single-Layer dataset used for the ρ -domain statistical study

Table 4.2 Scalable dataset used for the ρ -domain statistical study

Sequence	$BL - 2x$	$BL - 1.5x$	EL
BasketballDrive	960x540p508b	1280x720p508b	
Cactus	SDR - BT.709	SDR - BT.709	SDR - BT.709
BOTerrace	960x540p60 8b	1280x720p60 8b	1920x1080p608b
	SDR - BT.709	SDR - BT.709	SDR - BT.709
Kimono1	960x540p248b	1280x720p248b	1920x1080p248b
	SDR - BT.709	SDR - BT.709	SDR - BT.709

As mentioned previously, the encoding are carried out in the RA configuration for both HEVC and SHVC. In this configuration, the sequence is composed by a number of GOP, each of these GOP consists of 8 pictures identified by a picture order count (POC), as illustrated in the Figure 4.1a. The encoding starts with an IRAP picture followed by the first GOP. Then, and after approximately 1s, the dependencies are broken and a GOP starting with an IRAP is introduced, and so on. This way, a new access point appears every one second which is mandatory in broadcast applications. For SHVC, the same scheme is used by duplicating two temporally aligned GOP where the EL GOP also uses the underlying frame as reference for prediction. For each RA point in the BL, dependencies are also broken at the EL where the temporally aligned frame is considered as P frame, which only uses intra and inter-layer predictions. In this configuration, the QP used to encode each level of the hierarchical GOP is estimated according to the asignement described in Table 4.3.

Table 4.3 QP computation in RA GOP for HEVC and SHVC

POC								
Encoding order								
Temporal layer								
$QP_{SL}(POC)$	$OP0+4$	$OP0+3$	$OP0+4$	$OP0+2$	$OP0+4$	$OP0+3$	$OP0+4$	OP0+1
$QP_{BL}(POC)$	QP_{SL}							
$QP_{EL}(POC)$	QP_{BL} -2							

In this Table, *QPSL* refers to the QP used in HEVC encoding. For SHVC, a fixed ∆QP is applied between BL and EL with lower quantization in the EL to promote the quality in the EL. This computation is based on a reference QP identified as QP0, used for IDR frames only. The pictures of the lower levels are favoured with lower QP value since they are used as reference, and thus are crucial for encoding the other ones with higher temporal levels.

4.2.2 Bitrate distribution in HEVC and SHVC RA bitstreams

To have a clear idea on how the residual information affects the bitstream, we collect statistics on the bitrate distribution in the HEVC RA bitstreams. For each temporal layer, we provide the bitrate distribution for sequences described in Tables 4.1 and 4.2. We split the bitstream according to the functions described in the tabular form syntax specification of HEVC. This split enables to classify the bitrate into several categories provided in Table 4.4. The *NALU Header* represents the bits used for signaling the information at Network Abstraction Layer Units (NALU) level. The *Slice Header* represents the proportion of bits consumed for slice segments signaling, which includes SE included in the *slice_segment_header()* section of the bitstream. The *SAO* represents the bits dedicated to signaling the SAO syntax. The *CU*

syntax represents the bits consumed for description of the quadtree structure, included in the *coding_quatree()* and *coding_unit()* sections of the bistream. *Prediction Syntax* and *Motion Vectors* respectively refer to the bits used for signaling in *prediction_unit()* and *mvd_coding()* sections of the bitstream. The *Transform Syntax* refers to the bits used for encoding information includes in the *transform_tree()* and *transform_unit()* parts of the bitstream. In the same way, the *Residual* categories refers to the information includes in the *residual_coding()* field of the bitstream. Eventually, the remaining and non considered sections of the bitstreams are grouped in the category *Others*. In addition this information on bitrate distribution, we will also provide the ρ -domain curves for each classes for both residual bitrate and global bitrate, to investigate ρ -domain validity.

Table 4.4 Description of categories consider in the statistical study and their references

In Figure 4.2 and 4.3, an overview of the collected statistics is provided for the *BasketballDrive* sequence coded with HEVC and SHVC EL, respectively at different *QP*0 values. Additional results are provided in Appendix A.1. We can first notice that as expected, the residual part of the bitstream remains the most important relative to the other categories. However, the amount of bits dedicated to residual and thus directly linked to the residual quickly falls under 50% for low bitrate cases and especially for the highest temporal levels of the GOP. We can also notice that this distribution remains similar in both HEVC and SHVC EL encodings. It appears that ρ -domain modeling should be efficient combined with an header bits estimation. As a first recommendation, ρ -domain is probably more suitable for high-bitrate application. In the next Section, the ρ -domain is modeled versus both residual only and global bitrate to have further information on how efficient is this approach at different bitrates.

Fig. 4.2 Average bitrate distribution for BasketballDrive RA bitstreams in HEVC

Fig. 4.3 Average bitrate distribution for *BasketballDrive* RA bitstreams in SHVC EL in CTC encoding

Fig. 4.4 Illustration of a ρ -domain modeling for BasketballDrive random-access bitstreams

4.2.3 ρ -domain analysis for RA encoding in HEVC and SHVC

In order to assess the ρ -domain linearity, we measure the number of residual zero coefficients divided by the whole number of coefficients (noted here ρ), for each sequence of the HEVC and SHVC selected dataset. In the Figure 4.4, we provide an example for *BasketballDrive* sequence where each point represents a frame in the sequence. Then, we draw the ρ -domain curve that fits these points while minimizing the least-square errors. To assess the accuracy the of this model with both the residual bitrate and the global bitrate (bitrate of all categories in the bitstream), we provide in blue the curve for global bitrate while the residual bitrate is shown in red. In addition to these curves, the coefficients of determination of the estimated linear models are also provided in Table 4.5, and computed as follows:

$$
E^2 = 1 - \frac{\sum_i (f_i - \tilde{f}_i)}{\sum_i (\tilde{f}_i - \overline{f})}
$$
\n(4.2)

where f_i is the actual samples, \overline{f} is the average value of these samples and \tilde{f}_i the samples linearly estimated. These results are representative of the dataset, and additional results are provided in Appendix A.1.3.

Table 4.5 Coefficients of determination E^2 for linear ρ -domain modeling applied on the BasketballDrive random-access bitstreams

Coding	Considered	QP ₀			
Scheme	B itrate	22	27	32	37
HEVC SL	Global	0.9921	0.9797	0.9683	0.9567
	Residual	0.9994	0.9990	0.9996	0.9995
SHVC EL	Global	0.9959	0.9810	0.9532	0.9189
	Residual	0.9989	0.9969	0.9972	0.9976

We can notice on the curves provided in Figure 4.4 that the linear behavior of the ρ -domain is efficient and accurate for estimation on residual bitrate, as expected. Moreover, the models based on global bitrates seem linear and suit well to the estimated curves. Regarding to the coefficients of determination, we can observe that the linearity of the modeling is good and always above 0.99 for residual part only. The modeling at global-bitrate level is also satisfying, especially for QP=22 and QP=27 (high bitrates).

4.2.4 Analysis and discussion

In the previous Sections, we have seen both the bitrate distribution in HEVC and SHVC EL bitstreams and also the ρ -domain modeling considering the global or residual only bitrates. Regarding to the bitrate distribution, we can notice that logically the residual part is more significant for lowest levels in the hierarchical GOP for both HEVC and SHVC EL. However, we can notice for higher QP which are related to lowest bitrates that the residual part fall under 50% of the whole bitrate which may suggests that ρ -domain modeling on the whole bitrate could be inefficient (less accurate).

However, the provided modeling shows interesting coefficients of determination. We can observe that considering residual part only always provide an high quality modeling with a coefficient of determination greater than 0.99. If we consider global bitrate for modeling we can notice that the linear trend is well maintained along the QP0 with a coefficient of determination from 0.9921 to 0.9567 for HEVC and from 0.9959 to 0.9189 for SHVC, which is probably appropriate for providing an accurate rate-control scheme. As a recommendation, using ρ -domain is probably more appropriate for high bitrate applications where the amount of residual is substantial and thus where a good coefficient of determination can be achieved.

To conclude this section, we can first state that ρ -domain models for HEVC and SHVC applications can be logically considered, even with the reticence about the amount of residual in these new standards. We have showed that this model is appropriate for both residual and global bitrates even if at low bitrate case seems less appropriate for global bitrate modeling. In addition, the state-of-the-art on rate control techniques shows that ρ -domain based schemes are often based on LUT1D which involve an "*a posteriori*" modeling since the tables are updated after selecting the encoding parameters. Thus, we can conclude that there is a work to do in developing deterministic schemes for ρ -domain based encoding parameter estimation. This will be investigated in the next section for QP estimation.

4.3 Quantization parameter estimation

The estimation of encoding parameters which enable to reach a targeted bitrate is crucial in rate-control algorithms. This way, the first step of our work consists in proposing an encoding parameter estimation scheme for both HEVC and SHVC. Our proposed method models the distribution of transformed and quantized coefficients and computes the QP which enables to reach the targeted ρ in each CTU. The parameters of our method are derived from the neighborhood CTUs to provide an accurate QP estimation in the current CTU.

4.3.1 Preliminaries

In [8], a particular mixture of Laplacian distribution for inter residual modeling is provided. It was demonstrated that transformed coefficients are distributed separately depending on the

HEVC CU depth. In Equation 4.3, the Laplacian probability density function is defined with the distribution parameter μ_{depth} of the coded block depth.

$$
f_{depth}(x) = \frac{\mu_{depth}}{2} \times e^{-\mu_{depth} \times |x|}
$$
\n(4.3)

The variance of the samples set at a specific depth can be computed as $\sigma_{depth}^2 = 2/\mu_{depth}^2$. In Figure 4.5, we can observe an example of distributions at different depths computed for HEVC residuals. We choose to use this distribution in our method since it provides an accurate modeling of the transformed and quantized coefficients with a simple μ_{depth} computation through variance.

Fig. 4.5 Depth-dependent Laplacian distribution model measured in HEVC

As mentioned in Section 3.3.2, ρ -domain for video coding was introduced in [1], linking ρ the rate of zero coefficients after transformation and quantization with the resulting bitrate R (see Equation 4.1). Experiments show that this ρ -domain is linear for a given picture, where θ denotes the slope parameter. This behavior was identified for HEVC in [3] and the previously introduced experiments on HEVC and SHVC have confirmed this assumption. To achieve a certain ρ in the quantized residual coefficients, we have to find the threshold value which sets ρ rate of coefficients to zero. This threshold value is called the dead-zone boundary (DZB). With the ρ -domain linear approach, we can easily compute the ρ we have to reach for a given bitrate target. By combining it with Laplacian distribution, we can determine the DZB producing this ρ , and eventually the related quantization parameter. To link ρ with this DZB value, we have to integrate the Laplacian distribution in the interval

[−*DZB*,+*DZB*]; thus we have:

$$
\rho_{depth} = \int_{-DZB}^{+DZB} f_{depth}(x)dx = 1 - e^{-\mu_{depth} \times |DZB|}
$$
\n(4.4)

These approaches, introduced in the existing literature, are used in our proposed method to efficiently perform the QP estimation at CTU-level. This way, our method is based on the following information collected while encoding. The parameter θ and μ_{depth} respectively refer to the ρ -domain slope and to the Laplacian distributions with the depth parameter which varies from 0 to 3 in HEVC. The number of depth coded per residual is noted *Ndepth* (with $N_{depth} \in \{0, 1, 2, 3\}$). Finally, N'_{size} is the number of coded residuals per size (32x32, 16x16, 8x8 and 4x4). All these parameters are collected per CTU while encoding. We eventually have one array per parameter. All these parameters will be used to predict encoding parameters of future CTU. The ρ -domain slope θ is computed with the resulting number of bits produced after CTU encoding and the measured ρ (Equation 4.5).

$$
\theta = \frac{R}{1 - \rho} \tag{4.5}
$$

The Laplacian parameter μ_{depth} should be computed with the variance estimator, but to limit the complexity we use a biased estimator. Indeed, we only measure the number of zero coefficients, that we divide by the whole number of coefficients, in order to get the central value (equal to $\mu/2$). We will observe in experimental results that this biased approach does not impact the accuracy of the proposed method. The two other parameters *Ndepth* and *N* ′ *size* are simply measured while encoding.

Fig. 4.6 Possible prediction configurations, a) no predictor, b) above CTU, c) left CTU.

4.3.2 Proposed method

Considering the previously introduced parameters, we can now describe the first axis developed in our method. When a new CU has to be processed, the first step of the proposed

solution is the estimation of indicators based on direct neighbourhood. These indicators are: the most probable residual size, the most probable ρ -domain slope, the most probable depth and the most probable μ . In Figure 4.6, we can observe several cases for CTU based prediction. Derivation of these indicators is only performed from neighbour's value. If left neighbour is available, it is used for reference (c), otherwise the above CTU is used (b). A predetermined initial values are used for initial CTU (a). If a CTU has no indicator due to a lack of residual, we assign the last existing measure (in the frame or in the previous frame). This way, we estimate for the current CTU: the residual size N, the ρ -domain slope θ and the depth. After computing these indicators and retrieving targeted bitrate R , the first step consists in computing the targeted ρ :

$$
\rho = 1 - \frac{R}{\theta_{est}} \tag{4.6}
$$

Once $ρ$ is available, it can be combined with the most probable $μ$ to obtain the dead-zone boundary, derived from Equation 4.4:

$$
DZB = -\frac{1}{\mu} \times \ln(1 - \rho) \tag{4.7}
$$

The final step consists in computing the QP which enables to reach this DZB on residuals. The HEVC standard specification gives a particular Equation for the scaling process, written with logical operations. This Equation can be re-written in the following way:

$$
D(x,y) = \frac{T(x,y) \times S(x,y) \times L(QP \pmod{6}) \times 2^{\frac{QP}{6}}}{2^{bdShift}} + \frac{1}{2}
$$
(4.8)

with (x, y) the indices of the coefficient position in the residual block, $T(x, y)$ the scaled coefficient, $D(x, y)$ the transformed coefficient, $S(x, y)$ a scaling factor (equal to 16 if no scaling list is used), *bdShift* described in Equation 4.9 and a scaling factor $L = \{40, 45, 51, 57, 64, 72\}$ for $k=0$ to 5.

$$
bdShift = BitDepth + \log_2(N) - 5 \tag{4.9}
$$

When $D(x, y) = DZB$, the *DZB* is reached and thus $T(x, y) = 1$. Assuming that we use a 8-bit depth, without using a scaling list, we have:

$$
F(QP) = L(QP \pmod{6}) \times 2^{\frac{QP}{6}} = \left(DZB - \frac{1}{2}\right) \times 2^{\log_2(N) - 1}
$$
 (4.10)

The function F has to be inverted to get the appropriate QP. It must be noted that the $\frac{QP}{6}$ is an integer division. Combining this integer division with *L*[*QP*%6] enables to derive the following equation for QP estimation:

$$
QP = 6 \times \log_2\left(\frac{1}{40} \times \left(DZB - \frac{1}{2}\right) \times 2^{\log_2(N) - 1}\right) \tag{4.11}
$$

We also hold the QP value in the interval described in Equation 4.12, with $\Delta_{OP} = 5$.

$$
QP_{frame} - \Delta_{QP} \le QP \le QP_{frame} + \Delta_{QP} \tag{4.12}
$$

In this approach, the Lagrangian multiplier λ to be used for the RDO process is determined in the same classical way as the reference software [9], based on the selected QP. This λ computation is defined as follows:

$$
\lambda_{Mode} = \alpha \times W_k \times 2^{(QP-12)/3.0}
$$
\n(4.13a)

$$
\lambda_{Pred} = \sqrt{\lambda_{Mode}} \tag{4.13b}
$$

where α , W_k are parameters derived based on the position of a picture in the GOP or its reference-status. To summarized the method, encoding parameters are firstly derived from neighbourhood. Then, the related targeted ρ is computed as well as the dead-zone boundary, based on Equation 4.6 and 4.7. Lastly, the quantization parameter is computed with Equation 4.11 by fulfilling the interval described in Equation 4.12. In order to have feedback for the next CTU encoding, we collect the parameters described in previous section, after each CTU encoding.

4.3.3 Experimental procedure and dataset

To check the algorithm, we have implemented it in the HM 13.0 Reference Software. In Figure 4.7a, we can observe the experimental procedure we considered. This test is performed in RA configuration. A fixed QP encoding is first performed in the classical encoder, the bit budget generated for each CTU is then used as targeted bitrate for our algorithm. The method described in the previous Section is used to produce the appropriate QP. In Table 4.6, the accuracy of the proposed method is provided as well as the PSNR difference between our method and the fixed QP anchor. A positive PSNR difference means that our method provides a better objective quality while negative value involves a deterioration.

In addition, we have duplicated this algorithm for each layer in a scalable scheme and implemented it in the SHM 5.1 encoder. In a similar way as HEVC experimental procedure, the targeted bit budget generated for each CTU in both BL and EL are transmitted to the

Fig. 4.7 Experimentation used to evaluate the QP estimation scheme for HEVC and SHVC

encoder where our method is enabled. The performance are recorded in Tables 4.7 for BL and EL encoding. In the following Section, the results are discussed and analysed.

4.3.4 Analysis and discussion

Regarding to HEVC performance, provided in Table 4.6, we can notice that the average 103% accuracy is satisfying considering the approximations and choices made. There is a slight bitrate mismatch since there is no targeted bitrate adjustment while encoding. This is not an issue since the aim of the proposed algorithm is only to estimate the QP. Regarding objective visual quality metrics, we can notice that the proposed algorithm does not substantially change the PSNR with a -0.01 dB average difference.

For SHVC results, provided in Table 4.7, we can observe that both layers are showing a satisfying bitrate error of 100% and 104% for BL and EL respectively. The slight bitrate overhead can be explained by the same reasons as HEVC scheme, namely that no corrections are performed while encoding. The proposed algorithm is only focused on the QP estimation side and then could be attached to a bitrate management scheme. With regard to visual quality, there is not substantial variation in objective quality measurement with only +0.06 dB and +0.08 dB of average differences. We can notice that PSNR variations are in harmony with bitrate ones since a bitrate increase leads to PSNR improvement. In the same way, when

a bitrate drop occurs, the associated PSNR difference tends to be negative.

To summarize, we propose a single-pass approach using ρ -domain for HEVC and SHVC. The proposed algorithm is really straightforward and enables a very good accuracy with no substantial with no PSNR deterioration. Moreover, this method can be associated to any existing bitrate-allocation algorithm to achieve efficient HEVC/SHVC rate control. The SHVC method can be potentially enhanced by considering inter-tools, which is a track for future work.

	Bitrate [kbps] QP-Ref Sequence					
		Reference	Reached	Error	\triangle PSNR [dB]	
Basketball	22	15277.16	15670.46	$+3\%$	$+0.03$	
	27	5223.90	5343.10	$+2\%$	$+0.02$	
	32	2401.90	2401.90	0%	0.00	
	37	1242.59	1284.26	$+3\%$	-0.02	
	22	36027.81	34669.14	$-4%$	$+0.05$	
BQTerrace	27	5571.87	5835.83	$+5\%$	-0.02	
	32	1410.73	1427.82	$+1\%$	0.00	
	37	501.49	518.48	$+3\%$	-0.01	
Kimono1	22	4034.17	4095.23	$+2\%$	-0.01	
	27	1776.67	1854.69	$+4\%$	-0.04	
	32	828.26	872.63	$+5\%$	-0.06	
	37	402.14	423.07	$+5\%$	-0.05	
ParkScene	22	5640.31	5836.40	$+3\%$	-0.05	
	27	2172.29	2267.63	$+4\%$	-0.05	
	32	904.84	952.03	$+5\%$	-0.07	
	37	392.83	415.03	$+5\%$	-0.03	
Tennis	22	4212.69	4241.41	$+1\%$	$+0.03$	
	27	1888.17	1947.98	$+3\%$	-0.01	
	32	911.79	945.08	$+4\%$	-0.02	
	37	485.90	498.63	$+3\%$	-0.01	
Mean				$+3\%$	-0.01	

Table 4.6 Achieved bitrate precision for SL HEVC encoding

				Bitrate [kbps]		
Sequence	QP-Ref	Layer	Reference	Reached	Error	\triangle PSNR [dB]
		BL	6418.28	6597.45	$+3%$	0.00
BasketballDrive	22	EL	25593.34	25634.92	0%	-0.03
	27	BL	3064.25	3177.42	$+4%$	$+0.05$
		EL	6287.34	6228.71	$-1%$	-0.06
	32	BL	1506.12	1565.13	$+4%$	$+0.09$
		EL	2630.39	2601.94	$-1%$	-0.05
	37	BL	789.23	816.78	$+3%$	$+0.08$
		EL	1264.73	1277.06	$+1%$	$+0.01$
	22	BL	5708.96	5843.85	$+2%$	$+0.03$
		EL	75885.53	74103.70	$-2%$	-0.04
	27	BL	1757.15	1796.03	$+2%$	0.00
BQTerrace		EL	11166.24	11362.25	$+2%$	$+0.01$
	32	BL	683.21	710.33	$+4%$	$+0.06$
		EL	2196.69	2161.51	$-2%$	$+0.01$
	37	BL	294.00	308.13	$+5%$	$+0.07$
		EL	739.32	737.77	0%	$+0.01$
	22	BL	2492.75	2603.37	$+4%$	$+0.11$
		EL	4600.37	4303.09	$-6%$	-0.07
	27	BL	1153.82	1223.57	$+6%$	$+0.12$
Kimono		EL	1858.74	1833.57	$-1%$	0.00
	32	BL	536.59	568.69	$+6%$	$+0.14$
		EL	881.74	887.13	$+1\%$	$+0.02$
	37	BL	254.89	270.33	$+6%$	$+0.11$
		EL	425.96	431.63	$+1%$	$+0.06$
	22	BL	5766.57	5942.96	$+3%$	$+0.02$
		EL	28420.49	28332.06	0%	-0.02
	27	BL	2634.60	2738.44	$+4%$	$+0.04$
Cactus		EL	5669.56	5629.15	-1%	-0.01
	32	BL	1264.68	1333.81	$+5%$	$+0.10$
		EL	2304.01	2354.57	$+2%$	$+0.06$
	37	BL	622.76	654.30	$+5%$	$+0.06$
		EL	1112.51	1132.51	$+2%$	$+0.04$
	22	BL	3006.71	3113.68	$+4%$	$+0.07$
		EL	8107.10	8312.39	$+3%$	$+0.18$
	27	BL	1298.74	1352.70	$+4%$	$+0.10$
ParkScene		$\mathop{\rm EL}$	2958.76	3051.15	$+3%$	$+0.10$
	32	BL	562.22	590.61	$+5%$	$+0.11$
		EL	1228.14	1255.90	$+2%$	$+0.06$
	37	BL	241.21	253.37	$+5%$	$+0.05$
		EL	525.39	542.77	$+3%$	$+0.08$
Per-Layer		BL			$+4%$	$+0.06$ dB
Mean		EL			0%	$+0.08$ dB
Global Mean					$+2\%$	$+0.07$ dB

Table 4.7 Achieved bitrate precision for SHVC encoding

4.4 Deterministic ρ -domain parameter estimation

To efficiently model the ρ -domain slope, and as mentioned in Section 3.3.2, the common methods update the model during the encoding process based on the *'a-posteriori'* estimation. This approach supposes that the model of the upcoming frame is estimated based on the models of the already encoded frames. In this Section, we provide a new deterministic model for ρ -domain slope estimation based on the spatial resolution of the video signal. This model can also be enhanced in terms of accuracy according to the rate of skipped-block in the pictures. First, the relation between the frame-level ρ -domain models and the video resolution is explored to assess our approach. Then, a model is proposed based on these observations. The accuracy of the model is evaluated on the dataset described in Table 4.1. Finally, an architecture is proposed for integration of this tool in existing ρ -domain based rate-control schemes.

4.4.1 On the relationship between slope and resolution

In Section 4.2, the statistical study on ρ -domain is provided and enable to build a huge dataset for each frame of each sequence for several QP. We have evaluated, for each frame in the dataset, the ρ -domain model that minimizes the sum of least-squares errors on each of these frame. In Figure 4.8, the estimated ρ -domain slopes are provided where the bitrate is represented in bits per frame (i.e. bpf) and where the frames belonging to the same Class (i.e. same resolution) are represented in the same color and for all QP.

Fig. 4.8 Frame-based ρ -domain for the whole dataset

We can notice that the ρ -domain slope decreases with the spatial resolution since less bitrate is required for lower spatial resolution. However, each class seems to be located around a fixed location. These results are interesting since it appears that the ρ -domain slope θ could be potentially estimated in a deterministic way. In order to evaluate this aspect, we have displayed the average measured ρ -domain slope for each sequence versus the spatial resolution of the sequence in number of samples. It appears that the obtained curve seems perfectly linear, and thus confirms our first assumptions.

Fig. 4.9 ρ -domain slope versus the spatial resolution

In the Figure 4.9, we can observe the measured slope with the estimated linear model on this dataset. For this model, the coefficient of determination is equal to 0.9980 which enables to state that the linearity of the slope value as a function of the resolution is significantly marked. This potential for deterministic ρ -domain slope estimation would enable to perform a more accurate initialization in rate-control modules and thus improves its accuracy.

4.4.2 Proposed models

In this context, we propose a deterministic ρ-domain estimation which avoids *'a-posteriori'* modeling. So as to formulate this model, we start from the lowest possible level, i.e. the transformed and quantized residual and thus extends the formulation to the frame-level to match with our initial assumption (i.e. frame-level ρ -domain slope as a function of the resolution is linear). In HEVC, the residual information is recorded into square TB of size $N \times N$. In Figure 4.10, we provide the *ρ*-domain data cloud for each TB size in HEVC

 $(4 \times 4, 8 \times 8, 16 \times 16$ and $32 \times 32)$, and from sequences of Classes A and B. In addition, we provide the linear models computed based on the function $R(\rho) = 4 \times S \times (1-\rho)$ where *S* is the residual surface (i.e. N^2). The samples represent all the residual produced in classes A and B, without dissociating QP and frames. We can observe in Figure 4.10a and 4.10b that the ρ -domain relationship is logically valid and seems deterministic and unchanged, whatever the considered Class of sequence. In addition this formulation provides a very good coefficient of determination greater than 0.99. Thus, the ρ -domain can be estimated as described in the following Equation, for a given TB *i*:

$$
R_i = 4 \times N_i^2 \times (1 - \rho) + \Delta_i \tag{4.14}
$$

where Δ_i and N_i respectively refers to a TB level tweaking parameter and the TB block size. Based on this residual-level equation, we can derive a frame-level equation. First, let ℜ*^C* be the set of TB which composes the channel *C* (where $C \in \{Y, U, V\}$) of a given frame. Thus, we can write the frame level ρ -domain equation of the channel *C* as the sum of TB that it contains:

$$
R_C = \sum_{i \in \Re_C} R_i \tag{4.15}
$$

This equation can be combined with Equation 4.14 to obtain the following formulation:

$$
R_C = 4 \times \sum_{i \in \mathfrak{R}_C} (1 - \rho_i) \times N_i^2 + \sum_{i \in \mathfrak{R}_C} \Delta_i
$$
 (4.16)

Since ρ_i represents the rate of zero coefficient in the i^{th} TB, it can be replaced by its number of non-zero coefficient Z_i divided by its whole number of coefficients N_i^2 , and thus turns the

Fig. 4.10 ρ -domain modeling of transformed/quantized residual as a function of the size

first term in the following way:

$$
\sum_{i \in \mathfrak{R}_C} (1 - \rho_i) \times N_i^2 = \sum_{i \in \mathfrak{R}_C} \frac{(N_i^2 - Z_i)}{N_i^2} \times N_i^2 = \sum_{i \in \mathfrak{R}_C} N_i^2 - \sum_{i \in \mathfrak{R}_C} Z_i
$$
(4.17)

With this new formulation, we can express the Equation 4.16 as follows:

$$
R_C = 4 \times (S_C - Z_C) + \Delta_{FC}
$$
\n
$$
(4.18)
$$

with:

$$
S_C = \sum_{i \in \mathfrak{R}_C} N_i^2 \qquad Z_C = \sum_{i \in \mathfrak{R}_C} Z_i \qquad \Delta_{FC} = \sum_{i \in \mathfrak{R}_C} \Delta_i \tag{4.19}
$$

where S_C , Z_C and Δ_{FC} respectively refers to the whole TB surface, the number of zero coefficients and the global tweaking parameter of the given channel. To exploit this formulation at frame level, we just need to sum all the channels in the frame:

$$
R_F = \sum_{C \in \{Y, U, V\}} R_C = R_Y + R_U + R_V \tag{4.20}
$$

which is equivalent to:

$$
R_F = \sum_{C \in \{Y, U, V\}} \left(4 \times (S_C - Z_C) + \Delta_{FC} \right) \tag{4.21}
$$

or

$$
R_F = 4 \times (S_F - Z_F) + \Delta_F \tag{4.22}
$$

with

$$
S_F = \sum_{C \in \{Y, U, V\}} S_C \qquad Z_F = \sum_{C \in \{Y, U, V\}} Z_C \qquad \Delta_F = \sum_{C \in \{Y, U, V\}} \Delta_{FC} \tag{4.23}
$$

where S_F and Z_F are the TB surface and the number of zero coefficients in the whole frame, and where Δ_F refers to a frame-level tweaking parameter. Finally, we can express the frame-level ρ -domain Equation as:

$$
R_F = 4 \times S_F \times (1 - \rho_F) + \Delta_F \tag{4.24}
$$

This final relationship would enable to model the ρ -domain accurately at frame level, whatever the considered content. If we draw the ρ -domain curves according to the Equation 4.24, in a first time without considering the deviation parameter, we get the results provided in Figure 4.11. We can notice on this figure that the proposed ρ -domain equation fits well

to the dataset. However we can observe, that there could be an overhead between the predicted ρ -domain curve and the measured points, especially on the highest ratios. Indeed, we consider here that the frame surface S_F contains all the pixels for each channel, in other words that all pixels are represented in the bitstream through the residual syntax. This is not always true because several textures are skipped and thus not coded which leads to reconsider the formulation. This way, we propose to weight the whole frame surface by a coefficient *rSkip* which represents the rate of skipped surface in a given frame, this surface is defined as follows:

$$
S_F' = (1 - r_{skip}) \times S_F \tag{4.25}
$$

In this configuration, our ρ -domain model is modified to consider r_{Skip} and then to modulate the slope value:

$$
R_F = 4 \times S'_F \times (1 - \rho_F) + \Delta_F \tag{4.26}
$$

or

$$
R_F = 4 \times (1 - r_{skip}) \times S_F \times (1 - \rho_F) + \Delta_F \tag{4.27}
$$

This more accurate representation should enable a better ρ -domain modeling and thus a more efficient rate control. Indeed, the slope is modulate according to the *rSkip* value and thus should be able to be graphically moved toward the appropriate points in Figure 4.11. In the following Section, this model is evaluated on our dataset to evaluate its accuracy.

Fig. 4.11 Per-frame collected ρ -domain versus deterministic modeling.

4.4.3 Experiments and analysis

To evaluate the accuracy of the proposed model, we consider as reference the actual ρ -domain curve. For each sequence in our dataset, we sort and group the data by *rSkip* intervals of 1% and we estimate the ρ -domain model which minimizes the sum of least-squares on each of these interval. Then, we estimate the ρ -domain slope using our proposed models. In Figure 4.12b, we provide the ρ -domain slope estimation for r_{skip} -based method together with the ρ -domain slopes measured for each interval.

We can observe that the estimated slope based on r_{Skip} indicator is close to the real dataset, however some mismatches are still present. Indeed, the *rSkip* indicator does not consider the residual level skip mechanism signalled by the *cbf_luma*, *cbf_cb* and *cbf_cr* flags. Compared to the no-Skip approach, the final model is more accurate and more appropriate for low-bitrate applications where a lots of residual are skipped. To evaluate the proposed model, we provide in Table 4.8 the accuracy of this model compared to the measures realised on the dataset. The "no-skip" values are related to slope estimation with Equation 4.24 while "*rskip*-based" is related to Equation 4.27.

Class						Average
no-skip	586.42%	1151.13% 475.28%		1557.25% 924.56%		738.92%
r_{Skip} -based	92.50%	99.55%	91.60%	90.50%	90.10%	92.85%

Table 4.8 Average accuracy of the proposed no-skip and *rSkip*-based model

First, we can notice that the no-skip method logically provides bad performance since the ρ -domain slope is accurate only for high-bitrates where a lot of residual information is coded in the bitstream. In Figure 4.12a, this accuracy in high-bitrate areas is illustrated in the circled areas which are related to low *rskip*. Graphically, the *rskip*-based method should enable bring the slope close to the real values.

Second, we can observe that the *rskip*-based method is logically better, with an accuracy is maintained above 90% for each Class in the dataset. The mismatch can be for corrected for instance with the tweaking parameter Δ_F , or by including the residual-level skip mechanism (*cbf_luma*, *cbf_cb* and *cbf_cr*) in *rSkip* computation. However, a first initialization can be performed without considering the *rSkip*, by using model described in Equation 4.24. For more accurate modeling, the *rSkip* value can be considered, and initialized according to content complexity for instance.

(b) Deterministic slope estimation based on *rSkip* rate

Fig. 4.12 Illustration of ρ -domain curves and their deterministic estimation

4.4.4 Architecture suggestion for deterministic slope estimation

As a track for future work in this field, we provide in this Section some suggestions of encoder architecture using the previously described technologies. For HEVC, an example of architecture is provided in Figure 4.13.

Fig. 4.13 Architecture suggestions for *rSkip*-based rate control

In this suggested architecture the encoding process is controlled by a ρ -domain based RC unit. During the whole encoding process (i.e after each frame), the skip-rate is measured and stored into a database which collects statistics on *rSkip* values for each temporal layer for instance. This database is then used for estimating the skip-rate value for the current frame. This skip-rate estimation is transmitted to the rate-control unit which uses this value to update the ρ -domain rate model. In a similar way, this method could be implemented in scalable encoders where statistics coming from the BL can be used in the EL.

4.5 Conclusion

In this Chapter, algorithms addressed to ρ -domain based rate-control schemes have been developed. A deep and exhaustive statistical study has been carried out to show and attest that ρ -domain is still valid for HEVC and SHVC (at diffrent bitrates) and then can be efficiently considered. These algorithms enables to avoid multi-pass encoding or pre-processing stage by enabling on-the-fly efficient rate-control.

First, A CTU-Level QP estimation scheme has been proposed for HEVC and SHVC which enables a good accuracy for various bitrates with no quality degradation. Second, a deterministic way of estimating the ρ -domain slope have been proposed. This method proves that $ρ$ -domain can be efficiently modelled based on a formulation which considers the frame resolution modulated by the skip-rate in the frame, which enables an accurate ρ -domain slope initialization and control during the encoding. These technologies aim at being attached to a bitrate-allocation scheme to form a complete RC scheme, as illustrated in architecture suggestion.

However, several tracks remain for improving and integrating these tools to existing encoders. Indeed, the QP estimation scheme could be further studied in the SHVC case by exploring inter-layer aspects. These aspects would help the EL rate-control unit to be more accurate, based on information coming from BL. In addition, the deterministic approach developed in the second part could be integrated in the QP estimation scheme to improve its accuracy. Nevertheless, this work proposes some innovative and efficient tools which can be integrated in existing encoders even if some prospectives for improvement remain. In the following Section, the adaptive rate control schemes will be investigated for SHVC and especially applied to the deployment of UHDTV services.

References

- [1] Y.-K. Kim, Z. He, and S.-K. Mitra, "A Novel Linear Source Model and a Unified Rate Control Algorithm for H.263 / MPEG-2 / MPEG-4," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2001.
- [2] B. Li, H. Li, and L. Li, "λ-Domain Rate Control Algorithm for High Efficiency Video Coding," *IEEE Transactions on Image Processing*, vol. 23, pp. 3841–3854, Sept. 2014.
- [3] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Quadratic ρ-Domain Based Rate Control Algorithm for HEVC," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.
- [4] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP Based Rate Control for High Efficiency Video Coding," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, Dec. 2013.
- [5] Y. Y.-J., H. Kim, S.-H. Jung, D. Jun, Y. Kim, J. Choi, and S.-J. Ko, "A New Rate Control Method for Hierarchical Video Coding in HEVC," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2012.
- [6] F. Bossen, "JCTVC-L1100: Common Test Conditions and Software Reference Configurations," Jan. 2013.
- [7] F. Bossen, D. Flynn, and K. Suehring, "JCTVC-L1010: HEVC HM 10 Reference Software," Feb. 2013.
- [8] B. Lee and M. Kim, "Modeling Rates and Distorsions Based on a Mixture of Laplacian Distributions for Inter-Predicted Residues in Quadtree Coding of HEVC," *IEEE Signal Processing Letters*, vol. 18, pp. 571–574, Oct. 2011.
- [9] C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G. Sullivan, "JCTVC-T1002: High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Improved Encoder Description Update 2," Feb. 2015.

Chapter 5

Adaptive rate control for SHVC

5.1 Preamble

As mentioned in the state-of-the-art Section 3.5, there is a need for enhanced and efficient ARC schemes in SHVC. To develop such algorithms, we first have to explore the impact of bitrate ratio between layers over SHVC performance in 2-layers configuration. The results and observations from this experimental study will enable to model the SHVC encoder performance according to the bitrate ratio and then design an adaptive rate control algorithm which dynamically adjusts the bitrate ratio to optimize the encoding process, for a given global constant bitrate R_G . In this Section, we firstly provide the statistical study about the ratio impact on the SHVC performance in different coding configurations and video contents. Then, we develop an algorithm which dynamically allocates bitrate within layers by finding the optimal balance between these layers. First, a simple method is tested without considering any bitrate and quality instructions on the layers. Second, an enhanced method where bitrate and quality constraints are considered is proposed. Finally, we provide experiments on both approaches with results and analysis.

5.2 Notations and definitions

Before introducing the statistical study and the proposed algorithm, we first give some definitions and notations that will be used in the rest of this chapter. Let us consider the global targeted bitrate R_G , defined in Equation 5.1, as the sum of BL and EL bitrates.

$$
R_G = R_{BL} + R_{EL} \tag{5.1}
$$

where R_{BL} and R_{EL} are the BL and EL bitrates, respectively. The bitrate ratio (noted τ) is defined in Equation 5.2 as the BL bitrate divided by the global bitrate *RG*.

$$
\tau = \frac{R_{BL}}{R_G} \tag{5.2}
$$

In addition, we consider several instructions for 2-layers encoding such as quality and bitrate. These parameters are given in Table 5.1 with their respective units.

Notation	Definition	Unit
τ	Bitrate ratio as defined in Equation 5.2	П
R_G	Global targeted bitrate as defined in Equation 5.1	
R_{BLmin}	Minimal desirable bitrate in the BL	
R_{ELmin}	Minimal desirable bitrate in the EL	$[bit][s^{-1}]$
R_{BLmax}	Maximal authorized bitrate in the BL	
R_{ELmax}	Maximal authorized bitrate in the EL	
Q_{BLmin}	Minimal desirable PSNR in the BL	
Q_{ELmin}	Minimal desirable PSNR in the EL	
Q_{BLmax}	Maximal authorized PSNR in the BL	[dB]
O_{ELmax}	Maximal authorized PSNR in the EL	

Table 5.1 Principal notations for proposed ARC schemes

We also introduce the dual-Picture which is the aggregation of two temporally aligned BL and EL pictures. In a similar way, the dual-GOP is defined as the aggregation of two temporally aligned BL and EL GOP. The concepts of dual-Picture and dual-GOP are illustrated in Figure 5.1 and 5.2, respectively. It must be noted that this concept exists in HEVC under the name of Action Unit (AU). Eventually, the content is also characterized by its framerate *f^r* and its length in frames *N^f* .

Fig. 5.1 Illustration of dual-Picture

Fig. 5.2 Illustration of dual-GOP

5.3 Statistical study

5.3.1 Dataset, bitrates and ratios

In order to study the impact of bitrate ratio over performance through BD-BR and BD-PSNR scores [1], we will measure the encoding performance for nine tested ratio, from 0.1 to 0.9 by step of 0.1 ($\tau \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$). Each of these ratio has to be applied on four global targeted bitrates $R_G \in [R_0, R_1, R_2, R_3]$. For a given ratio τ and a global targeted bitrate R_T , the BL and EL targeted bitrates are computed as follows:

$$
R_{BL} = \tau \times R_G \tag{5.3a}
$$

$$
R_{EL} = (1 - \tau) \times R_G \tag{5.3b}
$$

This study is performed on 1.5x and 2x spatial scalability, CGS and SDR-to-HDR scalability. Thus, we consider the dataset provided in Table A.1 of Appendix A.2. The sequences specified in the SHM Common Test Conditions (CTC) [2] are used to evaluate the impact of bitrate ratio for 1.5x and 2x spatial scalability (i.e. Class A and B). In addition, the UHD-1 dataset [3] provided by the European Broadcasting Union (EBU) is also investigated in 2x spatial scalability configuration. The sequences proposed by Technicolor [4] combining CGS and 2x spatial scalability are also tested. Finally, SDR to HDR scalability using CGS is also tested based on the sequences considered during the CfE on WCG and HDR video coding [5]. In addition, five sequences used for evaluating SDR-to-HDR scalability are provided. They were shot during the French stage of the Volvo Ocean Race 2015 which took place in *Lorient*, under a collaborative project of French-based audiovisual companies. In order to cover a wide range of configurations, we have selected the bitrates recorded in Table 5.2. The previously specified ratios (from 0.1 to 0.9) are applied to these global bitrates to build per-layer targeted bitrates according to Equation 5.20a and 5.20b.

Sequences/classes	Scalability	R_0	R_1	R_2	R_3
Class A	Spatial 2x	5 Mbps	10 Mbps	15 Mbps	20 Mbps
Class B	Spatial 1.5x	0.5 Mbps	2 Mbps	3.5 Mbps	5 Mbps
Class B	Spatial 2x	0.5 Mbps	2 Mbps	3.5 Mbps	5 Mbps
EBU	Spatial 2x	5 Mbps	10 Mbps	15 Mbps	20 Mbps
Technicolor	CGS / Spatial 2x	5 Mbps	10 Mbps	15 Mbps	20 Mbps
HDR CfE	CGS / SDR-to-HDR	0.5 Mbps	2 Mbps	3.5 Mbps	5 Mbps
Volvo Ocean Rave	CGS / SDR-to-HDR	5 Mbps	10 Mbps	15 Mbps	20 Mbps

Table 5.2 Selected bitrates for the statistical study

5.3.2 Method to build comparative tests

Scalable coding technology can be evaluated through several ways with different comparisons with single layer HEVC encoding. Three SHVC performance evaluation scenarios are provided in Table 5.3. In scenario n° 1, the EL produced by SHVC is compared to HEVC equivalent single-layer encoding of the EL. In this scenario the scalable scheme is expected to be better thanks to IL predictions. In scenario $n^{\circ}2$, the whole scalable stream is compared to the equivalent Simulcast HEVC encoding; scalability is also expected to be better for the same reasons as in scenario n° 1. In scenario n° 3, the whole SHVC stream is compared to HEVC encoding of its highest layer. In this case, the SHVC stream is expected to be less efficient, since it transports two services (BL & EL).

Table 5.3 RD evaluation of SHVC coding scheme with respect to HEVC

Scenario	SHVC	HEVC	Expected performance
			SHVC vs HEVC
	EL	EL (HEVC encoding of the EL)	Gain
	BL+EL	$BL + EL$ (Simulcast HEVC encoding)	Gain
	$BI + EI$	EL (HEVC encoding of the EL)	Loss (Overhead)

In this study, we provide the R-D performance scores (i.e. BD-BR and BD-PSNR) for the scenario $n^{\circ}3$ where scalability brings overhead. Then, we display the performance as a function of the ratio to observe the SHVC coding performance. Due to variable behavior through the sequences, we provide separated performance for each dataset and tested configuration. The encodings were run on the SHM9.0 [6] reference software in scalable main profile, using Random-Access configuration. The targeted bitrates are reached thanks to the $R-\lambda$ [7] rate control algorithm, with CTU-level granularity enabled. The single-layer (SL) HEVC encoding are carried out by the same software in SL mode, which is equivalent to an HM16.3 [8] encoding.

5.3.3 Overview of the results

In this section we provide the SHVC performance for the considered video contents in different scalability configurations including spatial, CGS and CGS+HDR. Figure 5.3 illustrates the SHVC performance in terms of BD-PSNR and DB-rate for four cases, representative of the dataset. The exhaustive results with a brief analysis are provided in Appendix A.2 .

Fig. 5.3 Illustration of the ratio impact for SHVC vs SL-HEVC scenario

We can first notice in Figure 5.3a the case where there is a worst ratio point, located at τ =0.4 in this example. Indeed, we can notice that from this point, the ratio raises –then the loss of bitrate in the EL– is compensated with better and more efficient inter-layer predictions in the EL, which leads to a decreasing overhead. In Figure 5.3b, the case where scalability is not efficient enough is illustrated. We can observe that the bigger the ratio is, the bigger the overhead is. Indeed, *ParkDancers* contains a lot of thin textures such as grass, these textures are definitively lost in the down-sampling process and cannot be reconstructed from the BL, even with an high BL bitrate. The behavior illustrated in Figure 5.3c for *Tilbull* refers to the case where there is an optimal ratio point. We can notice that the performance curve is concave. Thus, there is a best ratio point approximately located at 0.5 which is optimal. Finally, the Figure 5.3d illustrates the case where the ratio raising will always reduce the overhead. In a similar way, experiments for scenarios n^o1 and 2 lead to results showing that tweaking the bitrate ratio would lead to significant variations in performance.

5.3.4 Tracks for performance improvement

To illustrate the potential gain that a variable bitrate ratio adjustment can provide, we consider the examples described in Table 5.4 where bandwidth and quality constraints are specified. For each of these examples, we can find an optimal ratio which respects all of the constraints.

	\vert Ex. \vert Sequence	$\mid Q$ BLmin \mid	R_{ELmin}	R_{ELmax}	R_G
a)	\parallel Studio Dancer \parallel		42 dB 9 Mbit/s 12Mbit/s 15 Mbit/s		
b)	\parallel Wind Wool		44 dB \mid 10 Mbit/s \mid 14 Mbit/s \mid 20 Mbit/s		
c)	\parallel Fountain Lady	39 dB	5 Mbit/s	7 Mbit/s \parallel 10 Mbit/s	

Table 5.4 Examples of application

These constraints are defined in the Table 5.1 and are linked with a minimal quality on BL and a limited bandwidth on the EL. All these constraints can be represented as a particular ratio interval. This interval can be computed in a straightforward way, as described in the following Equation.

$$
\tau_G = \left[\frac{\Gamma_{BL}(Q_{BLmin})}{R_G}, 1\right] \cap \left[1 - \frac{R_{ELmax}}{R_G}, 1 - \frac{R_{ELmin}}{R_G}\right]
$$
(5.4)

where $\Gamma_{BL}(Q)$ is the function linking the BL quality using PSNR metric (dB) with its bitrate. We can observe in Figure 5.4 the representation of the ratio interval with its optimal points.

Fig. 5.4 Illustration of potential gains that dynamic ratio adjustment can provide

In Example a), a minimal quality of $42dB$ is required for the BL, with $\{9,12\}$ Mbit/s available bitrate range for the EL, and a global bitrate of 15Mbit/s. In this case, we can notice that the optimal ratio for quality optimization differs from the bitrate optimization one. Indeed, selecting $\tau \approx 0.4$ enables to reduce bitrate losses from 23% to 17%, while selecting $r \approx 0.23$

enable 0.1dB of PSNR improvement.

In Example b), the bitrate reduction in the EL caused by ratio rising is not compensated by IL prediction gain. In this case, more IL prediction rising does not improve the EL quality, then the lower bound of the τ_G is the optimal ratio.

In Example c), the optimal ratio is the upper-bound of τ_G and reduces BD-PSNR and BD-BR. Regarding to all these results and observations, we can conclude that ARC schemes using dynamic ratio adjustment can potentially provide significant performance improvements. These results have been published in [9] and [10] and are the basis of the following work.

5.4 Fixed-interval adaptive rate control

5.4.1 Preliminary

To perform ARC on the top of the current RC implementation, we choose to work at GOPlevel granularity. This way, our goal is to dynamically adjust the distribution between both BL and EL GOP included in the dual-GOP. We choose to work at GOP level granularity because it perfectly fits with the λ -domain RC architecture. Before considering complex methods with quality and bitrate constraints, the most simple approach consists in working in a fixed ratio interval, and to optimize the encoding by selecting the ratio that enables the highest SHVC coding performance in this interval. In this section, we develop this approach where the encoding is optimized according to a fixed ratio interval defined as follows:

$$
\Phi_0 = [(1 - \omega) \times \tau_0, (1 + \omega) \times \tau_0]
$$
\n(5.5)

where τ_0 is the center of the interval fixed to 0.35 in our experiment. This value refers to the usual bitrate ratio between HDTV and UHDTV resolutions, close to the average ratio generated by using the CTC (i.e 0.37). The coefficient ω is a weighting factor set to 0.25. With this configuration, our method dynamically allocates the ratio in the interval [0.26, 0.44].

5.4.2 Proposed algorithm

GOP-level adaptive rate control

The proposed algorithm substitutes both GOP-level bit allocation of BL and EL in the SHM9.0 and replaces it by a dual-GOP management algorithm, as illustrated in Figure 5.5, where the encoding constraint is to select the ratio in Φ_0 . The underlying frame and CTUlevel algorithms are unchanged and are used to achieve the allocated bitrate of each layer. To

Fig. 5.5 Block diagram of the SHVC encoder for 2x spatial scalability (HD to UHD), with the proposed method represented in red

initialize the encoding process, the average targeted bitrate per dual-Picture is first computed as described in Equation 5.6 and is later considered as an encoding constant.

$$
R_{DualPic} = \frac{R_G}{f_r} \tag{5.6}
$$

Before each dual-GOP encoding, the algorithm firstly updates the targeted bitrate in the upcoming one. The average number of bits per dual-Picture for the rest of the sequence *TDualPic* is firstly updated:

$$
T_{DualPic} = \frac{R_{DualPic} \times (N_{Coded} + SW) - (B_{BL} + B_{EL})}{SW}
$$
\n(5.7)

where *NCoded* is the number of already coded dual-Pictures and *SW* the smoothing window width. Then, the global targeted number of bits for upcoming dual-GOP is computed as:

$$
T_{DualGOP} = T_{DualPic} \times N_{DualGOP}
$$
\n
$$
(5.8)
$$

where *N*_{DualGOP} is the number of dual-Pictures in the upcoming dual-GOP. In order to efficiently allocate this budget within both layers, the optimal bitrate ratio τ*opt* has to be determined. This optimal ratio is the solution of an optimization problem which can be formulated as:

$$
\tau_{opt} = \underset{\tau \in \Phi_0}{\arg \max} G(\tau) \tag{5.9}
$$

where Φ_0 is the previously defined fixed ratio interval and $G(\tau)$ the estimated performance function which links the ratio τ with a performance indicator. In order to smooth the bitrate variations, the optimal ratio value is clipped within the following interval:

$$
\max(\tau_{last} \times 0.8, \tau_{min}) \le \tau_{opt} \le \min(\tau_{last} \times 1.2, \tau_{max})
$$
\n(5.10)

with τ_{last} the ratio measure on the last dual-GOP, and (τ_{min}, τ_{max}) the minimal and maximal authorized ratios. Once the optimal ratio is set, the related targeted number of bits for each layer T_{BL} and T_{EL} can be computed as:

$$
T_{BL} = \tau_{opt} \times T_{DualGOP} \tag{5.11a}
$$

$$
T_{EL} = (1 - \tau_{opt}) \times T_{DualGOP}
$$
 (5.11b)

These bitrates feed the underlying frame and CTU level algorithms which accurately undertake targeted bitrates reaching. After each dual-GOP, the parameters of performance function $G(\tau)$ are updated, as described in the following Section.

Estimation of gain function *G*

The performance function *G* has to be estimated to solve the optimization problem described in Equation 5.17. In Figure 5.6, the BD-BR gains in the EBU data set are provided together with EL quality according to the ratio. We can notice that the EL quality in PSNR is highly correlated with the bitrate losses.

This is logical since BD-BR score in scenario n° 3 is based on EL-PSNR. Since Φ_0 only concerns a short portion of the [0,1] interval, performing a parabolic estimation on this interval may be inaccurate because of GOP-level local fluctuation. This way, we consider a simple first order polynomial to locally represent the gain function on Φ_0 :

$$
G(\tau) = Q_{EL}(\tau) \triangleq \alpha \times \tau + \beta \tag{5.12}
$$

After the ith dual-GOP, the achieved ratio τ_{Gi} and the quality Q_{Gi} in the EL GOP are measured. The global YUV-PSNR of each frame *k* in the EL-GOP of the i*th* dual-GOP is computed as follows [11]:

$$
q_i(k) = \frac{6 \times Y_{PSNR}(i,k) + U_{PSNR}(i,k) + V_{PSNR}(i,k)}{8}
$$
(5.13)

Fig. 5.6 BD-BR and EL-PSNR as a function of τ

Then, the average YUV-PSNR value in the EL GOP of the i*th* dual-GOP is computed with Equation 5.14.

$$
Q_{Gi} = \frac{1}{N_{GOP}(i)} \sum_{k=1}^{N_{GOP}(i)} q_i(k)
$$
\n(5.14)

where $N_{GOP}(i)$ is the number of picture in the EL-GOP of the ith dual-GOP. The effective GOP ratio is computed as:

$$
\tau_{Gi} = \frac{R'_{BL}(i)}{R'_{BL}(i) + R'_{EL}(i)}
$$
\n(5.15)

where $R'_{BL}(i)$ and $R'_{EL}(i)$ are respectively the BL and EL bitrates measured on the ith dual-GOP. Considering that a buffer of *N* previous samples (τ_i, q_i) is maintained, the α and β parameters can be estimated via the least-squares method by solving the following system:

$$
\begin{pmatrix} \sum \tau_{G_i}^2 & \sum \tau_{G_i}^1 \\ \sum \tau_{G_i}^1 & \sum \tau_{G_i}^0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sum \tau_{G_i} Q_{G_i} \\ \sum Q_{G_i} \end{pmatrix}
$$
\n(5.16)

With the estimated $\hat{\alpha}$ and $\hat{\beta}$ the optimization problem in Equation 5.17 can be easily solved, and the appropriate ratio is applied to the upcoming dual-GOP bitrate.

5.4.3 Experimental procedure and dataset

To evaluate the performance of the proposed method, this straightforward ARC algorithm has been implemented in the SHM9.0 and tested on the EBU UHD-1 data set which provides ten 3840x2160p sequences used as EL. The BL are built with the down-sampling tool provided in the reference software. The proposed algorithm is compared to the SL-Equivalent UHDTV HEVC encoding which refers to the scenario n° 3, and the resulting BD-BR performance is noted *GARC*. The SHM9.0 using constant 0.35 ratio to build bitrate targets is also compared to SL-Equivalent UHDTV HEVC encoding, and the resulting BD-BR performance is noted G_{Ref} . The BD-BR between proposed and reference algorithm (i.e R- λ in fixed-ratio mode) is also computed and noted G_X . The whole results are recorded in Table 5.5 and the R-D curves of several sequences are provided in Figure 5.7. A positive value for G_{Ref} , G_{ARC} and *G^X* means losses.

Sequence	G_{Ref}	G_{ARC}	G_X
Candle Smoke	18.97%	12.89%	$-6.08%$
Fountain Lady	15.68%	16.16%	$+0.48\%$
Lupo Boa	11.31%	10.46%	$-0.85%$
Lupo Confetti	13.35%	11.60%	$-1.75%$
Park Dancers	23.80%	22.22%	$-1.58%$
Pendulus Wide	30.56%	26.49%	-4.07%
Studio Dancer	18.92%	14.57%	$-4.35%$
Waterfall Pan	20.72%	14.61%	$-6.11%$
Wind Wool	19.64%	9.83%	$-9.81%$
Veggie Fruits	31.33%	22.91%	$-8.42%$
Average	20.43%	16.17%	$-4.25%$

Table 5.5 Performance of the proposed fixed-ratio ARC

5.4.4 Analysis and discussion

We can notice that the proposed ARC scheme improves performance compared to fixed-ratio encoding. The average losses are reduced by almost 20% in average, with a 15.5 standard deviation. All sequences are improved except for *Fountain Lady* which is slightly reduced by 0.48%. In this particular case, the fixed ratio value corresponds to the optimal ratio. The best optimization is reached in *Wind Wool* with 50% of losses reduction. The crossed BD-BR *G^X* is logically good with an average bitrate reduction of 4.25%, with a 9.8% peak and a -0.48% worst case. Regarding to computational complexity, the proposed method does not bring significant complexity since it substitutes the GOP-level existing method. The only additional complexity is the matrix inversion performed in Equation 5.16 by using a 2nd order Gaussian elimination, this step is only performed once before each dual-GOP encoding so around 93 times in the whole encoding (for this 10 seconds duration test set). The encoding time may change since a different ratio means a different targeted bitrate for each layer. If a

smaller ratio than reference 1/2 √ 2 is used, then a bigger bitrate is targeted in the UHD EL which leads to longer encoding time. If a bigger ratio than reference $1/2$ √ 2 is selected, a lower bitrate will be targeted in the EL which results in a smaller encoding time.

Fig. 5.7 Overview of fixed-ratio ARC RD Curves

5.5 Constrained adaptive rate control

5.5.1 Preliminary

We have seen in the previous Section that efficient and straightforward ARC can be performed based on adaptive ratio adjustment. In this Section, we develop a more advanced approach where encoding is subjected to several constraints, defined in Table 5.1. Indeed, the proposed method considers all the encoding constraint and estimate the ratio interval which respects all of them. Then, the optimal ratio is selected in this interval, so as to improve coding efficiency.

5.5.2 Proposed algorithm

The proposed algorithm aims at dynamically adjusting the bitrate ratio τ while encoding to optimize the performance of SHVC and with respects to bitrate and quality constraints, under a global targeted bitrate. We choose to re-use and modify the current algorithm implemented in the SHM reference software, but the proposed method can be simply transposed to other scalable video coders. In the current SHM reference software, the $R-\lambda$ algorithm is duplicated in each layer. In Figure 5.5,our proposed method is illustrated in red where the encoding

constraints refer to BL/EL quality and bandwidth requirements. We can observe that the reference implementation provides separate RC in each layer, with separated targeted bitrates. To summarize, our method substitutes the GOP-Level RC scheme currently implemented in the $R-\lambda$ method. However the frame-level and CTU-level algorithms are maintained. This way, several parameters are updated through encoding such as the number of bits already written in the BL and EL noted *B_{BL}* and *B_{EL}* respectively and the number of already coded dual-picture *NC*. In a similar way as fixed-interval ARC method, the average targeted number of bits per dual-Picture *RDualPic* is firstly computed with Equation 5.6. Then, the updated targeted number of bits per dual-picture for the rest of the sequence *TDualPic* is computed before each GOP, as described in Equation 5.7. The targeted number of bits for the upcoming dual-GOP *TDualGOP* is computed with Equation 5.8. Let *G* be the function linking the bitrate ratio with a performance indicator, then the optimal ratio estimation is the solution of an optimization problem which can be formulated as:

$$
\tau_{opt} = \underset{\tau \in \Phi_G}{\arg \max} \ G(\tau) \tag{5.17}
$$

where Φ_G is the global ratio interval, which respect constraints of both BL and EL. Let Φ_{BL} and Φ_{EL} respectively be the intervals respecting BL and EL constraints including quality and bitrate, then Φ_G can be computed as:

$$
\Phi_G = \begin{cases} \Phi_{BL} , & if \ \Phi_{BL} \cap \Phi_{EL} = \emptyset \\ \Phi_{BL} \cap \Phi_{EL} , & else \end{cases}
$$
\n(5.18)

This way, a minimal BL service is provided when being compliant with all the constraints is impossible. Once τ_{opt} is set, it is clipped into the interval defined in Equation 5.10 to smooth the bitrate variations. Then, and as described in Equations 5.20a and 5.20b, the optimal ratio is applied to dual-GOP targeted bitrate to get the targeted bitrates for BL and EL. To reach the per-layer bitrate, the classic $R-\lambda$ GOP-level RC algorithm is used in each layer. Eventually, the quality and bitrate produced by each layer in the dual-GOP are collected to update $G(\tau)$ as well as the R-D functions $\Gamma_{BL}(D)$ and $\Gamma_{EL}(D)$ used in the next Sections.

Estimation of Φ*BL*

Let Γ*BL* be the GOP-level BL R-D functions which links the BL distortion with its bitrate per frame. Since quality and bitrate constraints are specified, the BL interval (Φ*BL*) can be also computed as the intersection of quality and bitrate sub-intervals $\Phi_{Q_{BL}}$ and $\Phi_{R_{BL}}$:

$$
\Phi_{BL} = \Phi_{Q_{BL}} \cap \Phi_{R_{BL}} \tag{5.19}
$$

The quality interval ($\Phi_{Q_{BL}} = [\tau_{Q_{BLmin}}, \tau_{Q_{BLmax}}]$) can be computed as follows:

$$
\tau_{QBLmin} = \frac{1}{T_{DualPic}} \times \Gamma_{BL} \left(\frac{Q_{BLmin} \times N_F - Q_{BL} \times N_C}{N_F - N_C} \right)
$$
(5.20a)

$$
\tau_{QBLmax} = \frac{1}{T_{DualPic}} \times \Gamma_{BL} \left(\frac{Q_{BLmax} \times N_F - Q_{BL} \times N_C}{N_F - N_C} \right)
$$
(5.20b)

The bitrate interval ($\Phi_{R_{BL}} = [\tau_{R_{BLmin}}, \tau_{R_{BLmax}}]$) is computed as follows:

$$
\tau_{R_{BLmin}} = \frac{B_{BLmin} - B_{BL}}{T_{DualPic} \times (N_F - N_C)}
$$
(5.21a)

$$
\tau_{R_{BLmax}} = \frac{B_{BLmax} - B_{BL}}{T_{DualPic} \times (N_F - N_C)}
$$
(5.21b)

If the constraints Q_{BLmin} and R_{BLmin} are not specified, then the related $\tau_{Q_{BLmin}}$ and $\tau_{R_{BLmin}}$ are not computed while encoding and are set to zero. In the same way, unspecified values for Q_{BLmax} and R_{BLmax} led to avoiding $\tau_{Q_{BLmax}}$ and $\tau_{R_{BLmax}}$ computation and force them to one.

Estimation of Φ*EL*

The EL interval Φ*EL* is also computed as the intersection between quality and bitrate constraints intervals:

$$
\Phi_{EL} = \Phi_{Q_{EL}} \cap \Phi_{R_{EL}} \tag{5.22}
$$

The quality interval ($\Phi_{Q_{EL}} = [\tau_{Q_{ELmax}}, \tau_{Q_{ELmin}}]$) can be computed as follows, with Γ_{EL} the GOP-level EL R-D function :

$$
\tau_{QELmin} = 1 - \frac{1}{T_{DualPic}} \times \Gamma_{EL} \left(\frac{QELmin \times N_F - Q_{EL} \times N_C}{N_F - N_C} \right)
$$
(5.23a)

$$
\tau_{Q_{ELmax}} = 1 - \frac{1}{T_{DualPic}} \times \Gamma_{EL} \left(\frac{Q_{ELmax} \times N_F - Q_{EL} \times N_C}{N_F - N_C} \right)
$$
(5.23b)

The bitrate interval ($\Phi_{R_{EL}} = [\tau_{R_{ELmax}}, \tau_{R_{ELmin}}]$) can be computed as follows:

$$
\tau_{R_{ELmin}} = 1 - \frac{B_{ELmin} - B_{EL}}{T_{DualPic} \times (N_F - N_C)}
$$
(5.24a)

$$
\tau_{R_{ELmax}} = 1 - \frac{B_{ELmax} - B_{EL}}{T_{DualPic} \times (N_F - N_C)}
$$
(5.24b)

In the same way as BL, interval computations are bypassed if these constraints are not specified. If the constraints Q_{ELmin} and R_{ELmin} are not specified, then the related $\tau_{Q_{BLmin}}$ and

τ*RBLmin* are not computed while encoding and set to zero. Unspecified values for *QELmax* and *RELmax* lead to avoid $\tau_{Q_{EImax}}$ and $\tau_{R_{EImax}}$ computation and force them to one.

Estimation of performance function $G(\tau)$

As mentioned in the previous sections, the proposed method is based on the optimization of the performance function $G(\tau)$. Due to performance computation in SHVC for scenario n^o3, the performance function is naturally correlated with the EL quality function since EL distortion is considered in performance computation. This correlation is illustrated in Figure 5.6. Based on this statement, the function $G(\tau)$ can be characterized as a quadratic function, described in Equation 5.25. This quadratic model is more appropriate than the first-order model used in the fixed-ratio ARC, especially on wide ratio intervals.

$$
G(\tau) = Q_{EL}(\tau) \triangleq a \times \tau^2 + b \times \tau + c \tag{5.25}
$$

After the *i*th dual-GOP, the achieved ratio τ_{Gi} and the quality Q_{Gi} in the EL GOP are measured in the same way as fixed-ratio ARC scheme, following the steps defined in Equations 5.13, 5.14 and 5.15. To estimate the parameters (a, b, c) used in Equation 5.25, a LUT linking the samples (Q_G, τ_G) is updated in the encoder. Based on these samples, a least-squares method is performed in Equation 5.26 on the recorded samples in order to estimate the current model parameters \hat{a} , \hat{b} and \hat{c} .

$$
\begin{pmatrix}\n\Sigma \tau_{Gi}^4 & \Sigma \tau_{Gi}^3 & \Sigma \tau_{Gi}^2 \\
\Sigma \tau_{Gi}^3 & \Sigma \tau_{Gi}^2 & \Sigma \tau_{Gi}^1 \\
\Sigma \tau_{Gi}^2 & \Sigma \tau_{Gi}^1 & \Sigma \tau_{Gi}^0\n\end{pmatrix}\n\begin{pmatrix}\n\hat{a} \\
\hat{b} \\
\hat{c}\n\end{pmatrix} =\n\begin{pmatrix}\n\Sigma \tau_{Gi}^2 Q_{Gi} \\
\Sigma \tau_{Gi}^1 Q_{Gi} \\
\Sigma \tau_{Gi}^0 Q_{Gi}\n\end{pmatrix}
$$
\n(5.26)

With the estimated \hat{a} , \hat{b} and \hat{c} the optimization problem in Equation 5.17 can be easily solved, and the appropriate ratio is applied to the upcoming dual-GOP bitrate.

Estimation of quality functions Γ*BL*(*D*) and Γ*EL*(*D*)

The proposed method enables the specification of bitrate constraints at the encoder input, which leads to R-D function modeling while encoding. Commonly, the R-D functions are modelled through quadratic, exponential or polynomial functions to provide accurate representation on a wide range of bitrates. We choose to use the polynomial $log(R)$ -D functions, used in the computation of BD-BD and BD-PSNR. This way, the functions are

defined as following:

$$
\Gamma_{BL}(D) = \log_{10}(R_{BL}) \triangleq a_0 \times D^3 + a_1 \times D^2 + a_2 \times D + a_3 \tag{5.27a}
$$

$$
\Gamma_{EL}(D) = \log_{10}(R_{EL}) \triangleq b_0 \times D^3 + b_1 \times D^2 + b_2 \times D + b_3 \tag{5.27b}
$$

In the same way as performance function estimation, a LUT of *N* previous samples is hold and updated while encoding for both layer functions. Two least-squares estimations are also performed after each GOP in order to update the $\Gamma_{BL}(D)$ and $\Gamma_{EL}(D)$ functions.

Models initialization

In order to begin the encoding process with the most appropriate models for performance functions including $G(\tau)$, $\Gamma_{BL}(D)$ and $\Gamma_{EL}(D)$, several models have been extracted from the statistical study. These models, illustrated in Figure 5.8, suit well the selected dataset and represent a good starting point for a quick convergence towards the optimal models of each sequence. These models are firstly used in their initial versions. Once the first dual-GOP is encoded, these models are scaled according to the measured ratio, BL/EL bitrate and distortion values. Thereafter, the previously described updating process based on LUT is performed for the rest of the sequence.

Fig. 5.8 Initial and scaled versions of $G(\tau)$ and $\Gamma_{BL}(D)$

5.5.3 Use-cases, experimental procedure and dataset

Two industrial SHVC UC can be addressed by the proposed ARC scheme. First of all, an Hybrid Broadcast Broadband (HBB) scheme is considered where the HDTV BL is transmitted via terrestrial or satellite broadcasting while the UHDTV is sent via broadband network (UC1). In this UC, a minimal quality is required for broadcast customers who only receive the HDTV signal. In addition, the UHDTV customers may have bandwidth limitation

imposed by the IP network, then an EL bitrate constraint can also be specified. Moreover, a constant global bitrate is desired for the sum of both layers. These parameters are available at the input of the encoder and are specified in Table 5.1 as *RG*, *QBLmin*, *RELmin* and *RELmax* respectively for global bitrate, minimal BL quality, minimal and maximal EL bitrates.

Configuration	Q_{BLmin}	R_{ELmin}	R_{ELmax}	R_G
	38 dB	2.5 Mbps	4 Mbps	5 Mbps
	40 dB	5 Mbps	8 Mbps	10 Mbps
	41 dB	10 Mbps	14 Mbps	15 Mbps
	42 dB	10 Mbps	16 Mbps	20 Mbps

Table 5.6 Tested configurations for UC1 and UC2

Secondly, a DVB UC (UC2) is studied where both BL and EL are respectively associated with UHD-1 Phase 1 and UHD-1 Phase 2 signal. These phases are related to DVB plans for introducing UHDTV services in several phases, with a mandatory Backward Compatibility (BC) with legacy UHD-1 Phase 1 receivers. In this case the BL is conforming to an UHD-1 Phase 1 signal while the EL is UHD-1 Phase 2. Then, the BL is an UHD signal in a BT.709 colorspace and coded over 10 bits in a SDR dynamic range while the EL is BT.2020-10 bits in HDR dynamic range. In this UC, we can imagine that each phase can be transmitted to a scalable layer to address all receivers. In this case, the Phase-1 and Phase-2 would be placed in the same multiplex, with a mandatory constant bitrate for the sum of both layers. As for UC1, parameters are specified at the input of the encoder and are identified in Table 5.1, with *R^G* the global bitrate, *QBLmin*, *RELmin* and *RELmax* respectively for global bitrate, minimal BL quality, minimal and maximal EL bitrates. The tested configurations are recorded in Table 5.6.

The proposed method has been implemented in the latest SHM9.0 reference software. We run both the classic SHVC encoder with and without our method. The constraints for both use-cases are recorded in Table 5.6 for HBBTV and DVB UCs. To evaluate the performance of the proposed ARC, we compare it with a fixed-ratio reference which respects both BL and EL constraints. Thus, we choose to select as a reference the first fixed-ratio encoding which satisfies the constraints of the given configuration. If no compliant bitstream can be built, the ratio producing the BL minimal quality constraint is selected, this choice guarantees a minimum QoS for BL users. The proposed algorithm is compared to the SL-Equivalent UHD HEVC encoding, and the resulting BD-BR performance is noted *GARC*. The fixed-ratio encoding is compared to SL-Equivalent UHD HEVC encoding, and the resulting BD-BR performance is noted G_{Ref} . The BD-BR between proposed and reference algorithm is also computed and noted G_X . It must be noted that these scores reflect average values which can

be computed on different bitrate intervals, this is also the reason why G_X is not necessarily equal to $G_{Ref} - G_{ARC}$. .A positive value for G_{Ref} , G_{ARC} and G_X means losses. For SDRto-HDR scalability, we also provide BD-BR performance based on more appropriate HDR metrics, computed in the linear light field. Thus, we provide mPSNR [12], tPSNR [5] and deltaE [13] based metrics, computed thanks to the HDR Tools software [14]. In addition, R-D curves using the HDR-VQM [15] metric which is more correlated to the subjective tests is provided in Figure 5.9 for HDR content. The accuracy of the proposed method for each use case is also provided in Table 5.9 in terms of mean error and standard deviation. The UC 1 is evaluated on EBU and Technicolor sequences which provides HD to UHD scalability and the results are provided in Table 5.7. The second UC is evaluated on Volvo Ocean Race content which is suitable for UHD1-P1 to UHD2-P2 scalability, the results for all metrics are provided in Table 5.8.

5.5.4 Analysis and discussion

Use-case 1: HBBTV service

As mentioned previously, this UC deals with HD to UHD spatial scalability. Thus, the performance is evaluated for sequences included in EBU-UHD1 and CTC-2x-CGS, and the results are provided in Table 5.7. First, we can notice that our method reduces the average overhead from 32.13% (*GRe f*) to 24.60% (*GARC*), which represents a reduction of 20.45% and a BD-Rate improvement of -7.51% (G_X). More specificaly, we can notice that our method significantly outperforms the fixed ratio encoding for 10 sequences ($G_X < -1\%$), is equivalent for 3 sequences ($|G_X| \leq 1\%$) and is less efficient for 2 sequences ($G_X > 1\%$). However, we can notice that for Birthday, the overhead versus SL encoding is well reduced from 37.44% to 32.27%, but this is not translated into a gain $(G_X = +0.13\%)$. This strange behavior is due to the way of achieving BD-BR computation [1]. Indeed, the PSNR interval used to compute the average BD-BR gain between ARC and fixed-reference R-D curves may differ from the ones used to compute G_{ARC} and G_{Ref} .

Use-case 2: DVB UHD1-P1 to UHD1-P2

For this experiment, the results based on a YCbCr space PSNR are recorded in Table 5.8 with linear-light HDR based metrics are provided. First, in a PSNR-based point of view we can notice that our method is clearly more efficient than fixed-ratio encoding since average overhead is reduced from 11.04% to 2.28% with a related BD-BR improvement of -8.30% (*GX*). We can observe that our method clearly outperforms the fixed-ratio encoding for 4 sequences ($G_X < -1\%$) and is less efficient for one sequence ($G_X > 1\%$). In addition, our

Sequence	G_{Ref}	G_{ARC}	G_X
Candle Smoke	7.94%	5.41%	$-2.53%$
Fountain Lady	31.22%	29.65%	-1.57%
Lupo Boa	12.75%	8.61%	$-4.14%$
Lupo Confetti	16.70%	20.09%	$+3.40\%$
Park Dancers	61.87%	35.98%	-25.89%
Pendulus Wide	59.79%	43.05%	$-16.73%$
Studio Dancer	20.35%	19.77%	-0.57%
Waterfall Pan	35.04%	18.67%	-16.36%
Wind Wool	30.75%	14.09%	-16.66%
Veggie Fruits	39.90%	22.97%	-16.93%
Birthday	37.44%	32.27%	$+0.13%$
BirthdayFlashP1	24.33%	27.94%	$+2.99\%$
BirthdayFlashP2	24.25%	25.50%	$+0.91%$
Parakeets	24.77%	15.20%	-5.78%
TableCar	54.78%	49.79%	-13.00%
Average	32.13%	24.60%	-7.51%

Table 5.7 ARC performance for Use-Case 1

method enables to avoid the overhead for 4 sequences $(G_{ARC} < 0\%)$.

Table 5.8 UC2 performance based on HDR metric

Sequence	G_{Ref}	G_{ARC}	$G_X(PSNR)$	$G_X(mPSNR)$	$G_X(tPSNR)$	$G_X(deltaE)$
Boat1	10.20%	-5.47%	-14.69%	-20.39%	-9.32%	-47.35%
Boat2	-1.07%	-5.34%	-4.87%	-15.35%	-18.50%	$-39.23%$
Crew	20.82%	23.91%	1.01%	2.59%	-13.71%	-28.50%
Crowd	14.47%	-1.35%	$-12.78%$	-17.69%	$-27.72%$	-50.81%
Pontoon	10.78%	-0.33%	$-10.19%$	$-14.42%$	-47.49%	-52.77%
Average	11.04%	2.28%	-8.30%	-13.05%	-23.35%	-43.73%

For HDR-based metrics, we can observe that the proposed method is more efficient for all tested metrics with -13.05%, -23.35% and -43.73% of average gain for mPSNR, tPSNR and deltaE, respectively. These metrics are currently discussed in MPEG and are questionable because none of these is correlated enough with subjective quality assessment. In Figure 5.9 the R-D curves for UC using the HDR-VQM metric are provided to illustrate the performance of the proposed method where the lower the distorsion value is, the better the subjective quality is supposed to be. We can notice that the results obtained with other metrics are confirmed and that our proposed method outperforms the fixed-ratio encoding for 4 sequences in the dataset. It can be observed that the Crew sequence, where losses were observed, is still underperforming compared to the fixed-ratio encoding.

Fig. 5.9 R-D curves of UC2 using HDR-VQM as distorsion metric

Discussion on accuracy

In Table 5.9, the accuracy of the proposed methods for configurations in Table 5.6 is provided for both UCs. A mean value over 100% for *QBLmin* and *RELmin* means that the constraint is fully respected. The accuracy related to *RELmax* is not provided since this constraint is always respected through the experiments. First, it must be noted that the selected configurations are generic and are thus potentially not fully reachable for some sequences. However, we can notice that this case is treated in Equation 5.18. Indeed, if $\Phi_{BL} \cap \Phi_{EL} = \emptyset$, then the BL interval is only selected so as to provide, at least, an appropriate BL service. The accuracy provided in Table 5.9 illustrates this BL priority since we can observe that the BL minimal level of quality is well maintained regarding to the mean and standard deviation values. We can notice, especially for configurations 1, 2 and 3 of UC 1, that the level of quality is observed (with tolerable standard deviation) to the detriment of the minimal restricted bitrate on the EL. This underflow in the EL channel could be filled by regrouping several EL in a channel by using StatMux.

Discussion on potential performance improvement

We have seen that the proposed method enables good performance, in average. However, the fixed-ratio encoding seems better for several sequences such as *BirthdayFlashP1* or *Lupo Confetti* for example. On one hand, we can observe in the Figures provided in Annex A.2 that several sequences do not offer great improvement prospective. For the configurations described in Table 5.6, we can observe that the EL bandwidth constraints enable an initial Φ_{EL} approximately equal to [0.2, 0.5] which can be reduced according to the BL quality constraint. For several sequences such as *Birthday* or *BirthdayFlashP2*, we can observe in Figure A.16 that the performance curve on this initial interval is totally flat and then any ratio should produce the same performance.

Fig. 5.10 Alternative model initialization for $G(\tau)$

On the other hand this can be explained by no suitable initialization model for $G(\tau)$, $\Gamma_{BL}(D)$ and $\Gamma_{EL}(D)$. For long sequences included in EBU UHD-1 or Volvo Ocean Race content, the use of an unoptimized model is not necessarily critical since the encoder has 10s of content to converge towards appropriate models. For short sequences, including for example the Technicolor CGS dataset, the encoder has only 5s which can be not enough to build appropriate models, and thus not enough to perform an efficient ratio selection. In Figure 5.10,

Indicator Use-Cases	Configuration 1			Configuration 2	Configuration 3		Configuration 4		
		Q_{BLmin}	R_{ELmin}	Q_{BLmin}	R_{ELmin}	Q_{BLmin}	R_{ELmin}	Q_{BLmin}	R_{ELmin}
UC 1	mean	104.41%	96.35%	104.75%	100.69%	101.80%	103.41%	103.79%	116.72%
	std	7.07%	6.78%	5.79%	9.63%	5.81%	10.91%	5.19%	20.92%
UC ₂	mean	111.56%	102.09%	110.39%	102.01%	107.98%	102.78%	107.00%	132.77%
	std	8.98%	12.90%	8.16%	16.56%	8.24%	4.16%	8.78%	20.81%
Average Value	mean	107.98%	99.22%	107.57%	101.35%	104.89%	103.09%	105.39%	124.74%
	std	8.02%	9.85%	6.97%	13.09%	7.02%	7.53%	6.48%	20.86%

Table 5.9 Accuracy of the proposed method

an alternative model which is more suitable for several sequences is proposed. For instance, in *Crew* where it enables to reduce the overhead from 23.91% to 21.57% and in *BirthdayFlashP2* where the overhead is reduced from 27.94% to 21.98%. The selection of an optimal model can be implemented in a pre or look-ahead processing stage or before a scene transition, and would increase the performance of the proposed method.

5.6 Conclusion

In this second contribution, we propose an innovative adaptive rate control scheme for SHVC, particularly suitable for new UHD services deployment. We investigate the impact of bitrate ratio over SHVC performance, for several combinations of scalability including, spatial, color-gamut and dynamic-range. This study emphasizes the fact that a dynamic ratio adjustment has a strong interest in SHVC performance improvement.

We firstly investigate a simple approach which consists in selecting the optimal bitrate ratio in a fixed ratio interval. For this straightforward approach, the experiments show an average reduction of 20% on scalable overhead for the EBU-UHD1 data set compared to the fixed-ratio reference coding, with the best and worst performance of 50% and -3% , respectively. In terms of BD-BR performance, this algorithm enables on average a significant gain of -4.25%. However, this approach is interesting but does not consider bitrate and quality constraints on each layer which are mandatory in an industrial context.

Then, we define the BL and EL quality and bandwidth constraints as well as a global targeted bitrate (BL+EL), submitted to the encoder as mandatory parameters. Based on these parameters, our proposed method dynamically computes an interval of possible ratios which respects all constraints and chooses the optimal ratio that minimizes the SHVC coding loss. The algorithm is implemented and tested on two use-cases. First, we deal with spatial scalability from HD to UHD service, for Hybrid Broadcast and Broadband delivery, with a minimal level of quality required on the BL and a restricted bandwidth on the EL. Second, we consider the UHD introduction phases defined by DVB. We provide a scalable scheme insuring backward compatibility between UHD1-Phase1 (UHD-BT.709-SDR) and UHD1-Phase2 (UHD-BT.2020-HDR), with the same quality and bandwidth constraints. Our experiments show that an average improvement of -7.51% and -8.30% can be respectively achieved for HBB and DVB use-cases.

The next step in this work will focus on performing subjective tests to get a confirma-

tion result, especially for HDR where all used metrics are contested. The future work will be dedicated to developing the post-processing stage mentioned during discussions, and to apply this algorithm to statistical multiplexing of several scalable programs.

References

- [1] G. Bjøntegaard, "VCEG-M33: Calculation of Average PSNR Differences Between RD-Curves," Apr. 2001.
- [2] V. Seregim and Y. He, "JCTVC-Q1009: Common SHM Test Conditions and Software Reference Configurations," Mar. 2014.
- [3] EBU, "https://tech.ebu.ch/testsequences/uhd-1."
- [4] P. Andrivon and P. Bordes, "JCTVC-N0163: AhG14 Wide Color Gamut Test Material Creation," July 2013.
- [5] A. Luthra, E. Francois, and W. Husak, "MPEG-M35464: Draft Call for Evidence (CfE) for HDR and WCG Video Coding," Feb. 2015.
- [6] J. Chen, J. Boyce, Y. Ye, and M.-M. Hannuksela, "JCTVC-T1007: SHVC Test Model 9 (SHM 9) Introduction and Encoder Description," Feb. 2015.
- [7] L. Li, B. Li, and J. Zhang, "JCTVC-M0037: Rate Control by R-Lambda Model for SHVC," Apr. 2012.
- [8] K. McCann, C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G.-J. Sullivan, "JCTVC-S1002: High Efficiency Video Coding (HEVC) Test Model 16 (HM 16) Improved Encoder Description," Nov. 2014.
- [9] T. Biatek, W. Hamidouche, J.-F. Travers, and O. Deforges, "Toward Optimal Bitrate Allocation in the Scalable HEVC Extension: Application to UHDTV," *IEEE International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, 2015.
- [10] T. Biatek, W. Hamidouche, J.-F. Travers, and O. Deforges, "ISO/IEC JCT1/SC29/WG11-m36967: On the Performance Evaluation of Scalable Technology in Future Video Coding," Oct. 2015.
- [11] J.-R. Ohm, G.-J. Sullivan, H. Schwarz, T.-K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards–Including High Efficiency Video

Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 22, pp. 1669–1684, Dec. 2015.

- [12] J. Strom and J. Samuelsson, "ISO/IEC JCT1/SC29/WG11-m35060: Fast Calculation of mPSNR HDR Video Error," Oct. 2014.
- [13] CIE, "Colorimetry Part 6: CIEDE2000 Colour-Difference Formula," 2012.
- [14] A.-M. Tourapis and D. Singer, "ISO/IEC JCT1/SC29/WG11-m35156: HDRTools: A Software Package for Video Processing and Analysis," Oct. 2014.
- [15] M. Narwaria, M. Perreira da Silva, and P. Le Callet, "HDR-VQM: An Objective Quality Measure for High Dynamic Range Video," *Signal Processing: Image Communication*, vol. 35, pp. 46–60, 2015.

Chapter 6

Investigations on statistical multiplexing of scalable programs using SHVC

6.1 Preamble

In this Chapter, we initiate some investigations on statistical multiplexing (SMX) for scalable video coding using SHVC. As mentioned in the Chapter 3.2.5, the SMX solution takes advantage of video scenes diversity within the programs pool. Thus, more bitrate is allocated to encode complex programs and less bitrate for programs that can be easily encoded. This enables to significantly improve the coding efficiency, and especially in a subjective quality way. Two main approaches exist in the literature for SMX, corresponding to open-loop and closed-loop schemes. On one hand, the close-loop approach has been investigated in [1] where each encoder received a feedback from the multiplexer (MUX). In this approach, a more accurate and efficient bitrate control is enabled since null-packets can be detected. This approach can be unfavourable since it requires a communication protocol between encoders and the MUX, and thus it is not codec-constructor agnostic. On the other hand, the open-loop approach developed in [2] and [3] decouples the bitrate allocation stage from the stream multiplexing and thus enables to interconnect encoders from various constructors. Our proposed method falls within the open-loop scheme which enables to re-use any existing SHVC implementation jointly with any existing MUX. Figure 6.1 illustrates an overview of the proposed SMX architecture with SHVC encoders in the open-loop paradigm.

We consider in this chapter the scheme illustrated in Figure 6.1. In this scheme, the SMX unit communicates with the rate-control (RC) of the SHVC encoders to compute BL and EL bitrates which fit into the two channels of bandwidth B_{ch1} and B_{ch2} . In addition to

the transmission of targeted bitrates for each program, the SMX unit also gets some basic feedback information from RC units such as achieved bitrate and quality in each layer. This scheme can be connected to a wide variety of RC algorithms which have to be configured in variable bitrate (VBR) mode.

Fig. 6.1 Overview of the architecture using proposed SMX

Concerning the joint bitrate allocation strategy in SMX, several solutions have been investigated. For previous standard such as AVC or MPEG-2, several algorithms have been proposed for instance in [4] and [5] which allocate the bitrates according to a complexity indicator to minimize the quality variation among programs. Contribution proposed in [6] includes a look-ahead processing to detect scene-transition which leads to an improved quality. A number of solutions such as [7] and [8] include temporal bitrate smoothing exploiting inter-frame dependencies to improve coding quality. For SVC, the majority of these contributions exploit the fine granular scalability (FGS) available in SVC. In [9], a complexity indicator is used to evaluate the bitrate allocated to each program, then adaptation engines refine the SVC bitstream of each program to achieve the allocated bitrate. Similar approaches have been proposed in [10] and [11]. In [12], an alternative approach based on coarse grain scalability (CGS) in SVC is proposed for multicast applications. In this method the BL and EL are separately multiplexed in two channels according to a complexity estimator which weights the programs. Finally, the contribution [13] highlights the impact of smoothing on SMX, for AVC and SVC dyadic spatial scalability. We can notice that no contribution jointly adjusts the global bitrate of each program (BL+EL) in CGS mode with the bitrate ratio between layers. In Section 5.3, a deep study on the impact of the bitrate ratio between layers on SHVC performance is provided for two-layers schemes $(L = 2)$, the ratio

being defined as:

$$
\tau_p = \frac{R_{p,1}}{R_{p,1} + R_{p,2}}\tag{6.1}
$$

where $R_{p,1}$ and $R_{p,2}$ are the BL and EL bitrates, respectively of the program p. It appears that the ratio adjustment has a strong impact on coding performance, and that this parameter is crucial in SHVC RC schemes. This assessment is used in Section 5.4 where the ratio is adjusted in a fixed interval, under a global bitrate target ($R_p = \sum_{l=1}^{L} R_{p,l}$), so as to optimize coding performance. This approach is further enhanced in Section 5.5 where bitrate and quality instructions are required at each layer.

In addition, the compression performance is not considered in the existing solutions, which means that scalability is never used in an optimal way. For these reasons, we propose an innovative statistical multiplexing scheme for SHVC which enables the joint encoding of several scalable programs where the BL and EL are transmitted over separated channels with different bandwidths. Our method smooths the quality during encoding to avoid rough variations and provides a consistent quality among programs. Finally, an optimal ratio selection is performed to insure better R-D performance compared to SL HEVC encoding of each program.

6.2 Pre-encoding based statistical multiplexing

6.2.1 Notations and definitions

In the proposed solution, we treat the two-layers SHVC case where the BL and EL of *N^P* programs are transmitted over two channels of bandwidths B_{ch1} and B_{ch2} . At each period *T*0, the proposed algorithm first re-evaluates the global bitrates of the programs and then optimises the SHVC ratio of each program. In fact, the proposed SMX modulates the bitrate and the ratio of each program to optimize the SHVC coding performance while limiting rough quality variations in time and providing homogeneous quality among programs. In the Algorithm 1, the proposed SMX is presented in few steps. For each new period (T_0) , the presence of a new configuration is firstly checked. If a scene transition appears in at least one program, then the global bitrate of each program (ie. the bitrate allocated to each SHVC program R_p) is re-allocated (Section 6.2.2). Then, the optimal ratio combination minimizing the coding overhead of SHVC compared to HEVC SL is selected (Section 6.2.3). Finally, the encoding is performed with the selected set-up and the targeted bitrates in all layers are achieved thanks to the $R-\lambda$ CBR control algorithm which is used in a VBR mode.

6.2.2 Global bitrate allocation

When a new content is detected (scene transition in at least one program), the global bitrates of the programs are re-evaluated. As mentioned in Section 6.2.1, the two layers of the *N^P* programs have to be transmitted over two separated channels of capacities B_{ch1} and B_{ch2} , thus the sum over the global bitrate R_p with $p \in \{1,...,N_p\}$ of the N_p programs shall be below the channels bandwidths:

$$
\sum_{p=1}^{N_P} R_p \leq B_{ch1} + B_{ch2} \tag{6.2}
$$

Let $\Gamma_{p,l}(\tau_p, R_p)$ be the R-D function of the *l*th layer of the p^{th} program linking the distortion *D* (i.e. PSNR) with the global bitrate R_p for a given ratio τ_p . To ensure distortion consistency among all the programs, we choose to select the combination of global bitrates R_p with p $\in \{1,...,N_P\}$ that minimises the variance between the programs qualities:

$$
\underset{\{R_1,\ldots,R_{N_P}\}}{\arg\min} \sum_{l=1}^{2} Var(E_{p,l}) \qquad \text{s.t. } (6.2) \tag{6.3}
$$

where $E_{p,l}$ is the vector defined as $E_{p,l} = [\Gamma_{1,l}(\tau_1, R_1), ..., \Gamma_{N_p,l}(\tau_{N_p}, R_{N_p})]$ and $Var(E_{p,l})$ is the function computing the variance of the vector $E_{p,l}$. At this step, τ_p values are set to $B_{ch1}/(B_{ch1} + B_{ch2})$ (ie. equal ratio distribuion between layers).

6.2.3 Optimal ratio selection

Once the global bitrate is allocated, the optimal ratio can be adjusted. As mentioned in Section 6.2.1, the sum of bitrates in the BL has to fit in a channel of bandwidth *Bch*1. Hence, the selected ratios have to respect the following relation:

$$
\sum_{p=1}^{N_P} \tau_p \times R_p \leq B_{ch1} \tag{6.4}
$$

Let $G_p(\tau_p)$ be the function linking the bitrate ratio τ_p to a performance indicator such as PSNR of each program (describes in Section 6.2.4). Thus, the optimal ratio balance between programs can be estimated as the solution of the following optimization problem:

$$
\underset{\{\tau_p\}_{p=1..N_p}}{\arg \max} \sum_{p=1}^{N_p} G_p(\tau_p) \qquad \text{s.t. (6.4)} \tag{6.5}
$$

To avoid coarse quality variations, we shall constraint the quality variation while selecting the optimal ratio in Equation 6.5. As we will see in the next Section, the optimization of performance leads to improve the quality of EL. Thus, in order to enables flexibility in ratio choice, we choose to strongly regulate the quality variations in the BL as follows:

$$
|Q_{BL}(\tau_p, t - T_0) - Q_{BL}(\tau_p, t)| \le \Delta_Q \tag{6.6}
$$

where Δ ^{*Q*} is the maximum authorised variation and $Q_{BL}(\tau_p,t)$ is the average quality achieved in the BL for the *t*th period and p th program, computed thanks to the function $\Gamma_{p,1}(\tau_p, R_p)$. Once this step is achieved, the encoding can be performed for the next period with the estimated set-up $\{(R_p, \tau_p)\}_{p=1..N_P}$. This method requires an accurate functions modeling, further described in the next section.

6.2.4 Functions modeling

In the Section 5.5, several modeling of functions $\Gamma_{p,l}(\tau_p,R_p)$ and $G_p(\tau_p)$ have been proposed. If the whole SHVC scheme (BL+EL) is evaluated against the equivalent HEVC SL encoding of the EL, then $G_p(\tau_p)$ can be approximated as the function linking the bitrate ratio τ_p with the distortion of the EL, which leads to the following quadratic formulation:

$$
G_p(\tau_p) = Q_{EL}(\tau_p) \triangleq \alpha_p \times \tau_p^2 + \beta_p \times \tau_p + \gamma_p \tag{6.7}
$$

where $(\alpha_p, \beta_p, \gamma_p)$ are the model parameters for the p^{th} program. The R-D functions can be modelled as a $3th$ order $log(R)$ -based polynomial defined as follows:

$$
\Gamma_{p,l}(\tau_p, R_p) \triangleq a_{p,l} \times Rh_{p,l}^3 + b_{p,l} \times Rh_{p,l}^2 + c_{p,l} \times Rh_{p,l} + d_{p,l} \tag{6.8}
$$

where $Rh_{p,l} = \ln(R_{p,l})$ and $R_{p,l}$ is computed based on the global bitrate R_p of program p and the related ratio τ_p . $(a_{p,l}, b_{p,l}, c_{p,l}, d_{p,l})$ are the model parameters of the *l*th layer of the *p*th program. For each of these models, a look-up-table can be build and use as basis for model estimation. For the R-D functions, this leads to solve a linear equation system under the form $AX = B$ where the vector $X = [\hat{a}_{p,l}, \hat{b}_{p,l}, \hat{c}_{p,l}, \hat{d}_{p,l}]$ for performance estimation of $\Gamma_{p,l}$ and the vector $X = [\hat{\alpha}_p, \hat{\beta}_p, \hat{\gamma}_p]$ for G_p estimation. These models are updated in a similar way as Section 5.5.2, which leads to solve the following equation system for performance function:

$$
\begin{pmatrix}\n\Sigma_{i=1}^{N} \tau_{p,i}^{4} & \Sigma_{i=1}^{N} \tau_{p,i}^{3} & \Sigma_{i=1}^{N} \tau_{p,i}^{2} \\
\Sigma_{i=1}^{N} \tau_{p,i}^{3} & \Sigma_{i=1}^{N} \tau_{p,i}^{2} & \Sigma_{i=1}^{N} \tau_{p,i}^{1} \\
\Sigma_{i=1}^{N} \tau_{p,i}^{2} & \Sigma_{i=1}^{N} \tau_{p,i}^{1} & \Sigma_{i=1}^{N} \tau_{p,i}^{0}\n\end{pmatrix}\n\begin{pmatrix}\n\hat{\alpha}_{p} \\
\hat{\beta}_{p} \\
\hat{\gamma}_{p}\n\end{pmatrix} =\n\begin{pmatrix}\n\Sigma_{i=1}^{N} \tau_{p,i}^{2} Q_{p,i} \\
\Sigma_{i=1}^{N} \tau_{p,i}^{1} Q_{p,i} \\
\Sigma_{i=1}^{N} \tau_{p,i}^{0} Q_{p,i}\n\end{pmatrix}
$$
\n(6.9)

where $(\tau_{p,i}, Q_{p,i})$ are the pair of EL measured distortion and achieved ratio in the p^{th} program, stored into a LUT1D and of size *N*. The R-D functions are computed in a similar way, by solving the following system:

$$
\begin{pmatrix}\n\sum_{i=1}^{M} Rh_{p,l,i}^{6} & \sum_{i=1}^{M} Rh_{p,l,i}^{5} & \sum_{i=1}^{M} Rh_{p,l,i}^{4} & \sum_{i=1}^{M} Rh_{p,l,i}^{3} \\
\sum_{i=1}^{M} Rh_{p,l,i}^{5} & \sum_{i=1}^{M} Rh_{p,l,i}^{4} & \sum_{i=1}^{M} Rh_{p,l,i}^{3} & \sum_{i=1}^{M} Rh_{p,l,i}^{2} \\
\sum_{i=1}^{M} Rh_{p,l,i}^{4} & \sum_{i=1}^{M} Rh_{p,l,i}^{3} & \sum_{i=1}^{M} Rh_{p,l,i}^{2} & \sum_{i=1}^{M} Rh_{p,l,i}^{1} \\
\sum_{i=1}^{M} Rh_{p,l,i}^{3} & \sum_{i=1}^{M} Rh_{p,l,i}^{2} & \sum_{i=1}^{M} Rh_{p,l,i}^{1} & \sum_{i=1}^{M} Rh_{p,l,i}^{1} \\
\sum_{i=1}^{M} Rh_{p,l,i}^{0} & \sum_{i=1}^{M} Rh_{p,l,i}^{0} & \sum_{i=1}^{M} Rh_{p,l,i}^{0}\n\end{pmatrix} = \begin{pmatrix}\n\sum_{i=1}^{M} Rh_{p,l,i}^{3} & \sum_{i=1}^{M} Rh_{p,l,i}^{2} \\
\sum_{i=1}^{M} Rh_{p,l,i}^{2} & \sum_{i=1}^{M} Rh_{p,l,i}^{1} \\
\sum_{i=1}^{M} Rh_{p,l,i}^{0} & \sum_{i=1}^{M} Rh_{p,l,i}^{0} \\
\sum_{i=1}^{M} Rh_{p,l,i}^{0} & \sum_{i=1}^{M} (h_{p,l,i}^{0} & \sum_{i=1}^{M
$$

where $(Rh_{p,l,i}, Q_{p,l,i})$ are the pair of measured distortion and achieved bitrate in logarithmic scale in the *l th* layer of the *p th* program, stored into a LUT1D and of size *M*. The equations systems 6.9 and 6.10 can be easily computed for example by performing a Gauss-Jordan elimination.

Fig. 6.2 Illutration of a N-periods database with M ratios, 5 global bitrates and 2 programs

6.2.5 Pre-encoding based framework

In this first investigation, we apply our SMX method to a pre-encoding scheme where dualchunks (i.e for BL and EL) are available in a database for each program encoded with SHVC in two layers at 10 ratios. Each chunk refers to an independently decodable piece (starting with an IDR picture and of duration T_0) of bitstream which can thus be assembled to form a multiplexed bitstream. Figure 6.2 illustrates an example of database containing two programs. This database is used to build the models presented in Section 6.2.4. Using a pre-encoded bitstream enables to constitute the best achievable multiplexes, and thus gives an overview of the best reachable performance.

Scenario		Hybrid Mobile/DTT Delivery		
	DTT	Mobile	Delivery	
T ₂ Profile	Fixed SFN	Portable SFN	Fixed MFN	
Channel Bandwidth	8 MHz	8 MHz	8 Mhz	
iFFT	8k Ext.	8k Ext.	32k Ext.	
Constellation	256 QAM	16 QAM	256 QAM	
Code-Rate	3/5	1/2	2/3	
LDPC Frame Length	64800 Bits	64800 Bits	64800 Bits	
Pilot Pattern	PP ₄	PP ₂	PP7	
Guard-Interval	1/16	1/8	1/128	
Available Bitrate	33.17 Mbps	12.34 Mbps	40.21 Mbps	

Table 6.1 DVB-T2 Parameters of the Tested Use-Case

Table 6.2 Source coding parameters for SMX evaluation

Use-case	Encoder	B itrates	Configuration	Sequences
Hybrid database	SHM 9.0	$R_p \in \{10, 11, , 15, 16\}$ Mbps $\tau_p \in \{0.1, 0.2, , 0.8, 0.9\}$	2-layers (HD/UHD) Random-access	Programs generated
Hybrid references $3x$ CBR	SHM 9.0	$R_p = 13$ Mbps $\tau_p = \frac{B_{ch1}}{B_{ch1}+B_{ch2}}$		based on EBU UHD-1
Single DTT $3x$ CBR	SHM 9.0	$R_p = 13$ Mbps	Single-layer (UHD) Random-access	dataset $[14]$

6.2.6 Use-cases, experimental procedure and dataset

Our algorithm is experimented on the realistic use-case of hybrid mobile/fixed delivery of three UHD programs jointly encoded using SHVC and illustrated in Figure 6.3. The BL and EL are multiplexed and respectively transmitted through the mobile and Digital Terrestrial Television (DTT) DVB-T2 channels described in Table 6.1. It must be noted that the mobile channel bandwidth is $B_{ch1} = 12$ Mbps while only a portion of the DTT channel is occupied with $B_{ch2} = 27$ Mbps, leaving 6 Mbps for additional services and metadata. In

this experiment, we generate bitstreams of the EBU UHD-1 dataset [14] at six global bitrates $R_p \in \{10, 11, 12, 13, 14, 15\}$ Mbps and nine ratios from 0.1 to 0.9 by step of 0.1.

Fig. 6.3 Overview of the considered hybrid-delivery system

The encoding are carried out thanks to the SHM9.0 reference software [15] using R−λ RC algorithm to reach the targeted bitrate. All the source-coding parameters are more accurately described in Table 6.2. The models used in our method are estimated according to the procedure described in Section 6.2.4 and the ∆*^Q* is set to 0.5 dB in default configuration and varied later for performance evaluation. For each program, we randomly combine 5 different sequences of the EBU UHD-1 dataset. An offset is applied to each program, thus the transitions between sequences are not temporally aligned, as illustrated in Figure 6.4.

Fig. 6.4 Illustration of random program generation

The proposed method is compared to the constant bitrate (CBR) encodings where bitrate of both layers are fairly allocated per channel (i.e fixed ratio encoding). The accuracy reached in each multiplex as well as the global value and the quality fluctuations of these two methods are computed. For quality fluctuations, σ_{PP} refers to the average PSNR

standard-deviation per program while σ_{Mux} refers to the average inter-program standarddeviation. More generally, σ*PP* indicates the quality fluctuations into programs while σ*Mux* indicates the quality homogeneity between programs in the whole multiplex. In addition, the overhead introduced by scalable schemes compared to the SL encoding of the highest layers is computed, for the two tested methods. The SL reference is defined as the separate CBR HEVC encoding of the three programs transmitted through the single DTT channel described in Table 6.1, where 39 Mbps are used to keep the same global bitrate (each program being encoded at 13 Mbps). The SL encoding are carried out by the SHM9.0 used in SL mode.

Table 6.3 Performance of hybrid – 3x VBR programs (proposed SMX)

Channel	Error		Quality variations		Average overhead
	Mean	Variance	σ_{PP}	σ_{Mux}	vs Single-DTT HEVC
Mobile (BL)	-0.12%	0.0007	1.49 dB	3.66 dB	
Fixed (EL)	$+1.26\%$	0.0013	0.82 dB	1.31 dB	7.65%
Global	$+1.02\%$	0.0010	1.15 dB	2.50 dB	

6.2.7 Analysis and discussion

In Tables 6.3 and 6.4, the statistics and metrics described in Section 6.2.6 are recorded for simulations with and without SMX, respectively. The provided values are computed as the average on 500 randomly generated combinations of programs to give a fair performance overview. Regarding to the accuracy, both CBR and VBR approaches are almost identical and provide global errors under 2%, with a variance close to zero. The proposed SMX method provides substantial performance improvement compared to the CBR configuration and enables an overhead reduction from 11.01% to 7.65% which represents a gain of 3.36%. These gains are counterbalanced by a less homogeneous quality for mobile receivers while the EL quality is improved. This is logical since our method limits the quality variations amplitude in the BL while maximizing EL performance, which does not put any constraints on the BL variance. On one hand, we can observe that the σ_{Mux} of our method suggests a less homogeneous quality than CBR approach for mobile receivers. On the other hand, our method outperforms the CBR approach for fixed receivers where it provides more smooth and

homogeneous quality. In addition, our method reduces the quality variations from 1.51 dB to 1.49 dB for mobile receivers (σ_{PP}). Therefore, consuming content in mobility on a small-size displays reduces the coding artefacts annoyance [16] and indicates that the tradeoff between overhead-reduction and BL services degradation is tolerable. For this particular use-case, our method reduces the SHVC overhead, improves the quality of EL services and leaves 6 Mbps on the DTT channel for metadata and additional services. In return, the BL services are reasonably less homogeneous but in a really good compromise. Finally, a subjective quality improvement should be noticeable since with SMX method the highest PSNR are reduced without annoying degradation while lowest PSNR are enhanced with more significant visual improvements.

Fig. 6.5 Quality and performance variations as a function of ∆*^Q*

Fig. 6.6 Illustration of quality variations in mobile channel

In Figure 6.5, the quality variations in mobile services are provided as a function of Δ ^{*Q*} (Maximum enabled BL quality variations during one period) as well as the performance. We can notice in Figure 6.5a that the quality variations increase with the ∆*^Q* which is expected regarding to the Equation 6.6 since more degrees of freedom are given to the ratio variations while encoding. In a similar way, the quality homogeneity also decrease strongly when

the ∆*^Q* is rising, as illustrated in Figure 6.5b. These behaviors are expected since rising the ∆*^Q* leads to more freedom in ratio adjustment, and thus leads to better performance at the expense on higher variations in the quality of the BL. Indeed, we can observe in Figure 6.5c that the overhead can be improved to the detriment of more quality variation in the mobile services. Therefore, the selected Δ ^{*Q*} = 0.5 represents a good compromise between compression efficiency and limitations of quality variations in the BL services. In addition, the quality variations in each program of the mobile channel are illustrated in Figure 6.6. We can observe that the huge quality variations are well reduced and limited in each program with our proposed method. This behavior is limited in program $n^{\circ}2$ which is logical since the variation interval is smaller than program n°1 or n°3 even in multi-CBR configuration. The non perfectly flat property of these curves is also expected since our method enables variations to optimize coding efficiency, and thus proposes a good trade-off between these two aspects. This confirms the smoothness of the proposed solution for mobiles services even if the σ_{Mux} is slightly higher in the Multi-CBR solution (Tables 6.3 6.4).

Fig. 6.7 Illustration of the multiplexes generated by the proposed SMX solution

The multiplexes dedicated to DTT and Mobile channels are provided, respectively in Figure 6.7a and 6.7b. We can observe that the available bandwidths are well respected in both channels with several slight overheads in the DTT channel. Finally, it should be noted that

the substantial gain achieved by the proposed SMX solution with three programs can be further improved in real SMX pools considering more than three programs (i.e. the gain will increase with the number of programs).

6.3 Conclusion

In this contribution, we propose a simple SMX solution adapted to databases which exploits the impact of the bitrate ratio between layers to perform a joint allocation on both a BL and EL to fit into specific DVB-T2 channels. The proposed approach has been evaluated on an industrial use-case of hybrid mobile/fixed delivery scheme where it enables in average an overhead reduction of 3.36%, with an enhanced smoothing in EL services and tolerable degradation in the BL ones. The proposed SMX enables an efficient deployment of hybrid and scalable UHD services with limited overhead.

The next step of this work should be dedicated to on-the-fly implementation of this method. First, a synchronized SHVC encoder implementation should be achieved. Second, a global control unit based on the proposed algorithm will be implemented to control optimal ratio selection considering channel bandwidths and quality variations smoothing. Finally, other algorithms for bitrate allocation in SMX could be investigated within this on-the-fly framework. For instance, a game-theory based approach could be investigated as a first track.

References

- [1] Q. Wu, "Statistical Multiplexing for MPEG-2 Video Streams," *Master's Thesis of Massachusetts Institute of Technology (MIT)*, 1999.
- [2] A. F. Antone and R. Arsinte, "A Study on the Optimal Implementation of Statistical Multiplexing in DVB Distribution Systems," *Informatics and IT Today*, vol. 1, pp. 19– 27, 2013.
- [3] UDCasts, "Open Statistical Multiplexing Architecture for Mobile TV," *White Paper*, 2007.
- [4] A. Gutha and D.-J. Reininger, "Multichannel Joint Rate Control of VBR MPEG Encoded Video for DBS Applications," *IEEE Transactions on Consumer Electronics*, vol. 40, no. 3, pp. 616–623, 1994.
- [5] L. Wang and A. Vincent, "Joint Rate Control for Multi-Program Video Coding," *IEEE Transactions on Consumer Electronics*, vol. 42, no. 3, pp. 300–305, 1996.
- [6] L. Boroczky, A.-Y. Ngai, and E.-F. Westermann, "Joint Rate Control with Look-Ahead for Multi-Program Video Coding," *IEEE TCSVT*, vol. 10, no. 7, pp. 1159–1163, 2000.
- [7] N. Changuel, B. Sayadi, and M. Kieffer, "Predictive Encoder and Buffer Control for Statistical Multiplexing of Multimedia Contents," *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 401–416, 2012.
- [8] C. Pang, O.-C. Au, J. Dai, and F. You, "Dependent Joint Bit Allocation for H.264/AVC Statistical Multiplexing Using Convex Relaxation," *IEEE TCSVT*, vol. 23, no. 8, pp. 1334–1345, 2013.
- [9] M. Jacobs, S. Tondeur, R. Van de Walle, T. Paridaens, and P. Schelkens, "Statistical Multiplexing using SVC," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2008.
- [10] Y. Wang, L.-P. Chau, and K.-H. Yap, "Joint Rate Allocation for Multiprogram Video Coding Using SVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 6, pp. 829–837, 2010.
- [11] W. Yao, L.-P. Chau, and S. Rahardja, "Joint Rate Allocation for Statistical Multiplexing in Video Broadcast Applications," *IEEE Transactions on Broadcasting*, vol. 58, no. 3, pp. 417–427, 2012.
- [12] J. Jeong, Y.-H. Jung, and Y. Choe, "Statistical Multiplexing using Scalable Video Coding for Layered Multicast," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, 2009.
- [13] G. Vand Der Auwera and M. Reisslen, "Implications of Smoothing on Statistical Multiplexing of H.264/AVC and SVC Video Streams," *IEEE Transactions on Broadcasting*, vol. 55, no. 3, pp. 541–558, 2009.
- [14] EBU, "https://tech.ebu.ch/testsequences/uhd-1."
- [15] J. Chen, J. Boyce, Y. Ye, and M.-M. Hannuksela, "JCTVC-T1007: SHVC Test Model 9 (SHM 9) Introduction and Encoder Description," Feb. 2015.
- [16] S.-H. Bae, T.-N. Pappas, and B.-H. Juang, "Subjective Evaluation of Spatial Resolution and Quantization Noise Tradeoffs," *IEEE Transactions on Image Processing*, vol. 18, no. 3, pp. 495–508, 2009.

Part IV

Conclusion

Chapter 7

Conclusion

In this thesis, we have explored several directions related to rate-control applications for HEVC and SHVC. First, we have developed a ρ -domain based QP estimation scheme as well as a ρ -domain analytical formulation suitable for rate-control applications. These methods can be joint to existing bitrate-allocation schemes. Second, the impact of the bitrate ratio between layers on SHVC performance has been studied. This statistical study was used as a basis for adaptive and optimal rate control algorithm. Finally, we have investigated a first and simple approach for statistical multiplexing of SHVC programs exploiting the impact of bitrate ratio in scalable encoders. In the next sections, we will first remind the objectives of the thesis. Then, the achieved works will be evaluated compared to the initial targets and discussed. Finally, several tracks for future work in this field will be proposed and discussed.

7.1 Initial targets

At the beginning, the scope of the thesis was to develop and study tools that enable the introduction of UHDTV services in broadcasting industry by using SHVC. Several tracks were possible such as AVC to HEVC transcoding, SHVC decoding or either rate-control aspect for SHVC. At that time, several works have already been achieved in AVC-to-HEVC transcoding and the decoding field was also well investigated since a real-time SHVC decoder was almost finished at the IETR Lab. Thus, we chose to focus our work on rate-control tools, and especially three tracks. First, we wanted to investigate ρ -domain rate-control schemes for SHVC, especially for their low-complexity benefits. Second, the impact of bitrate ratio between layers on SHVC coding performance should be investigated to see if an optimal or worst points exist, and potentially exploit that aspect. Finally, and in the perspective of scalable services introduction, the statistical multiplexing schemes had to be explored for

SHVC. In the following Section, the achieved work is compared to these initial targets and the encountered difficulties are discussed.

7.2 Achieved work

As mentioned in Chapter 4, the first step of the ρ -domain investigations was to perform a deep statistical study. First we have investigated the natural bitrate distribution in both HEVC and SHVC bitstreams and validated the ρ -domain modeling for these two standards for global and residual-only bitrate. To enhance the existing solutions, we have then explored an accurate method for encoding parameter estimation in HEVC and SHVC. Our method avoids the use of Look-up-tables and enables an analytic formulation of $\rho - OP$ relationship. This method was experimented and showed a very good accuracy for HEVC and SHVC RA encoding, with an average error below 4%. We have then observed that the ρ -domain curves seem pretty similar for sequences at the same spatial resolution, and that the slope can potentially be predicted in a deterministic way. We have investigated this track and thus found an accurate equation which consider the skip rate and thus which could be integrated in existing schemes to enhance the ρ -domain slope estimation. Experimental results show an accuracy above 90% for the deterministic formulation. These investigations were published and presented in an international conference [1] and protected by a granted patent [2]. However, the low complexity benefit of this method was hard to emphasize since the $R-\lambda$ approach is single-pass and particularly adapted to encoders that are based on RDO loops. Hence, we switched to the alternative way which was promising and of significant interest for deployment of scalable services.

This second part of the thesis is described in Chapter 5 and also begins with a deep statistical study. Indeed, we have investigated the impact of the bitrate ratio between layers in SHVC for different configurations to have a clear idea on what is the best way to evaluate scalability. This extensive study enabled to evaluate the potential gains that can be reached and was published in [3]. In the same time, we realised that the current way of estimating SHVC in standardization bodies such as MPEG was biased for some sequences and that the announced performance did not reflect the whole potential of SHVC. These observations were shared and discussed in MPEG [4] for spatial, color-gamut and SDR-to-HDR scalability. Based on these observations, we began the development of adaptive rate control schemes for SHVC which adjust the bitrate ratio while encoding under a global bitrate constraint. The first approach, published in [5], considered an optimization into a pre-defined ratio interval and enabled a substantial performance improvement of 4.25%. To improve this method, we have

added bitrate and quality instructions to be more realistic, which also enabled substantial overhead reduction compared to the SL HEVC encoding published in [6]. In addition, this enhanced method can be adapted to any two-layers scheme, whatever the type of scalability, since it enables a quality maximization in the EL. This second axe showed a strong interest for deployment of UHDTV services using SHVC and the integral work is protected by a patent application [7] and is currently submitted for review in the IEEE Transactions on Broadcasting journal.

Finally, the third part of the thesis described in Chapter 6 is dedicated to the statistical multiplexing of scalable programs using SHVC. In this work, the assessments of the Chapter 5 on the impact of bitrate ratio are used as a basis for the proposed solution. In the considered use-case, several programs are transmitted via two channels. The first one is dedicated to mobile reception and contains multiplexed BL streams. The second one targets the fixed receivers and contains the EL streams. Our method first allocates the global bitrate for each program based on program complexity and then select the optimal ratio combination which optimizes the coding performance. At the same time, this scheme aims at minimizing the quality variations among programs and has to respect bandwidth restrictions of each channel. The proposed algorithm has been tested in the case of a database containing pre-encoded bitstreams of different global bitrate and ratios, used as a basis to create our multiplexed streams. Our method builds the optimal bitstream by assembling the database available piece of bitstream with respect to bandwidths and quality restrictions. The proposed method enables to significantly reduce the cost of scalable services introduction, with an overhead reduced from 11.01% to 7.65%, while maintaining a good accuracy and an acceptable quality variations among programs. This work has been submitted in an international conference and is currently under review for publication.

To summarize, we have developed three contributions. The first one is dedicated to ρ domain based tools for rate-control applications which shows interesting results and let some space to further investigations described in the next Section. The second contribution is dedicated to adaptive rate control schemes for SHVC and is much more complete, deeply investigated and applied to many use-cases. The last one is dedicated to statistical multiplexing and enables to significantly limit the cost of hybrid delivery of SHVC programs. In addition to the publications and patent applications, the work on this thesis included the participation in several standardization bodies such as DVB or MPEG which enabled us to select relevant and realistic use-cases for our experiments and to emphasize its industrial interest. All these publications, patents and standardization contributions are detailed in Appendix C.

7.3 Prospectives and future works

7.3.1 ρ-domain based rate-control

As mentioned previously, the solutions developed around the ρ -domain RC can further be enhanced in several ways. First, the two proposed tools can be merged and the deterministic formulation could be integrated in the QP estimation scheme. Secondly, inter-layer tools in SHVC coding scheme could be investigated for the QP estimation to possibly predict the encoding parameters from the underlying layer information. Eventually, the existing R−λ approach could also be enhanced by considering inter-layer tools. Indeed, the current implementation only duplicates the HEVC algorithms including the RC algorithm and thus could be improved by exploiting inter-layer dependencies (SHVC).

7.3.2 Adaptive rate control for SHVC

For the adaptive rate control scheme, several tracks remain for improvement. A more fine granularity could be investigated, for instance at frame level, this might enables better R-D performance and more accurate rate-control. In addition, our method has been tested on HDR content and showed good performance for several objective metrics. All these metrics have been tested in MPEG and any of these is clearly optimal for HDR quality evaluation. An additional performance evaluation using HDR-VQM [8] –which is higly correlated to the subjective tests'– is proposed to assess the objective performance of our method. Finally, subjective tests would be relevant to complete the evaluation of our method. In addition, other types of scalability could be investigated such as codec scalability (AVC-BL and HEVC-EL) and even fidelity scalability.

7.3.3 Statistical multiplexing of scalable programs

The next step of this work should be dedicated to enhance the statistical multiplexing scheme introduced in Chapter 6. We have proved that using statistical multiplexing can provide significant gains, but obviously there is room for improvement. Our proposed method relies on database of pre-encoded bitstream, thus an on-the-fly method should be investigated as a next step, including a look-ahead processing to efficiently compute complexity indicators such as MSE or GPP. In addition, other use-cases and more complex approach could also be explored, for instance with a game-theory based approach.

References

- [1] T. Biatek, M. Raulet, J.-F. Travers, and O. Deforges, "Efficient Quantization Parameter Estimation in HEVC based on ρ-Domain," *European Signal Processing Conference (EUSIPCO)*, 2014.
- [2] T. Biatek, M. Raulet, and O. Deforges, "Method of Estimating a Coding Bitrate of an Image of a Sequence of Images, Method of Coding, Device and Computer Program Corresponding Thereto," in *European Patent WO/2015/090682*, 2015.
- [3] T. Biatek, W. Hamidouche, J.-F. Travers, and O. Deforges, "Toward Optimal Bitrate Allocation in the Scalable HEVC Extension: Application to UHDTV," *IEEE International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, 2015.
- [4] T. Biatek, W. Hamidouche, J.-F. Travers, and O. Deforges, "ISO/IEC JCT1/SC29/WG11 m36967: On the Performance Evaluation of Scalable Technology in Future Video Coding," Oct. 2015.
- [5] T. Biatek, W. Hamidouche, J.-F. Travers, and O. Deforges, "Adaptive Rate Control Algorithm for SHVC: Application to HD/UHD," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016.
- [6] T. Biatek, W. Hamidouche, J.-F. Travers, and O. Deforges, "Optimal Bitrate Allocation for High Dynamic Range and Wide Color Gamut Services Deployment using SHVC," *Data Compression Conference (DCC)*, 2016.
- [7] T. Biatek, "Procédé d'Allocation de Débit, Dispositif, Codeur et Programme d'Ordinateur Associés," in *French Patent Application FR1559029*, 2015.
- [8] M. Narwaria, P. da Silva, and P. Le Callet, "HDR-VQM: An Objective Quality Measure for High Dynamic Range Video," *Signal Processing: Image Communication*, no. 35, pp. 46–60, 2015.
Part V

Appendix

Appendix A

Statistical studies: exhaustive results

A.1 ρ -domain and bitrate distribution for HEVC and SHVC

A.1.1 Dataset

As mentioned in Section 4, we provide in this Appendix additional results for the statistical study on both the ρ -domain and the bitrate distribution in HEVC and SHVC random-access bitstreams. For this study, hundreds of Figure can be drawn for the whole datasets described in Tables 4.1 and 4.2. This, we choose to limite the amount of curves provided in this Appendix to one sequence per class for HEVC. For SHVC, we provide the results for three sequence of the dataset at 24, 50 and 60 frames per second. All these curves are representative of the whole considered dataset.

In Figure A.1, A.2, A.3 and A.4, the bitrate distribution for HEVC RA configuration is provided respectively for the sequences *Traffic*, *BasketballDrillText*, *Keiba* and *KristenAndSara*. For SHVC, the Figures A.5 and A.6 illustrates the bitrate distribution for RA configuration in 1.5x spatial scalability for sequences *BQTerrace* and *ParkScene*. These results have been built according to the procedure described in Section 4.2.2.

In Figure A.7, A.8, A.9 and A.10, the ρ -domain study for HEVC RA configuration is provided respectively for the sequences *Traffic*, *BasketballDrillText*, *Keiba* and *KristenAndSara*. For SHVC, the Figures A.11 and A.12 illustrates the ρ -domain study for RA configuration in 1.5x spatial scalability EL for sequences *BQTerrace* and *ParkScene*. These results have been built according to the procedure described in Section 4.2.3.

$\mathbf 0$ $Ivl.0$ $Ivl. 1$ $Ivl.2$ $Ivl.3$ (a) QP0 = 22 Distribution [%] 100 50 $\bf{0}$ $Ivl.0$ $Ivl. 1$ $Ivl.2$ $Ivl.3$ (b) $QP0 = 27$ Distribution [%] 100 50 $\mathbf{0}$ $Ivl.$ 0 $Ivl. 1$ $Ivl.2$ $Ivl.3$ (c) QP0= 32 Distribution [%] 100 50 $\mathbf 0$ $\overline{IVI.0}$ I vl. 1 $\overline{M.2}$ $Ivl.3$ (d) $QP0 = 37$ █CU Syntax
█Prediction Syntax
█Motion Vectors NALU Header Transform Syntax Slice Header Residual **Other** (e) Legend

A.1.2 Bitrate distribution in HEVC and SHVC RA configuration

Fig. A.1 Average bitrate distribution in HEVC RA for *Traffic* (Class A)

Fig. A.2 Average bitrate distribution in HEVC RA for *BasketballDrillText* (Class C)

Fig. A.3 Average bitrate distribution in HEVC RA for *Keiba* (Class D)

Fig. A.4 Average bitrate distribution in HEVC RA for *KristenAndSara* (Class E)

Fig. A.5 Average bitrate distribution in SHVC EL in RA for *BQTerrace* 1.5x

Fig. A.6 Average bitrate distribution in SHVC EL in RA for *ParkScene* 1.5x

A.1.3 ρ -domain Curves for HEVC

Fig. A.7 Illustration of a ρ-domain modeling in HEVC for *Traffic* (Class A)

Fig. A.8 Illustration of a ρ-domain modeling in HEVC for *BasketballDrillText* (Class C)

Fig. A.9 Illustration of a ρ-domain modeling in HEVC for *Keiba* (Class D)

Fig. A.10 Illustration of a ρ-domain modeling in HEVC for *KristenAndSarah* (Class E)

A.1.4 ρ -domain Curves for SHVC EL

Fig. A.11 Illustration of a ρ-domain modeling in SHVC EL for *BQTerrace* 1.5x

Fig. A.12 Illustration of a ρ-domain modeling in SHVC EL for *ParkScene* 1.5x

A.2 Impact of bitrate ratio over performance

(b) 1.5x spatial scalability

Fig. A.13 Impact of bitrate ratio for Class B dataset

Fig. A.14 Impact of bitrate ratio for EBU UHD-1 dataset – 2x Spatial Scalability

Fig. A.15 Impact of bitrate ratio for Class A dataset – 2x Spatial Scalability

Fig. A.16 Impact of bitrate ratio for Technicolor dataset $-2x$ Spatial Scalability & CGS

Fig. A.17 Impact of bitrate ratio for HDR Call for Evidence dataset – CGS & SDR-to-HDR

Fig. A.18 Impact of bitrate ratio for Volvo Ocean Race dataset – CGS & SDR-to-HDR

Appendix B

Adaptive rate control: exhaustive results

B.1 Fixed-interval adaptive rate control

In this Section, the performance of the algorithm described in Section 5.4 is provided for each sequence in the EBU UHD-1 dataset. The Single-Layer encoding is represented as well as the fixed ratio encoding and our method. The R-D curves with EL YUV-PSNR in Y-Axis are provided in Figure B.1 with the bitrate in X-Axis logarithmic scale.

B.2 Constrained adaptive rate control

In this Section, the performance of the algorithm described in Section 5.5 is provided for the two use-cases considered in Section 5.5.3. The Single-Layer encoding is represented as well as the fixed ratio encoding and our method. The R-D curves with EL YUV-PSNR and bitrate in X-Axis logarithmic scale are provided in Figures B.1 and B.2/B.3 for use-cases 1 and 2, respectively.

Fig. B.1 R-D curves of the EBU UHD-1 dataset encodings for evaluation of fixed-interval ARC proposed algorithm

Fig. B.2 R-D curves of the fixed ratio, single layer, and ARC encoding for use-case 1 (HD-BT.709 to UHD-BT.709)

Fig. B.3 R-D curves of the fixed ratio, single layer, and ARC encoding for use-case 1 (HD-BT.709 to UHD-BT.2020)

Fig. B.4 R-D curves of the fixed ratio, single layer, and ARC encoding for use-case 2

Appendix C

Publications and patents

C.1 Publications

[C1] T. Biatek, M. Raulet, J.-F. Travers and O. Deforges, "Efficient Quantization Parameter Estimation in HEVC based on ρ -Domain", Proceedings of the 22nd European Signal Processing Conference (EUSIPCO), pp. 296–300, Sept. 2014.

Abstract – This paper proposes a quantization parameter estimation algorithm for HEVC CTU rate control. Several methods were proposed, mostly based on Lagrangian optimization combined with Laplacian distribution for transformed coefficients. These methods are accurate but increase the encoder complexity. This paper provides an innovative reduced complexity algorithm based on a ρ -domain rate model. Indeed, for each CTU, the algorithm predicts encoding parameters based on co-located CTU. By combining it with Laplacian distribution for transformed coefficients, we obtain the dead-zone boundary for quantization and the related quantization parameter. Experiments in the HEVC HM Reference Software show a good accuracy with only a 3% average bitrate error and no PSNR deterioration for random-access configuration.

[C2] T. Biatek, W. Hamidouche, J.-F. Travers and O. Deforges, "Toward Optimal Bitrate Allocation in the Scalable HEVC Extension: Application to UHDTV", IEEE International Conference on Consumer Electronics Berlin (ICCE-Berlin), Sept. 2015.

Abstract – Scalable video encoders compress a single video sequence to produce a bitstream composed of several layers, corresponding to different temporal, spatial and quality representations of the input video sequence. This technique improves the coding efficiency compared to simulcast encoding of each representation by exploiting additional correlations through inter-layer predictions. The latest scalable video coding standard SHVC –the extension of the recent HEVC standard– announces up to

30% bandwidth reduction. However, this gain is valid under the common test conditions, established by the JCT-VC expert group, which are not necessarily relevant in broadcasting environment and do not include video sequences in UHD resolution. In this paper, we expose the results of an extended study about the optimum gains that scalability can bring in concrete broadcast use-cases. Indeed, we consider SHVC with HDTV/UHDTV as spatial enhancement layers. Then, we search the optimum balance of layers' dedicated bitratres in sense of coding efficiency and objective quality for different UHDTV video sequences and use-cases.

[C3] T. Biatek, W. Hamidouche, J.-F. Travers and O. Deforges, "Adaptive Rate Control Algorithm for SHVC: Application to HD/UHD", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar. 2016.

Abstract – Scalable video coding consists in compressing the video sequence into a layered bitstream where each layer refers to different spatial, temporal or quality representation of the video. Scalability enables compression gain compared to the simulcast encoding of layers thanks to inter-layer predictions. The scalable HEVC extension (SHVC) is the latest scalable technology promising up to 30% bitrate gains under the common test conditions, defined by JCT-VC. These conditions do not consider UHD and use fixed quantization step, which is not relevant in operational environment. In this paper, we propose an innovative adaptive rate control algorithm for SHVC. We consider HD as a base layer and UHD as an enhancement layer, with a constant global bitrate and a dynamic bitrate ratio adjustment between layers. The proposed algorithm is evaluated on a UHD data set where enables on average a BD-BR gain of 4.25% compared to a fixed-ratio encoding.

[C4] T. Biatek, W. Hamidouche, J.-F. Travers and O. Deforges, "Optimal Bitrate Allocation for High Dynamic Range and Wide Color Gamut Services Deployment using SHVC", Data Compression Conference (DCC), Mar. 2016.

Abstract – The scalable video coding enables to compress video contents into a hierarchical layered representation, each layer depicts an enhanced version of the underlying layer. SHVC is the scalable extension of HEVC and enables spatial, SNR, color-gamut, codec and bitdepth scalability. It has been proved, in the MPEG investigations prior to the recent Call for Evidence, that SHVC can support SDR-to-HDR scalability by using the color gamut scalability, when SDR and HDR signals are placed in different color gamuts. This way, SHVC can be used to address future backward compatible issues in the HDR and WCG services deployment. In this paper, we exploit the impact of bitrate ratio over performance in scalable schemes to design an adaptive rate control algorithm suitable for such deployment, considering adjustable quality and bandwidth constraints. Our method dynamically adjusts the bitrate ratio between two layers during encoding in the most quality-related optimal way under specified constraints. The proposed method is tested on scalable combinations of HD/UHD, R.709/DCI-P3/R.2020 and SDR/HDR video contents, and reduces the average overhead introduced by SHVC compared to the single-layer HEVC encoding by 23%.

[J1] T. Biatek, W. Hamidouche, J.-F. Travers and O. Deforges, Optimal Bitrate Allocation in the Scalable HEVC Extension for the Deployment of UHD Services, IEEE Transactions on Broadcasting, Vol. PP, Issue: 99 (To be published).

Abstract – Ultra High Definition (UHD) is the latest trend in broadcasting area, which enables new services with 3840x2160 resolution and comes with enhanced colorgamut, frame-rate, dynamic range and better audio system compared to the currently deployed HD services. The UHD format for broadcasting is already under standardization in the DVB consortium which plans to introduce UHD services in three phases. The increase in data brought by these services requires more efficient compression and transmission systems. The recent scalable video coding standard SHVC is a promising candidate to handle these three phases while ensuring backward compatibility. Moreover, delivering such contents over networks needs an accurate control of the output bitrate from encoder engines to match rigid constraints on bandwidth and QoS. Several contributions have already been proposed to jointly encode scalable stream, but without considering the impact of bitrate ratio between layers on the compression performance. In this paper, we first investigate the impact of the bitrate ratio between layers on the coding performance for several UHD scalable schemes including spatial, color-gamut and SDR-to-HDR scalability in SHVC. Based on this investigation, we propose an adaptive rate control algorithm which dynamically allocates the bitrate between two layers to optimize the performance under quality and bitrate constraints. The algorithm has been implemented in the latest SHVC reference software (SHM9.0) and tested on 15 video sequences within two industrial use-cases. The performance shows an average BD-BR improvement of 7.51% and 3.35% for these use-cases.

[C5] T. Biatek, W. Hamidouche, J.-F. Travers and O. Deforges, Pre-Encoding based Statistical-Multiplexing for Hybrid Delivery of UHD Services using SHVC, Picture Coding Symposium, Dec. 2016.

Abstract – The scalable video coding consists in encoding the video content into multiple representations, called layers, where each one refers to a particular version of the content. The scalable extension of the High Efficiency Video Coding standard SHVC is currently considered in ATSC and DVB to carry out layered and scalable programs which enables to target multiple equipments, and is also used to ensure backward compatibility with legacy receivers. In the case of hybrid delivery of HD and UHD services using SHVC, these services are encoded in two layers including base and enhancement layers, which are then broadcast over separated channels. In this paper, a statistical multiplexing method is proposed for broadcasting of UHD services in this hybrid scenario. This innovative method considers both variable bitrate among programs and optimal SHVC layers coding, which was not considered in the existing approaches. The proposed method enables to reduce the overhead introduced by SHVC compared to the single-layer encoding by 3.3% in average while maintaining smooth quality variations among programs.

C.2 MPEG/JCT-VC/JVET standardization contributions

[D1] T. Biatek, W. Hamidouche, J.-F. Travers and O. Deforges, "ISO/IEC JCT1/SC29/WG11 m3697: On the Performance Evaluation of Scalable Technology in Future Video Coding", Oct. 2015.

Abstract – Many interrogations and doubts have been recently raised in the Future-VideoCoding reflector concerning scalability. Facts and figures have been mentioned concerning the codec performance that a scalable scheme such as SHVC can achieve. In addition to this, the test conditions have been discussed and in particular the aspect of bitrates. In this contribution, we have investigated the impact of the bitrate ratio between layers on the performance. We present the results of an advanced study on SHVC performance. The results show that a clever ratio adjustment may lead to different and better performance. These results enable to reconsider the purpose and choices made in the current common test conditions (CTC) and may help in designing the future test conditions and performance evaluation methods, especially for scalable technologies.

[D2] T. Biatek and X. Ducloux , "ISO/IEC JCT1/SC29/WG11-m37051: HEVC Encoding Results on Future Video Coding Test Sequences", Oct. 2015.

Abstract – This document provides the encoding results of Netflix test sequences and some additional observations and analysis about the content and noticeable visual artifacts. These tests have been achieved within the research institute IRT b<>com on five 10sec sequences provided by Netflix according to the test sequence encoding work plan.

[D3] X. Ducloux, T. Biatek and J.-Y. Aubié, "ISO/IEC JCT1/SC29/WG11-m37052: Proposal of New UltraHD Test Material", Oct. 2015.

Abstract – A Call for Test Materials for Future Video Coding Standardization was issued at the 112th MPEG meeting in Warsaw. The Institute of Research & Technology b<>com is willing to provide some of its UltraHD video contents to MPEG for the purpose of development, testing and promulgation of video coding standards. This document gives a description of these video contents and compression performance with HEVC.

[D4] F. Henry, X. Ducloux T. Biatek and J.-Y. Aubié, "JCTVC-V0086: B-com Test Sequences for Video Coding Development", Oct. 2015.

Abstract – The Institute of Research & Technology b<>com is willing to provide five of its UltraHD video contents for the purpose of development, testing and promulgation of video coding standards. This document gives a description of these video contents and compression performance with HEVC.

[D5] T. Biatek and X. Ducloux, "JVET-B0024: Evaluation Report of SJTU Test Sequences", Feb. 2016.

Abstract – This document provides the evaluation results of five Ultra HD test sequences provided by SJTU. Both objective and subjective evaluation were performed. For objective evaluation, the HM-16.6 reference software was used as defined in N15791. All the obtained results of objective evaluation can be found in the attached Excel table. In addition, some examples of subjective distortions are presented. Encoded streams as well as decoded YUV sequences are available for evaluation or cross-check.

C.3 Patents

[P1] T. Biatek, M. Raulet, and O. Deforges, "Method of Estimating a Coding Bitrate of an Image of a Sequence of Images, Method of Coding, Device and Computer Program Corresponding Thereto", France, WO2015090682 (A1).

[P2] T. Biatek, "Procédé d'Allocation de Débit, Dispositif, Codeur et Programme d'Ordinateur Associés", France, Demande FR-1559029

AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE

Titre de la thèse:

Stratégies d'encodage pour codeur vidéo scalable

Nom Prénom de l'auteur : BIATEK THIBAUD

Membres du jury :

- Monsieur TRAVERS Jean-François
- Madame GUILLEMOT Christine - Monsieur DUFAUX Frédéric
- Monsieur WIEN Mathias
-
- Monsieur VIER Mannus
- Monsieur Cagnazzo Marco
- Monsieur DEFORGES Olivier
- Monsieur HAMIDOUCHE Wassim
- Monsieur LE CALLET Patrick

Président du jury :

C. guillemot

Date de la soutenance : 25 Avril 2016

Reproduction de la these soutenue

Thèse pouvant être reproduite en l'état Thèse pouvant être reproduite après corrections suggérées

Fait à Rennes, le 25 Avril 2016

Signature du président de jury

Unhifar Harl

Le Directeur, M'hamed DRISSI $\sqrt{\text{SW}}$

Résumé

High Efficiency Video Coding (HEVC/H.265) est la dernière norme de compression vidéo, finalisée en Janvier 2013. Son extension scalable, SHVC, a été publiée en Octobre 2014 et supporte la scalabilité spatiale, en gamut de couleur (CGS) et même en norme de compression (AVC vers HEVC). SHVC peut être utilisée pour l'introduction de nouveaux services, notamment grâce à la rétrocompatibilité qu'elle apporte par la couche de base (BL) et qui est complétée par une couche d'amélioration (BL+EL) qui apporte les nouveaux services. De plus, SHVC apporte des gains en débit significatifs par rapport à l'encodage dit simulcast (l'encodage HEVC séparés). SHVC est considérée par DVB pour accompagner l'introduction de services UHD et est déjà incluse dans la norme ATSC-3.0. Dans ce contexte, l'objectif de la thèse est la conception de stratégies de régulation de débit pour les codeurs HEVC/SHVC lors de l'introduction de nouveaux services UHD.

Premièrement, nous avons étudié l'approche ρ-domaine qui modélise linéairement le nombre coefficient non-nuls dans les résidus transformés et quantifiés avec le débit, et qui permet de réaliser des régulations de débit peu complexes. Après validation du modèle, nous avons conçu un premier algorithme de contrôle de débit au niveau bloc en utilisant cette approche. Pour chaque bloc et son débit cible associé, notre méthode estime de façon précise le paramètre de quantification (QP) optimal à partir des blocs voisins, en limitant l'erreur de débit sous les 4%. Puis, nous avons proposé un modèle d'estimation déterministe du ρ-domaine qui évite l'utilisation de tables de correspondance et atteignant une précision d'estimation supérieure à 90%.

Deuxièmement, nous avons investigué l'impact du ratio de débit entre les couches d'un codeur SHVC sur ses performances de compression, pour la scalabilité spatiale, CGS et SDR vers HDR. En se basant sur les résultats de cette étude, nous avons élaborés un algorithme de régulation de débit adaptatif. La première approche proposée optimise les gains de codage en choisissant dynamiquement le ratio de débit optimal dans un intervalle prédéterminé et fixe lors de l'encodage. Cette première méthode a montré un gain de codage significatif de 4.25% par rapport à une approche à ratio fixe. Cette méthode a été ensuite améliorée en lui ajoutant des contraintes de qualité et de débit sur chaque couche, au lieu de considérer un intervalle fixe. Ce second algorithme a été testé sur le cas de diffusion de programme HD/UHD et de déploiement de services UHD1-P1 vers UHD1-P2 (cas d'usage DVB), où elle permet des gains de 7.51% et 8.30% respectivement.

Enfin, le multiplexage statistique de programmes scalable a été introduit et brièvement investigué. Nous avons proposé une première approche qui ajuste conjointement le débit global attribué à chaque programme ainsi que le ratio de débit, de façon à optimiser les performances de codage. De plus, la méthode proposée lisse les variations et l'homogénéité de la qualité parmi les programmes. Cette méthode a été appliquée à une base de données contenant des flux pré-encodés. La méthode permet dans ce cas une réduction du surcoût de la scalabilité de 11.01% à 7.65% comparé à l'encodage a débit et ratio fixe, tout en apportant une excellente précision et une variation de qualité limitée.

Abstract

High Efficiency Video Coding (HEVC/H.265) is the latest video coding standard, finalized in January 2013 as the successor of Advanced Video Coding (AVC/H.264). Its scalable extension, called SHVC was released in October 2014 and enables spatial, bitdepth, color-gamut (CGS) and even standard scalability. SHVC is a good candidate for introducing new services thanks to backward compatibility features with legacy HEVC receivers through the base-layer (BL) stream and next generation ones through the BL+EL (enhancement layer). In addition, SHVC saves substantial bitrate with respect to simulcast coding (independent coding of layers) and is also considered by DVB for UHD introduction and included in ATSC-3.0. In this context, the work of this thesis aims at designing efficient rate-control strategies for HEVC and its scalable extension SHVC in the context of new UHD formats introduction.

First, we have investigated the ρ-domain approach which consists in linking the number of non-zero transformed and quantized residual coefficients with the bitrate, in a linear way, to achieve straightforward rate-control. After validating it in the context of HEVC and SHVC codings, we have developed an innovative Coding Tree Unit (CTU)-level rate-control algorithm using the ρ-domain. For each CTU and its associated targeted bitrate, our method accurately estimates the most appropriate quantization parameter (QP) based on neighborhood indicators, with a bitrate error below 4%. Then, we have proposed a deterministic way of estimating the ρ-domain model which avoids the implementation of look-up tables. The proposed method enables accurate model estimation over 90%.

Second, we have explored the impact of the bitrate ratio between layers on the SHVC performance for the spatial, CGS and SDR-to-HDR scalability. Based on statistical observations, we have built an adaptive rate control algorithms (ARC). We have first proposed an ARC scheme which optimizes coding performance by selecting the optimal ratio into a fixed ratio interval, under a global bitrate instruction (BL+EL). This method is adaptive and considers the content and the type of scalability. This first approach enables a coding gain of 4.25% compared to fixed-ratio encoding. Then, this method has been enhanced with quality and bandwidth constraints in each layer instead of considering a fixed interval. This second method has been tested on hybrid delivery of HD/UHD services and backward compatible SHVC encoding of UHD1-P1/UHD1-P2 services (DVB use-case) where it enables significant coding gains of 7.51% and 8.30%, respectively.

Finally, the statistical multiplexing of SHVC programs has been investigated. We have proposed a first approach which adjusts both the global bitrate to allocate in each program and the ratio between BL and EL to optimize the coding performance. In addition, the proposed method smooths the quality variations and enforces the quality homogeneity between programs. This method has been applied to a database containing pre-encoded bitstreams and enables an overhead reduction from 11.01% to 7.65% compared to constant bitrate encoding, while maintaining a good accuracy and an acceptable quality variations among programs.

N° d'ordre : **16ISAR 06 / D16 - 06 Institut National des Sciences Appliquées de Rennes** 20, Avenue des Buttes de Coësmes - CS 14315 - F-35043 Rennes Cedex Tél : 02 23 23 82 00 – Fax : 02 23 23 83 96