Region-of-interest-based video coding for video conference applications

Marwa Meddeb

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

EDITE – ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

Télécom ParisTech

Spécialité « Signal et Images »

présentée et soutenue publiquement par

Marwa MEDDEB

le 15 Février 2016

Codage vidéo par régions d'intérêt pour des applications de visioconférence

—

Region-of-interest-based video coding for video conference applications

Directeurs de thèse :
Béatrice Pesquet-Popescu (Télécom ParisTech)
Marco Cagnazzo (Télécom ParisTech)

Jury

M. Marc ANTONINI , Directeur de Recherche, Laboratoire I3S Sophia-Antipolis	Rapporteur
M. François Xavier COUDOUX , Professeur, IEMN DOAE Valenciennes	Rapporteur
M. Benoit MACQ , Professeur, Université Catholique de Louvain	Président
Mme. Béatrice PESQUET-POPESCU , Professeur, Télécom ParisTech	Directeur de thèse
M. Marco CAGNAZZO , Maître de Conférence HdR, Télécom ParisTech	Directeur de thèse
M. Joël JUNG , Ingénieur de Recherche, Orange Labs	Examinateur
M. Marius PREDA , Professeur Associé, Télécom SudParis	Examinateur

THÈSE

Télécom ParisTech
Grande École de l'Institut Télécom - membre fondateur de ParisTech
46 rue Barrault — 75634 Paris Cedex 13 — Tél. +33 (0)1 45 81 77 77 — www.telecom-paristech.fr

À mes parents Nabila et Jamel, à mes chers Chedly et Khaled, à tous ceux que ma réussite leur tient à cœur.

Remerciements

Je souhaite en premier lieu exprimer toute ma reconnaissance à ma directrice de thèse Béatrice Pesquet-Popescu, professeur à Télécom ParisTech de m'avoir permis d'intégrer l'équipe Multimédia et de m'avoir accordé sa confiance. Sa rigueur scientifique, sa clairvoyance et ses critiques pointilleuses m'ont beaucoup appris. Je remercie également Marco Cagnazzo, maître de conférence HdR à Télécom ParisTech pour la direction de mes travaux de thèse. Ses idées, ses conseils et ses encouragements tout au long de ces trois dernières années m'étaient d'une aide précieuse.

Je souhaite remercier Marc Antonini et François Xavier Coudoux d'avoir accepté de relire cette thèse et d'en être rapporteurs. Je tiens à remercier Benoît Macq d'avoir accepté d'être président du jury. Je remercie également Joel Yung et Marius Preda d'avoir accepté de lire mon manuscrit et d'assister à la présentation de ce travail en tant qu'examinateurs.

J'adresse toute ma gratitude à tous les permanents de Télécom ParisTech qui ont veillé au bon déroulement de ma thèse malgré toutes les difficultés que j'ai pu rencontrer. Je remercie Yves Grenier, Marie-Laure Chauveaux, Laurence Zelmar, Florence Besnard, Marianna Baziz et Fabrice Planche. Je remercie également tous les collègues du département TSI et spécialement Cyril Concolato et Jean Le Feuvre pour leur collaboration.

Je souhaite exprimer particulièrement toute mon amitié à Anielle Fiengo pour sa présence, Marco Calemme pour sa gentillesse, Giuseppe Valenzise pour ses conseils, Paul Lauga pour son humour, Elie Gabriel Mora pour son aide précieuse et Nassima Bouzakaria pour ses confidences. Je salue aussi tous les doctorants et postdoctorants que j'ai rencontrés pendant ces trois ans : Giovanni Trotta, Maxim Kapushin, Benoît Boyadjis, Emin Zerman, Aakanksha Rana, Cagri Ozcinar, Andrei Purica, Antoine Dricot, Edoardo Provenzi, Wei Fan, Shuo Zheng, Vedad Hulusic, Bogdan Ionut Cirstea, Francesca De Simone, Alper Koz, Yafei Xing, Pierrick Milhorat, Claudio Greco, Giovanni Chiercha, Giovanni Petrazzuoli, Marc Decombas, Cristina-Adriana Oprean, Hamlet Medina. Je les remercie tous pour les beaux moments d'échange et de partage.

Je souhaite exprimer ma gratitude à mon amie et ma confidente Magali Martin, adjointe relations européennes et internationales à INRIA pour son soutien permanent et ses encouragements incessants. Je tiens à remercier tous mes amis : Ahlem, Sirine, Meriem, Rim, Safa, Fenina, Myriam, Naziha, Chadi, Nissem, Ayoub, Sahar, Malek, Mehdi, et Chérif pour leur présence et pour toutes les sorties et soirées passées ensemble à oublier momentanément le travail et à se ressourcer. Je remercie les plus jeunes aussi : Marie, Alexandre et Emma pour avoir été très appliqués en cours de soutiens et pour la fraîcheur et la bonne humeur qu'ils m'ont procurées.

Finalement, je souhaite remercier ma famille en commençant par mes parents Nabila et Jamel qui m'ont appris à aimer la vie et faire de mon projet une passion et une fierté, à mon frère Chedly pour son regard admiratif et sa présence dans les moments les plus durs et à mon fiancé Khaled pour son amour, son soutien continu et sa confiance. Mes remerciements vont à tous les membres de ma famille et ma belle-famille spécialement Taoufik, Houda et Faker pour leur présence et leur soutien jusqu'au bout de cette thèse. Sans ma famille, sans son aide et son amour, cette thèse n'aurait jamais vu le jour. Je remercie donc du fond du coeur tous ceux qui ont cru en moi et m'ont conduit à ce jour mémorable.

Résumé

Les travaux effectués durant cette thèse de doctorat traitent le problème du codage vidéo basé sur les régions d'intérêt. Ils ont pour but d'améliorer l'efficacité de codage dans HEVC et de gagner en qualité de décodage globalement sur une séquence donnée mais aussi localement sur des régions d'un intérêt particulier. Par conséquent, nous proposons d'une part une modélisation précise du débit et de la distorsion, et d'autre part des méthodes de contrôle de débit basées sur les régions d'intérêt et adaptées au standard HEVC.

Dans la première partie de ces travaux, nous proposons de nouveaux modèles débit-distorsion pour HEVC. La modélisation proposée tient compte des caractéristiques du contenu et de l'encodeur. Dans une première approche, les modèles sont développés pour des blocs de type intra en tenant compte uniquement des dépendances spatiales entre les pixels du bloc. Dans une seconde approche, on utilise les caractéristiques statistiques des données à encoder pour obtenir des modèles plus efficaces pour le codage vidéo hybride (intra et inter). Nos expériences montrent qu'une bonne représentation des coefficients transformés des blocs d'une image donne une meilleure répartition du débit et un gain global importants.

Dans la deuxième partie de la thèse, nous proposons de nouveaux algorithmes de contrôle de débit pour HEVC qui introduisent le concept de la région d'intérêt. Assurer une allocation de bits par région et calculer le paramètre de quantification indépendamment sur des blocs de l'image de divers niveaux d'importance aident à améliorer la répartition du budget sur les différentes régions. Cela peut être utile dans de nombreuses applications où le traitement de l'image par régions d'intérêt est nécessaire, par exemple les systèmes de visioconférence. Les méthodes proposées montrent une amélioration de la qualité de la région d'intérêt tout en respectant la contrainte globale imposée par le réseau.

Mots clés : Codage vidéo, standard HEVC, région d'intérêt, tuile, contrôle de débit, modèle débit-distorsion.

Abstract

This PhD. thesis addresses the problem of region-of-interest-based video coding and deals with improving the coding efficiency in High Efficiency Video Coding standard. We propose both accurate rate distortion modeling approaches, and also region-of-interest-based rate control methods adapted for High Efficiency Video Coding.

In the first part, we propose new rate-distortion models for High Efficiency Video Coding at coding unit level. Proposed modeling takes into account content characteristics and encoder features. In a first proposition models are based on spatial dependencies between pixels of a coding unit while in a second proposition statistical characteristics of the data are used to derive more efficient models. We show the benefits that can be drawn from using content based rate-distortion modeling. A good fitting of transform coefficients per unit gives us important gains in coding efficiency.

In the second part, we propose novel rate control algorithms for High Efficiency Video Coding that introduces region-of-interest concept. Performing bit allocation per region and computing quantization parameter independently per units of various importance levels, help improving budget partitioning over regions of different interest. This can be useful in many applications where region-based processing of the frame is required such as videoconferencing systems. The proposed methods show an improvement in the quality of the region-of-interest while the budget constraint is respected.

Keywords: Video coding, High Efficiency Video Coding, region-of-interest, tiles, rate control, rate-distortion model.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AVC	Advanced Video Coding
BGG	Bernoulli Generalized Gaussian
CABAC	Context Adaptive Binary Arithmetic Coding
CBR	Constant Bit Rate
CTC	Common Test Conditions
CTU	Coding Tree Unit
CU	Coding Unit
dB	Decibel
DBF	De-Blocking Filter
DCT	Discrete-Cosine Transform
DFT	Discrete-Fourier Transform
DPCM	Differential Pulse Code Modulation
DST	Discrete-Sine Transform
GOP	Group Of Pictures
GPB	Generalized P- and B-picture
HEVC	High Efficiency Video Coding
HM	HEVC test Model
HVS	Human Visual System
IDR	Instantaneous Decoding Refresh
IEC	International Electrotechnical Commission
ISDN	Integrated Services Digital Network

ISO	International Standard Organization
ITU-T	International Telecommunication Union
JPEG	Joint Photographic Experts Group
JVT	Joint Video Team
KLT	Karhunen-Loeve Transform
LCU	Large Coding Unit
MAD	Mean Absolute Difference
MB	Macro-Block
MSE	Mean Square Error
MTU	Maximum Transmission Unit
NAL	Network Abstraction Layer
PDF	Probability Density Function
POC	Picture Order Count
PSNR	Peak Signal to Noise Ratio
PU	Prediction Unit
QP	Quantization Parameter
RD	Rate-Distortion
RDO	Rate-Distortion Optimization
ROI	Region Of Interest
SAO	Sample Adaptive Offset
SSIM	Structural Similarity Index
TM5	Test Model 5
TMN8	Test Model Near-term 8
TU	Transform Unit
UHD	Ultra High Definition
URQ	Unified Rate-Quantization

VBR	Variable Bit Rate
VCEG	Video Coding Expert Group
VCL	Video Coding Layer
VM8	Verification Model 8
VQA	Video Quality Assessment
WPP	Wavefront Parallel Processing

Introduction

Context

Video coding technologies defines different techniques that respond to audiovisual data transmission challenges. As digital pictures and videos are captured, processed and then transmitted over various communication channels of limited bandwidths and with different conditions, encoders are meant to compress the input data to fit network and storage requirements. On the other hand, the broadcasted multimedia data, finally presented to the human observer should have a good visual quality. In other words, the problem consists in the fact that signals with high information content are to be transmitted through low-capacity channels or stored into low-capacity media. In these conditions, high compression efficiency is required.

It is true that the evolution in network and in storage devices mitigates the needs of compression. But, it does not remove it. Indeed, the improvement in acquisition devices, in display, new formats and augmented needs of the users demand improved encoding algorithms. Standardization bodies such as the Moving Picture Experts Group (MPEG) under the authority of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) and the Video Coding Experts Group (VCEG) under International Telecommunication Union (IUT-T) have been developing video coding standards for a long time. Video coding technology has been re-standardized every several years, and each time coding performance has been improved.

The latest video compression standard, known as High Efficiency Video Coding (HEVC), is a successor to H.264/MPEG-4 AVC (Advanced Video Coding). It has been finalized in January 2013 and aims of substantially improving compression efficiency compared to H.264/AVC, reducing bitrate requirements by half with comparable image quality. Today, significant efforts are being devoted to adapt HEVC coding to specific needs of certain applications. Depending on the application requirements, HEVC aims to trade off computational complexity, compression rate, robustness to errors and processing delay time. Thus, non-normative tools in video coding standards are being developed. In particular, due to its importance, rate control is being researched.

Rate control schemes have been recommended by standard during the development to ensure a successful transmission of the coded stream in a limited bandwidth. Considering

the bit budget constraint, the controller performs, first, bit allocation per frame and/or per block. Then, using appropriate rate distortion models, optimal coding parameters such as quantization parameters are computed. Due to more complicated coding structure and the adoption of new coding tools, the statistical characteristics of transformed residues are significantly different. Thus, rate control techniques have evolved greatly with the development of video coding techniques. Different rate control methods have been implemented and tested over video encoders, some of them based on simple rate expressions such as in TM5 for MPEG-2, VM8 for MPEG4 and TMN8 for H.263 others on more complex mathematical representations such as in H.264/AVC and HEVC. The accuracy of these models has been enhanced by introducing the so-called complexity of the source and by considering advanced video coding features.

Furthermore, in various fields such as videoconferencing systems, video surveillance and telemedicine, the subjective visual quality mainly depends on some important areas, called regions-of-interest (ROIs). In recent years, a large number of research works have been done for ROI-based rate control. The main challenge of theses algorithms is to ensure a bit partitioning over regions that respects both the ROI and the budget constraints. Consequently, many contributions on H.264/AVC have introduced rate control algorithms that consider ROIs for bit allocation. However, when we started our work, all the existing RC algorithms developed for HEVC do not take into account the importance of particular regions. Therefore, the work done in this thesis falls in the ROI-based HEVC coding context.

The goal of our thesis is to develop new rate control tools on top of the HEVC reference software to further improve on the first hand the coding efficiency of the whole frame by proposing appropriate rate distortion models and on the second hand the quality of particular regions by performing ROI-based bit allocation. In collaboration with Telecom ParisTech, the young company AMIRIEL lunched a research program to build a high-definition videoconferencing solution designed for domestic use. The work in this thesis was performed in the Multimedia (MMA) group of the Signal and Image processing department (TSI) of Telecom ParisTech and the LTCI laboratory (UMR 5141).

Contributions

Various rate-distortion (RD) models aimed at performing semantic video coding in HEVC were developed during this thesis. They can be divided into two categories. The first category includes new rate-distortion models that takes into account statistics of the encoded sequence. We identify two contributions in this category:

- An RD model that accurately describes the relationship between the encoded bits and the corresponding distortion for intra-coded frames. The proposed model is adapted to independently decodable coding tree units (CTUs). It takes into account

spatial dependencies considering the characteristics of HEVC coded stream.

- Operational rate and distortion models considering signal characteristics at both high and low bit rates and for intra- and inter-coded frames. Model parameters are generated by fitting transform coefficient distribution to a Bernoulli Generalized Gaussian (BGG) model. The proposed models are used to minimize the RD cost per frame and compute the optimal distribution of the quantization parameter (QP) at CTU level.

The methods in the second category are ROI-based rate control schemes where appropriate bit allocation approaches and RD models are developed at region level. There are three contributions in this category:

- A method that adapts an existing rate control scheme for ROI-based bit allocation in HEVC inter-coded frames. Modifications in the quadratic region-based model implemented in H.264/AVC are introduced to adapt it to HEVC test model 9 (HM.9).
- A method using a ROI-based R-λ model. The two major steps of the rate control are modified: the bit allocation at both frame and CTU levels and the computation of QP by the proposed model for both I and B frames. First, bits are allocated per region. Then, independent R-λ models are derived for ROI and non-ROI to compute QPs for different units. The proposed approach was implemented in HM.10 and adapted to later version of HEVC test model 13 (HM.13).
- A method using a ROI- and tile-based R-λ model. Tiling is added to generate separate regions and ROI-based R-λ approach is used for bit allocation and QP computing. Independent rate allocation but also independent transmission and decoding of the ROI and the non-ROI are introduced to overcome the limitation of our ROI-based rate control algorithm.

Structure of the manuscript

This manuscript comprises three parts. The first part starts with a state-of-the-art in video coding, and a detailed overview of HEVC coding tools. Then, the second part describes the main features of rate control to end with a glossary of the proposed rate-distortion models. Finally, in the third part, ROI-based rate control concept is introduced and our contributions and experiments are detailed. More precisely, the manuscript is organized as follows:

Part one

- Chapter 1 presents a state-of-the-art in video coding. It highlights the Necessity and feasibility of data compression and describes fundamentals of image and video coding.

It ends with a list of useful evaluation metrics.

- Chapter 2 summarizes the evolution of video coding standards by presenting major features of each standard and justifying the usage of coding tools. A detailed presentation of the last in date compression standard, HEVC, concludes this chapter as it will serve as basis for comparison in manuscript.

Part two

- Chapter 3 presents the principles of rate control by introducing the rate distortion theory and explaining the concept of rate distortion optimization (RDO). Different rate-distortion models form the literature are described next, followed by important features of rate control in video coding.

- Chapter 4 details classic rate control algorithms by putting the stress on the evolution of their bit allocation processes and rate distortion models. It ends with a detailed description of HEVC rate control algorithms and their corresponding rate-distortion models (quadratic and R-λ models). A comparison of these two RD models concludes this chapter to motivate the choices we made during our researches.

- Chapter 5 presents our contributions related to rate-distortion modeling. It starts with a study of some models from the literature that we adapt to HEVC. Experiments are performed for model validation at CTU level to motivate the efficiency of the proposed model which takes into account spatial dependencies in an intra-coded frame. Then, we provide a study on RD modeling for HEVC considering appropriate probabilistic models for the transform coefficients. The method is then presented: basically parameters of the probability density function are estimated by maximizing the likelihood of the data under the model. The rate and distortion models are derived at both high and low bit rates. They are used to minimize the RD cost per frame and compute the optimal distribution of quantization parameters (QP) at CTU level. For both intra- and inter-coded frames, the method gives significant coding gains. The subjective and objective results are reported and interpreted, which concludes the chapter.

Part three

- Chapter 6 introduces an important tool in ROI-based rate control which is ROI detection and tracking. A state-of-the-art on this tool is presented to spot approaches needed in our work. the chapter details in a second part different controllers that have been proposed for H.264/AVC encoder. This review of available schemes helps us chose appropriate models to compare with.

- Chapter 7 presents our contributions related to ROI-based rate control. It starts by introducing our first contribution in this category which is an evolution of an ROI-based controller proposed for H.264/AVC standard using a quadratic RD model. Then, details our second contribution which is a novel rate control scheme that introduces ROI concept to HEVC R-λ model. This scheme has been introduced first in HM.10 for only inter-coded frames then improved in a later version of HEVC (HM.13) by considering ROI-based rate control for intra-coded frames. The evolution of this method is described and experimental results given. This chapter ends with a comparison of both contributions.

- Chapter 8 describes a novel ROI-based rate control method that uses tiling to perform not only independent bit allocation, but also region coding and transmission over the network. At video coding layer (VCL) tiling is performed to create separate regions, then, ROI-based rate control is done. At network abstraction layer (NAL), units of different regions are transmitted in separate streams. A packet loss model and error concealment algorithm are proposed to simulate the transmission and the decoding processes. Experiments are made to evaluates the impact of introduced features and the efficiency of the proposed approach. The obtained results of the method are reported and analyzed.

We end this manuscript with a summary of the proposed methods and their associated results, as well as some perspectives for future work in this field.

Part I

Background on video coding

Chapter 1

Image and video compression: State-of-the-art

Contents

The work done in this thesis is aimed at proposing new approaches of semantic video coding for High Efficiency Video Coding (HEVC). Consequently, we begin this thesis manuscript by introducing fundamentals of image and video compression.

In this chapter, necessity as well as feasibility of image and video compression are discussed. Then, basic concepts of image and video compression that could be useful in our researches are reviewed. Finally, proposed algorithm should be evaluated comparing to state-of-the-art approaches. We thus give details of metrics used to evaluate compression algorithm performance.

1.1 Necessity and feasibility

Image and video compression has been found to be necessary in data storage and transmission applications. In fact, the huge amount of data in these applications usually well-exceeds the capacity of todays hardware. Moreover, visual information is important for human being to help them perceive recognize and understand the surrounding word. Today, demands of video services increase considerably, we talk about high video quality like HDTV, 3D movies, video games, and so on.

Image and video compression methods have been proposed since a long time thanks to statistical and psychovisual redundancy of the data set. Generally, it is achieved by exploiting all these redundancies [2]:

- Statistical redundancy

  Statistical redundancy represent both statistical correlation between pixels within a frame (Spatial correlation) and statistical correlation between pixels from successive frames in a video (Temporal correlation) as illustrated in Fig.1.1.

  Most video coding methods exploit both temporal and spatial redundancies to achieve compression. In the spatial domain, there is usually a high similarity between pixels that are close to each other. Thus, the value of a pixel is typically close to the neighboring ones. In the temporal domain, there is usually a high similarity between frames of video that are captured at around the same time, especially if the temporal sampling rate (the frame rate) is high. Consequently, successive frames can be predicted from previous ones.

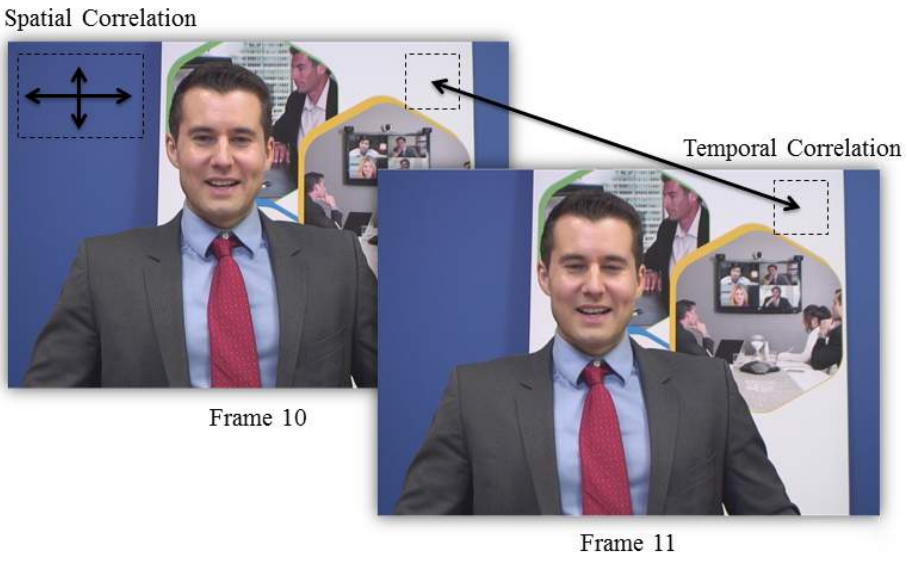


Figure 1.1: Spatial and temporal correlation in a video sequence

- Psychovisual redundancy

  Psychovisual redundancy stems from the fact that the human eye does not respond with equal intensity to all visual information. The human visual system (HVS) does not rely on quantitative analysis of individual pixel values when interpreting an image. An observer searches for distinct features and mentally combines them into recognizable groupings. In this process certain information is considered as psychovisually redudant as it is relatively less important than other.

1.2 Fundamentals of image and video compression

Image and video compression is a process in which the amount of data used to represent the image or the video is reduced to meet a bit rate requirement, while the quality of the reconstructed image or video and computational complexity satisfy application's requirements. Considering, previously described redundancies, this objective can be reached by removing redundancy and reducing irrelevance.

During the past two decades, various compression methods have been developed to address major challenges faced by digital imaging. These techniques can be classified broadly into lossless or lossy compression. Lossless compression gives only a moderate amount of compression but can completely recover the original data. On the other hand, lossy compression can achieve a high compression ratio, since it allows some acceptable degradation. Yet it cannot completely recover the original data.

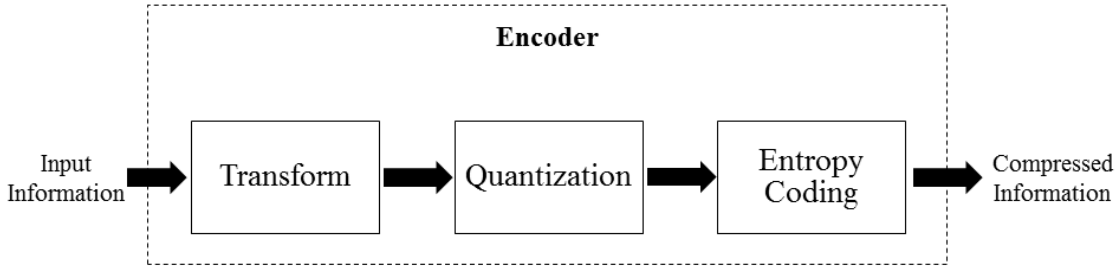


Figure 1.2: Lossy image compression scheme

Generally most lossy image compressors are implemented as three step algorithms as shown in Fig.1.2, the first two exploit both psychovisual and statistical redundancy, wile the third only statistical. The transform allows to concentrate information in as few coefficients as possible; then, quantization allows to allocate coding resource in such a way that the most relevant coefficient are finely represented, while the least important are coarsely quantized. Finally, lossless coding is used to remove residual statistical dependencies among quantize data.

1.2.1 Transform coding

The transform is the first step in image and video encoding as shown in Fig.1.2. The purpose of transformation is to convert the data into a form where compression is easier. A transform has the goal of reducing correlation [3]. The transformed values are usually smaller on average than the original ones which reduces the redundancy of representation. For lossy compression, the transform coefficients can be quantized according to their statistical properties, producing a much compressed representation of the original image data.

Linear transforms

In order to recover the original signal, the transform has to be invertible. In fact, if A is a linear transform matrix, A^{-1} is the matrix of its inverse transform. Considering an input 2-D signal X, transform coefficients in the encoder can be represented as follows:

$$\Theta = A\ X \tag{1.1}$$

At the decoder side, since transform coding implies the use of quanitzed transform coefficients $\hat{\Theta} = Q(\Theta)$, the reconstracted signal Y is:

$$Y = A^{-1}\ \hat{\Theta} \tag{1.2}$$

Practical transforms

Several linear and reversible transforms have been studied and used in transform coding, such as, the Karhunen-Loeve transform (KLT), the Discrete-Fourier transform (DFT), the Walsh transform, the Hadamard transform and the Discrete-Cosine transform (DCT) [4].

The optimal transform that decorrelates the data is the KLT [5]. It can compact the most energy in the smallest fraction of transform coefficients. However, it depends on the statistics of the encoded data. However, KLT assumes stationary signals. Knowing that images, video and music are not stationary, KLT makes no sense on them and can not be used in practice.

In practice, the DFT and DCT have been used. Studies have shown that DCT performs better than all other transforms [2]. For natural images, it compacts the energy of the signal in low frequency components. Then, it can be used for decorrelating the data. Differently from KLT, it is not data dependent. Moreover, DCT has been found to be efficient not only for still images coding but also for coding residual images in predictive coding.

1.2.2 Quantization

Quantization is essentially discretization in magnitude. It consists of the conversion of the input data from a large alphabet to a shorter one, which is an important step in lossy

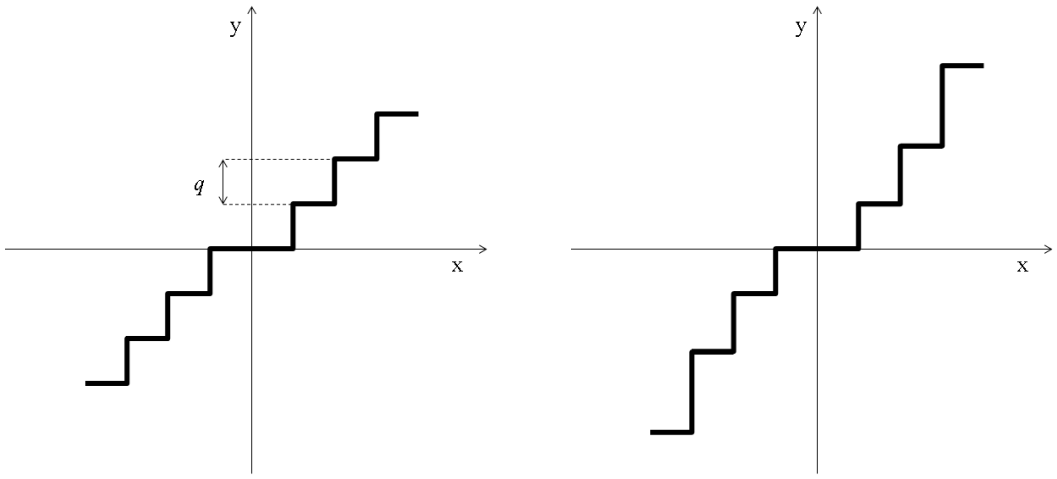
compression of digital image and video. It introduces a reconstruction errors that can be evaluated objectively or subjectively depending on the application needs and specifications [6]. In general, scalar quantization is a mapping from a set S to C, a discrete subset of cardinality N.

$$Q : x \in S \rightarrow C = \{y_1, \ldots, y_N\} \tag{1.3}$$

This means that the set S is divided into regions $R_i \subset \mathbb{R}$ for $i = 1 \ldots, N$, where $\cup_{i=1}^{N} R_i = S$. A region R_i is defined as

$$R_i = \{x \in S : Q(x) = y_i\} \tag{1.4}$$

To conclude, encoding consists in mapping the value $x \in R_k$ to the index k of the region to which it is associated. While, decoding consists in mapping the index k to the reconstruction value y_k.



(a) Uniform quantization (b) Non-uniform quantization

Figure 1.3: Input-output characteristic of uniform and non-uniform quantizer

There are three different types of scalar quantization techniques: uniform quantization, non-uniform quantization, and adaptive quantization.

Uniform quantization

If all the regions R_i have the same amplitude, and y_i is the center of the region R_i as represented in Fig.1.3(a), the quantization is then uniform. The length of the output interval is called the quantization step, denoted by q. Uniform quantization is only optimal for uniformly distributed signal. An alternative approach is to choose different quantization steps and perform non-uniform quantization.

Non-uniform quantization

In the case of non-uniform quantization (Fig.1.3(b)), Lloyd-Max algorithm gives the optimal regions and the optimal reconstruction values according to the statistics of the input signal x. Obviously, the range of values that is less probable is coarsely quantized than the most probable region [2].

Fixed quantizers may yield minimum mean square error (MSE) when assuming that signal is stationary. But this cannot be done in practical cases where signal is not stationary and is fluctuating. Thus, it is possible to adapt the properties of the quantizer to the level of the signal. This is called adaptive quantization.

Adaptive quantization

As said before, adaptive quantization attempts to make the quantizer design adapt to the varying input statistics to achieve better performance. Many researches have been done for performing adaptive quantization [7]. We distinguish two types of adaptive quantization. Forward adaptive quantization is based on a statistical analysis of the input signal to set up encoder and quantization. Then, side information are sent to the decoder quantize [8]. However, in backward adaptive quantization, statistical analysis is carried out with respect to the output quantization at both encoder and decoder sides.

1.2.3 Entropy coding

Entropy means the amount of information present in the data. For additional compression, an entropy coder encodes a given set of symbols with the minimum number of bits required to represent them. After the data has been quantized into a finite set of values, it can be encoded using an entropy coder. It is a lossless step of the compression scheme. Thus, the inverse process is able to retrieve the exact coefficients. In practice, Huffman coding, Lempel-Ziv (LZ) coding and arithmetic coding are the commonly used entropy coding schemes [9].

For example, Huffman coding utilizes a variable length code in which short code words are assigned to more common values or symbols in the data, and longer code words are assigned to less frequently occurring values. Many variations of Huffman's technique have been studied such as modified Huffman coding and dynamic Huffman coding [10].

1.2.4 Differential coding

As shown in Fig.1.2, an encoder consists in three major steps: Transform, quantization, and coding. The input information of the compressor can be a frame or a sequence. However, it is not necessarily the most suitable format for encoding. In a frame or a sequence, neighbor pixels are highly correlated as discussed in Section.1.1. They represent statistical redundancies that can be reduced using differential coding. Instead of encoding a signal

directly, the differential coding technique encodes the difference between the signal itself and its prediction. Therefore it is also known as predictive coding.

Simple pixel-to-pixel scheme

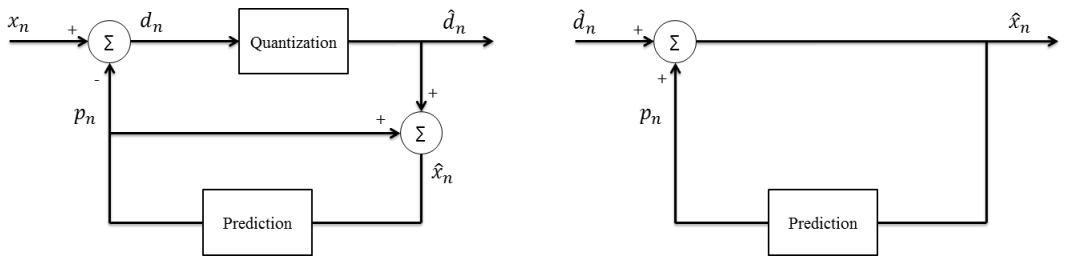
The principle of the differential coding is to encode the difference d_n between the actual sample x_n and the previous reconstructed sample $\hat{x}_{n-1}$ called predictor:

$$d_n = x_n - \hat{x}_{n-1} \tag{1.5}$$

In practice, in a simple pixel-to-pixel differential coding scheme the estimation of x_n depends on its own quantization error. This type of method is known as a Differential Pulse Coded Modulation Scheme (DPCM). In the general case, the scheme is more complex as the reference is predicted from the reconstructed signal.

General DPCM scheme

The general scheme of the DPCM system can be represented as in Fig.1.4. The used reference q_n is a function of the k previously decoded samples.



(a) Encoder (b) Decoder

Figure 1.4: DPCM system

Considering a prediction function f, the predictor is computed as following:

$$p_n = f(x_{n-1}, x_{n-2}, .., x_{n-k}) \tag{1.6}$$

Thanks to a good predictor, we have a smaller variance of the data to be quantized. Then, at the same rate, if d_n is quantized instead of x_n, we can obtain a smaller distortion.

Inter-frame differential coding

For image coding, DPCM is performed inside the same image considering spatial dependencies. They uses respectively pixels of the same scan line and pixels from different scan

lines to generate the predictor. While, video coding involves a third dimension. Thus, DPCM considers temporal dependencies between successive frames in addition to spatial redundancy. Consequently, motion estimation and compensation techniques have been introduced.

The objective of these techniques is to predict the frame I_k from a neighboring frame $\hat{I}_m$ already decoded, by estimating and compensating the objects' motion. Then, instead of sending I_k, estimated motion vectors are sent for constructing the prediction. For estimating the motion, the frame I_k is divided into blocks. A search is performed in the reference image $\hat{I}_m$ for the block that is the most similar to the current one, as represented in Fig.1.5. The vector v is called motion vector. The search is usually constrained to be within a reasonable neighborhood so as to minimize the complexity of operation.

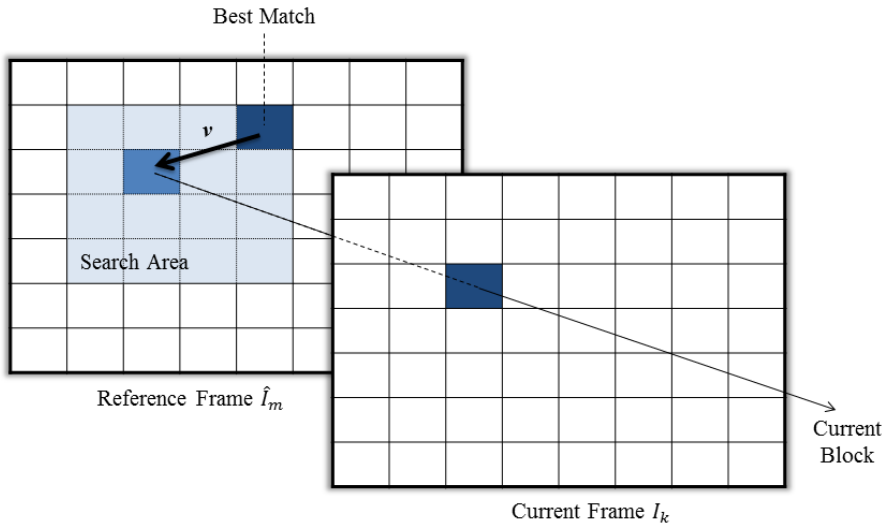


Figure 1.5: Motion estimation process

1.3 Evaluation criteria

In evaluating lossy compression methods, we first evaluate visual quality. If no difference in quality is noticed, the one that requires less data is considered to be superior to the other. Not only rate-distortion (RD) performance should be considered but also computational complexity is important metric to choose the best method. Among the evaluation criteria adopted by the community of image and video compression to highlight the performance of a proposed approach, we cite the visual quality metrics, RD curves and the measure of calculations' complexity.

1.3.1 Visual quality evaluation

As the definition of image and video compression indicates, image and video quality is an important factor in dealing with compression. In many multimedia and industrial applications, it is necessary to judge the quality of the compressed image or video with respect to the initial version of the data [11]. There are currently three major types of visual quality assessment:

- Subjective methods using a group of observers to describe the images quality.
- Objective methods that use the statistical properties of the signals. They are divided into perceptual and non-perceptual metrics.

Mean absolute difference (MAD)

The MAD is a measure of statistical dispersion equal to the average absolute difference between the original signal x and the reconstructed one x':

$$MAD = \frac{1}{N} \sum_{i \in N} |x_i - x'_i| \tag{1.7}$$

where N is the number of samples.

Mean Square Error (MSE)

It is common in image and video coding community to talk of losses information in terms of MSE. The latter is a simple difference between the original signal x and the reconstructed x':

$$MSE = \frac{1}{N} \sum_{i \in N} (x_i - x'_i)^2 \tag{1.8}$$

where N is the number of samples.

Peak-Signal-to-Noise-Ratio (PSNR)

Another significant metric derived from the MSE is the Peak-Signal-to-Noise-Ratio (PSNR)[12]. This metric represents a gain expressed in dB, and is obtained by the following equation:

$$PSNR = 10 \log_{10} \left(\frac{255^2}{MSE} \right) \tag{1.9}$$

The PSNR indicates the degree of similarity between the original signal and the reconstructed signal. It is generally calculated on the whole image, unit or group of units, it is representative of the average quality but do not account for local defects.

Structural Similarity Index (SSIM)

The structural similarity (SSIM) is a perception-based model that considers image degradation as perceived change in structural information. SSIM metric and its multi-scale extension (MS-SSIM) evaluate visual quality with a modified local measure of spatial correlation consisting of three components: mean, variance, and cross-correlation [13].

To overcome the limitations of MSE/PSNR, perceptual objective video quality assessment (VQA) has been developed. SSIM has been initially designed to measure the quality of still images and then adapted to video quality assessment. It has appeared to be the most widely spread method in recent years as an alternative quality measure of PSNR in the video coding community [14]. In our work, we used the SSIM developed in [15].

1.3.2 RD performance

A visual quality metric such as the PSNR is not enough to analyze a coding situation since it depends on another key parameter which is the bit rate measured in bpp. Thus, rate distortion curves have been adopted to assess the trade-off between the gain in bit rate and the quality of the reconstructed signal. A comparison between different coding approaches can be done using a representation of their RD curves or some metrics such as Bjontegaard metric. In fact, Bjontegaard metric (BD-PSNR) gives a single number that describes the distance between two RD-curve [16]. It is also useful to determine how big is the gain between two encoding versions, instead of curves that are more difficult to interpret.

1.3.3 Computational complexity

The computational complexity is another important performance parameter to be evaluated. We use different methods to calculate the complexity such as encoding time and memory occupancy. These criteria are used to calculate the complexity of a software and compare it with other scheme.

1.4 Conclusion

In this chapter, we start Section 1.1 by explaining the importance of video coding and its feasibility. We have then reviewed fundamentals of video coding in Section 1.2 and gave some evaluation metrics in Section 1.3.

Previously presented tools are, to this day, used in video coding standards. The next chapter will detail different image and video coding standards - from standard for continuous-tone still image coding to high efficiency video coding - for a better understanding of the evolution of coding tools. It ends with a detailed description of HEVC and its novelties.

Chapter 2

Overview of video coding standards

Contents

The two major international organization working on video compression standardization are ITU-T (International Telecommunication Union) and the ISO/IEC (International Standard Organization/International Electrotechnical Commission). The MPEG (Motion Picture Experts Group) was formed by ISO and IEC to set standards for audio and video compression. MPEG has standardized MPEG-1, MPEG-2 and MPEG-4. Meanwhile, the VCEG (Video Coding Expert Group) of the ITU-T, has finalized H.261 and H.263 standards. Finally, both organizations collaborated together in a Joint Video Team (JVT) and developped standard with better coding efficiency; H.264/AVC in 2003 and HEVC in 2013.

In this chapter, we make a summary of the evolution of video coding standards by presenting major features of each standard and justify the usage of tools described in

Section 1.2. Then, we give more details about the last in date compression standard, HEVC, which will serve as basis for comparison in manuscript.

2.1 Evolution of image and video coding standards

2.1.1 JPEG - Standard for continuous-tone still image coding

JPEG (Joint Photographic Experts Group) is a widely used method for compression of continous-tone still images (grayscale or color), standardized by ITU-T and ISO/IEC in 1992. To support a wide range of applications for storage transmission of digital images, the JPEG encoder specifies two different systems: a lossy coding of images based on the scheme represented in Fig.1.2, and a lossless compression using predictive methods [17].

For lossy compression the DCT is used in 8 by 8 non-overlapping blocks of the image. Transform coefficients are then uniformly quantized. The 64-element DCT gives 1 DC coefficient representing average color of the bloc and 63 AC coefficients representing color change across the block. Low-numbered coefficients (low-frequency color change) and high-numbered coefficients (high-frequency color change) are processed in zig-zag order. Then, a lossless coding of quantized coefficients is performed using a variable length dictionary (called Huffman coding) or an arithmetic encoder.

2.1.2 MPEG - Generic standard for coding moving pictures

To develop video coding standards, the ISO established the MPEG to find appropriate representation for moving pictures and associated audio information. They first standardize MPEG-1 in 1991 for digital media storage at 1.5 Mbits/s bit rate. Then, MPEG-2 came to complete the first standard by allowing greater input format flexibility and higher data rate. In 1999, MPEG-4 has been approved for standardization providing new profiles including higher compression efficiency and many novel coding concepts such as interactive graphics, object and shape coding [18].

New coding features have been introduced by MPEG to achieve high compression ratio when encoding moving picture. First, the group of pictures (GOP) concept divides the sequence into a sequence of frames of different types (I, P and B pictures) as shown in Fig.2.1. I-frames are self-sufficient intra-coded pictures. P-frames or predictive-coded pictures are coded using one directional motion compensation prediction from a previous frame. B-frames can be coded using either past or future anchor images. Second, the macro-block (MB) concept divides each frame in fixed-sized non-overlapped blocks. Each MB contain one luminance (Y) block of 16×16 pixels and two chronimance (C_b and C_r) blocks of 8×8 pixels [19].

MPEG encoding scheme starts with an inter prediction per block. DCT is then used for coding both intra pixels and predictive error pixels. At last, transformed coefficients are quantized and coded by a variable length coder. Furthermore, rate control method

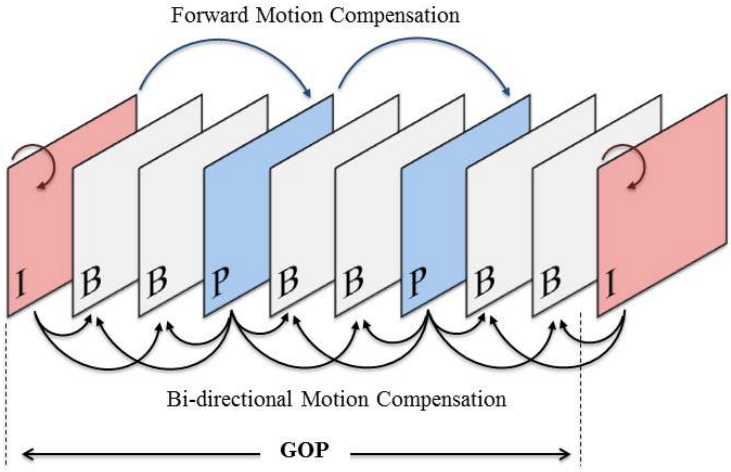


Figure 2.1: Example of GOP

has been introduced in test model 5 (TM5) for MPEG-2 to adapt the MB quantization parameter to bit rate limitation. TM5 algorithm will be detailed in Section 4.1.

2.1.3 H.261 and H.263 - Video coding standards for ISDN applications

H.261 standard was developed by ITU-T study group for low target rate applications suitable for transmission of color video over integrated services digital network (ISDN). It has been standardized in 1993 and present many common features with MPEG-1 standard. Later on, an improved version of H.261 has been designed for low bit rate applications. This standard is called H.263 and presents some novelties such as an advanced prediction mode that allows up to four motion vectors to be used per MB [20].

2.1.4 H.264/AVC - Advanced video coding for high coding efficiency

The ISO/IEC and ITU-T VCEG collaborated together in a joint video team (JVT) and developed a coding standard of high coding efficiency. Based on conventional block-based motion compensated hybrid video coding concept, H.264/AVC provides approximatively 50% bit rate saving for equivalent perceptual quality relative to MPEG-4 for high resolutions and a bit rate saving between 30-40% for low resolutions [21]. Experiments in [22] also demonstrated that the H.264 standard can achieve 50% coding gain over MPEG-2, 47% coding gain over H.263 baseline, and 24% coding gain over H.263 high profile encoders.

H.264 standard proposes a layered structure: a video coding layer (VCL) and a network abstraction layer (NAL). At VCL, the input video is compressed into a bitstream which is divided into NAL units that carries encoding informations [23]. In fact, NAL is a new concept that offers efficient transmission of the compressed stream. Except many common tools, the H.264/AVC includes many features able to improve coding efficiency. Variable block size with smaller block sizes offer more flexibility to the encoder. The MB size can

go from 4×4 to 16×16 pixels. Moreover, motion compensation uses multiple reference pictures and weighted prediction. Directional spatial prediction is adopted for improving intra-coding performance and new coding modes have been introduced for P-frames such as skip and direct modes. In these modes, the reconstructed signal is obtained directly from the reference frame with the motion vectors derived from previously encoded information. There are also other technical tools such as deblocking filters, flexible MB ordering, new entropy coding methods, etc [24].

H.264/AVC was designed for both low bitrate and high bitrate video coding in order to accommodate the increasing diversification of transport layer and storage media. Many works on new sets of extensions have been completed which gave a rise to a wide variety of H.264-based products and services including video telephony, video conferencing, TV, storage, video streaming, digital cinema and others [25].

2.2 High Efficiency Video Coding (HEVC)

In this section, we introduce the last developed video coding standard HEVC since it is used in our experiments. The main objective of HEVC is high efficiency video coding for ultra high definition (UHD) content. This new standard has introduced new tools to deliver the same video quality at half bit rate comparing to H.264/AVC [26]. We provide an overview of technical features of HEVC and review encoder performance using H.264/AVC as a reference [27].

2.2.1 Codec structure

HEVC is based on the same hybrid spatial-temporal prediction system as its predecessor H.264/AVC. Fig.2.2 shows the block diagram of the basic HEVC encoder. The main structure of the HEVC encoder looks like H.264/AVC one [28]. The main key features of HEVC can be summarized in the following 8 points:

- New coding, prediction and transform partitioning.
- Advanced motion vector prediction and introduction of motion sharing.
- Improvement of motion vector precision in inter-prediction.
- Up to 35 directional orientations for intra-picture prediction.
- Directional transform and quantization matrix adaptation.
- Simplified design of De-blocking filter (DBF) and sample adaptive offset (SAO) filter.
- Context adaptive binary arithmetic coding (CABAC) algorithm for entropy coding.
- Slice, tile structure and wavefront parallel processing (WPP) for parallel encoding.

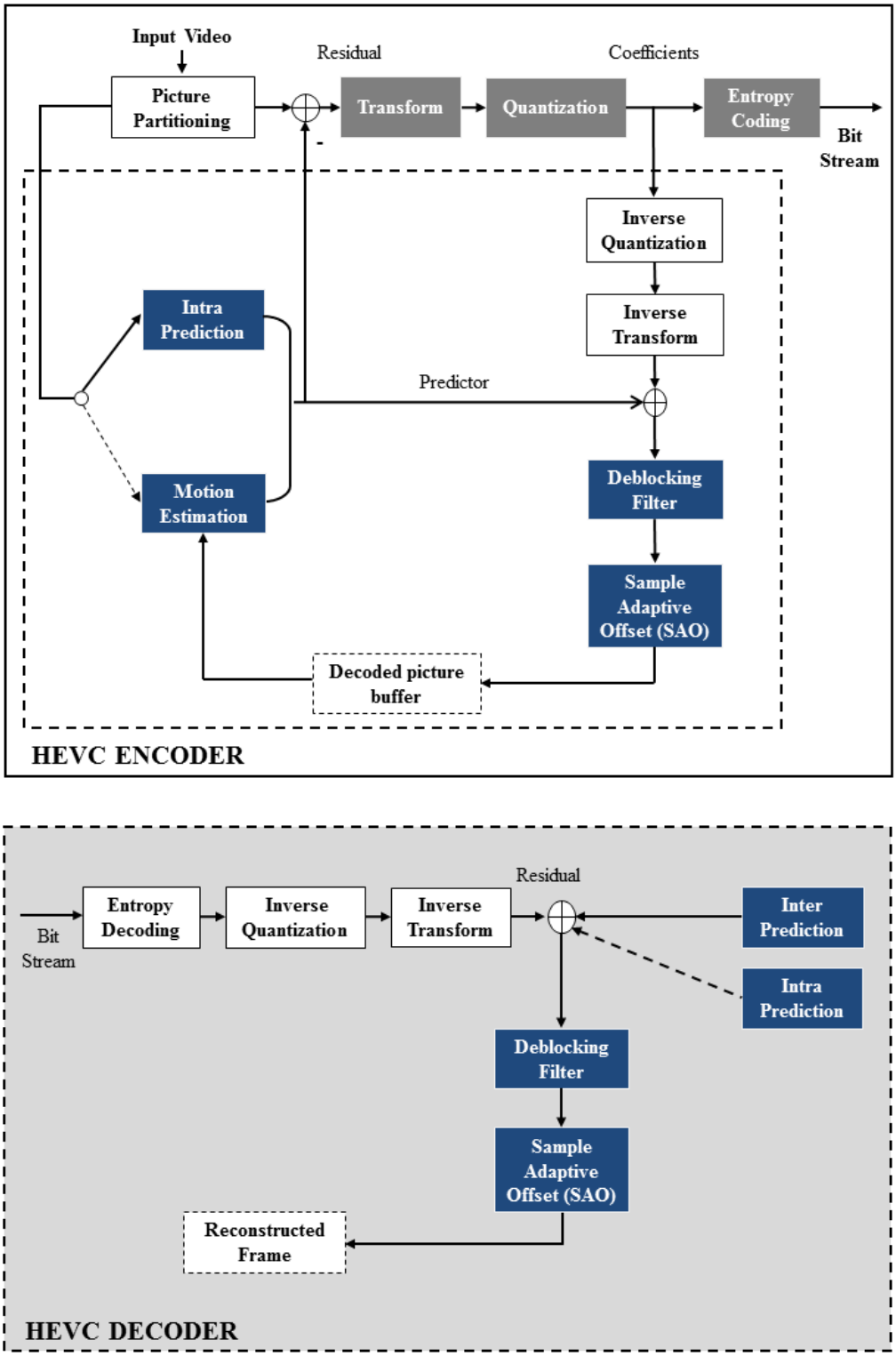


Figure 2.2: HEVC encoder and decoder structure

2.2.2 Main technical features

To perform perceptual video coding, some technical tools should be studied. We describe in this section HEVC encoder tools needed in our researches.

Coding, prediction and transform partitioning

Compared to the use of fixed size macroblock, different coding units' sizes enable the codec to be optimized for various content, applications and devices. It is especially useful for low resolution video services, which is still commonly used in the market. HEVC uses advanced quadtree-based approach [29]. It is based on choosing the size of largest coding unit (LCU) or coding tree unit (CTU) and maximum hierarchical depth to construct a hierarchical block structure that can be optimized in a better way for the targeted application as shown in Fig.2.3.

The coding unit (CU) is the basic unit of region splitting used for both intra and inter prediction modes. CUs ensure a sub-partitioning of an image into square regions of equal or variable size (8x8, 16x16, 32x32 or 64x64). The prediction unit (PU) defines a basic unit used for carrying the information related to the prediction processes. Each coding unit is divided into PUs to perform prediction. Asymmetric splitting can be performed to operate prediction efficiently as shown in Fig.2.3. PU splitting and the prediction type are two concepts that describe the prediction method which make PU basically the elementary unit for prediction. Finally, TUs are used for the transform and quantization processes and they are arranged in a quad-tree structure. These three new concepts help the coding to be as flexible as possible and to adapt the compression prediction to image peculiarities.

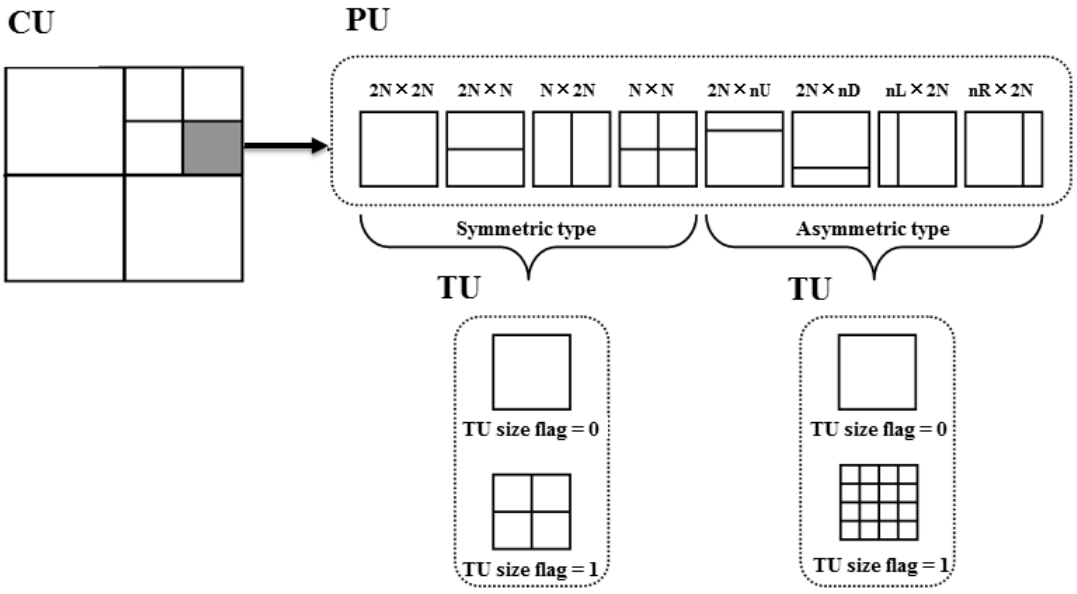


Figure 2.3: Hierarchical block structure

This partitioning respects some rules: transform unit can have same or smaller size than PU in intra prediction and same or smaller size than CU in inter prediction. For intra prediction, a PU is the same size as CU, or it can be further divided only if the CU is the Smallest CU. These flexible units used for compression have a hierarchy following quad-tree block structure as represented below.

Directional transform and uniform reconstruction quantization (URQ)

Transform coding in HEVC standard has been a very important field of study during its standardization. One of the most relevant changes with regards to previous coding standards is the replacement of the discrete cosine transform (DCT) in favour of the discrete sine transform (DST) for the 4×4 intra prediction luma residuals. This change provides approximately 1% bit rate reduction in intra-predictive coding.

Currently, HEVC selects the optimal residual in RD by choosing the best combination of transform size and intra-prediction mode - in other words, by performing rate distortion optimization (RDO). The URQ is used in HEVC, with quantization scaling matrices supported for the various transform block size. The residual is represented in the transform domain according to its TU size, that is, DST for 4×4 luminance component and DCT for all other cases. Transformed coefficients are quantized considering a particular quantization step q. A quantization parameter QP is used per sequence, frame or CTU to determine the q. In HEVC, QP can take 52 values from 0 to 51. The relationship between QP and equivalent q is:

$$q = 2^{\frac{QP-4}{6}} \tag{2.1}$$

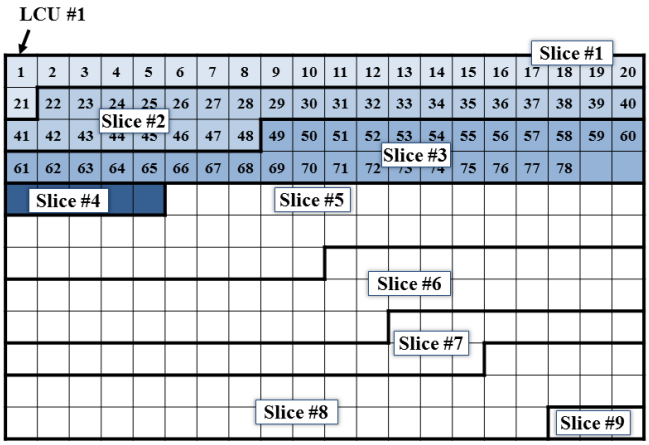
Slice and tile structures

In general cases, a bitstream is transmitted over a lossy channels. Once in the decoder side, a loss in the stream leads to an inability to reconstruct the signal. Error is then propagated, as all regions of the sequence are dependently decodable. To limit this propagation of error, new structures that break dependencies in processing such as slices and tiles have been introduced in HEVC [30].

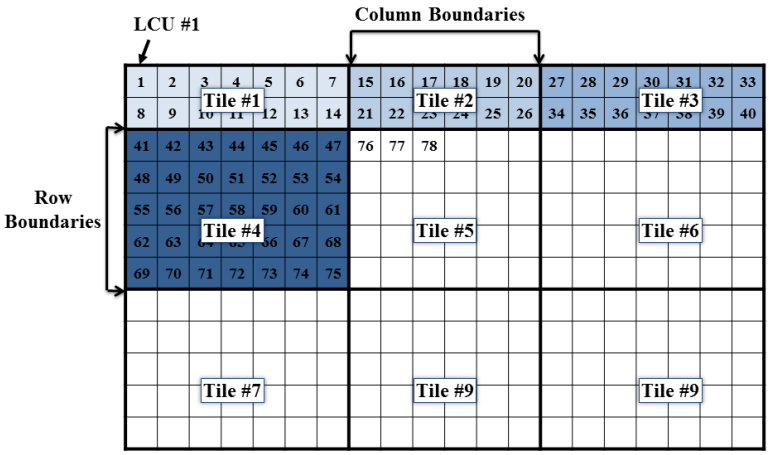
Slices define groups of independently decodable CTUs. The units in a slice follow a raster scan order as shown in Fig.2.4(a). However, tiles are rectangular independently decodable sets of CTUs as represented in Fig.2.4(b). This new feature introduced in HEVC standard offers a flexible classification of CTUs, a higher pixel correlation compared to slice and better coding efficiency as tiles do not contain header informations [28].

The benefits of tiling have been assessed in [30]. First, tiles offer better R-D performance in case of high level parallelization. Second, they facilitate improved maximum transmission unit (MTU) size matching comparing to traditional slices. Then, tiling can be used for additional region on interest (ROI) based functionality, to ensure that the ROI tiles are

independently decodable from the non-ROI tiles and that temporal and spatial predictions within the ROI do not refer to pixels outside the ROI [31]. ROI tile sections have been studied in different works to ensure a good fitting of the region and its corresponding tile [32]. They can be used for a tiled streaming for zoomable video, where all tiles are temporally aligned for an efficient bandwidth utilization and ROI quality improvement [33]. Moreover, tiles can be used in video conferencing application with multiple people for an efficient processing of the stream. In [34], faces are detected, placed in separate tiles and reassembled in a customized virtual scene.



(a) Slice partitioning



(b) Tile partitioning

Figure 2.4: Examples of slice-based and tile-based partitioning

2.2.3 Encoder control of the HEVC test model

Besides the development of the specification text, a test model document is maintained which describes the encoder control and algorithm implemented in the reference software.

Different versions of HEVC test model (HM) have been proposed and their corresponding implementation maintained in a subversion repository [35]. Three different versions of the standard have been approved by the IUT. Each version of the encoder contains profiles and levels that considers some evolutions by introducing extensions, reducing complexity and improving RD performance. It supports appropriate configurations and control for different applications. The latest update of the HEVC test model is HM-16.7 [36].

The HEVC profiles, tiers and levels

A profile defines a set of coding tools or algorithms that can be used in generating a conforming bitstream, whereas a tier and a level place constraints on certain key parameters of the bitstream, corresponding to decoder processing load and memory capabilities [28].

The first version of HEVC standard has been approved in 2013 and contains three profiles: Main, Main 10 and Main Still Picture. The second version came one year after and added 21 range extension profiles two scalable extension profiles, and one multi-view extension profile. Finally in 2015, the third approved version of HEVC encoder added the 3D Main profile [37].

The HEVC standard defines two tiers, Main and High. The Main tier was designed for most applications while the High tier was designed for very demanding applications. Moreover, HEVC standard introduces thirteen levels that support different resolutions from small picture sizes such as a luma picture size of 176×144 (support by the first level) to picture sizes as large as 7680×4320 called 4k and 8k resolutions (supported by the thirteenth level) [28].

HEVC encoder configurations

All versions are based on three kinds of temporal prediction structures: intra-only, low-delay and random access. The coding strategy changes from one configuration to another. It helps to choose which prediction to use for encoding one particular frame and fix the reference images for inter prediction. The reference picture list management depends on the configuration set before the coding. Moreover, the configuration file generated at the start of the encoding defines prediction parameter decision detailed in the software manual [38] such as the CTU maximum size, the quadtree depth, images hierarchy, quantization and filters' parameters, etc.

- Intra-only configuration

  Using the Intra-only coding configuration, each picture in a video sequence is encoded as instantaneous decoding refresh (IDR) picture unused for reference. Each GOP of the video contains only one image with an empty reference buffer, because, the decoding is based on intra prediction only. In fact, the I-frame period is equal to one. The encoding order is the display order and no temporal references are used.

- Low-delay configuration

  Two kinds of low-delay coding configurations have been defined for testing coding performance in low-delay mode: low-delay P and low-delay B. For these low-delay coding conditions, only the first picture in a video sequence is encoded as IDR picture, while the other successive pictures are encoded as generalized P- and B-pictures (GPB). The GPB are encoded using inter prediction. They use reference pictures with smaller picture order count (POC) than the current image and which are kept in two memory buffers List0 and List1. In fact, all reference pictures in these lists are temporally previous in display order relative to the current picture. The contents of the two lists are identical, and they are updated with sliding-window management process. The size of the reference picture list and the positions of references to chose are defined in the configuration file initially generated. The reference picture list combination is derived from the entries of List0 and List1 and used for reference index management and entropy coding.

- Random-access configuration

  For the random-access test condition, hierarchical B structure is used for coding. The images are not any more coded in their display order like in low-delay configuration. The QP levels are used to define an hierarchical ordering of the frames in each GOP. In this kind of configuration, the inter prediction can be made using frames of smaller and/or bigger POC. In fact, the configuration file contains negative and positive references' indexes. Intra frame period is defined such as an I-picture is inserted cyclically per about one second. The first intra picture of a video sequence is encoded as IDR frame, but, the other intra images are encoded as non-IDR intra pictures. The images located between successive intra pictures in display order are encoded as B-pictures.

2.2.4 HEVC performance and applications

The introduction of larger block structures has impact on motion vector compression, added to PU and TU structure, deblocking filters, and all previously described tools contribute to the improvement of coding performance of HEVC. It has shown to be especially effective for low bit rates, high resolution video content, and low-delay communication applications. According to multiple studies, HEVC should deliver up to 50% better compression than H.264 in interactive applications such as videoconferencing, which means similar quality at half the bitrate as presented in comparative works described in[9] and [26]. Furthermore, HEVC can perform real time coding thanks to reduced computational complexity. Alternatively, HEVC can also be used for interactive applications as it well performs coding of high resolution videos and enable larger resolution movies, whether 2K or 4K.

Essentially, these are the two benefits of HEVC in the streaming space. The first relates to encoding existing SD and HD content with HEVC rather than H.264, enabling cost savings and the ability to stream higher quality video to lower bit rate connections. The second relates to opening up new markets for ultra-high-definition (UHD) videos. Finally, the development of many extensions enhance the utility of HEVC standard and broaden its range of applications [40]. In fact, scalability extension to HEVC (SHVC) enables spatial and coarse gain SNR scalability and 3D and multiview extensions enable efficient compression of stereo and multiview video content.

2.3 Conclusion

In this chapter, we have presented an overview of all published video coding standard. Section 2.1 introduces main technical features of previous codecs such as MPEG-1, MPEG-2, H.261, H.263 and H.264/AVC. While, Section 2.2 exposes novelties of HEVC coding and main technical normative tools.

The rate control algorithm is often not standardized, since it can be independent of the decoder structure. However, it plays a critical role in perceptual video coding. Thus, it is important to study rate control in video coding, specially in high efficiency video coding. The next two chapters introduce rate control principles and detail different scheme adopted the state of the art. Among previously described standards, Section 4.2 focuses on rate control problem in HEVC.

Part II

Rate control and rate distortion modeling

Chapter 3

Rate control theory

Contents

In image and video coding systems, compression rate fluctuation depends on the amount of data we accept to lose. The spatial and temporal variability of video content cause variations in coding efficiency, resulting in important fluctuations in bit rate and quality of the encoder output. Rate control is therefore an important step in data coding and transmission.

In this chapter, we first introduce the basics of rate-distortion (RD) theory and describe the trade-off between lossy compression rate and the resulting distortion. We introduce the rate distortion function and explain the process of rate-distortion optimization (RDO). Then, we study different rate distortion models proposed in the literature. Finally, rate control key techniques used to maintain consistent quality under transmission channel constraints are detailed.

3.1 Rate distortion theory

Rate distortion theory was created by Claude Shannon [41]. It is the branch of information theory that describes the compromise between the compression rate and the resulting distortion for a lossy source coding to ensure a stable coding efficiency. Knowing that a decrease in the bit rate leads to an increase of the output distortion and vice versa, this theory addresses the problem of determining the minimal amount of information that should be communicated over a channel such that the source can be reconstructed at the receiver side with a given distortion.

RD theory gives an analytical formulation for the problem. It models how much compression can be achieved using lossy compression methods. Before being transmitted in the network, signals such as audio, image and video are compressed using techniques based on transform and quantization procedures that capitalize on the shape of a rate distortion function.

3.1.1 Rate distortion function

The amount of information at a discrete probability distribution p_i is measured by the entropy:

$$H = -\sum p_i \log p_i \tag{3.1}$$

and the mutual information between the source signal X and the reconstructed one $\hat{X}$ is defined as:

$$I(X;\hat{X}) = H(X) - H(X|\hat{X}) \tag{3.2}$$

The RD problem was introduced by Claude Shannon in a first paper in 1948 [41], and extensively studied in his 1959 paper [42]. As rate distortion function reflects the mutual information between source signals and decoded ones, the fundamental theorem of RD theory can be formulated as follows:

Theorem. *The rate distortion function for an i.i.d. source X with distribution $p(x)$ and bounded distortion function $d(x,\hat{x})$ is equal to*

$$R(D) = \min_{p(\hat{x}|x):\sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x|\hat{x}) \leq D} I(X;\hat{X}) \tag{3.3}$$

where $R(D)$ is the achievable rate at distortion D.

In fact, for a given source a closure of achievable rate distortion pairs (R, D) represents the rate distortion region. A rate distortion function $R(D)$ is defined as the infimum of rates, such that for a given distortion, the couple is in the rate distortion region of the source. $R(D)$ and $D(R)$ are two equivalent functions that include the same information and that are used in video coding to find the optimal (R, D) couple.

3.1.2 Rate distortion optimization

Video coding usually incorporates rate distortion optimization (RDO) techniques, to solve the previously described problem (3.3). For each image or region, the optimization task consists in finding the most efficient coded representation (prediction mode, motion vector, quantization level, etc.) in the rate distortion sense. This task is complicated by the fact that various options show varying efficiency at different bit rates and with different scene content [43]. The objective of rate distortion optimization is to find the minimum number of bits needed to represent the source data at a given distortion. Or, equivalently, the minimum expected distortion achievable at a particular rate as demonstrated in Fig.3.1 by horizontal and vertical arrows.

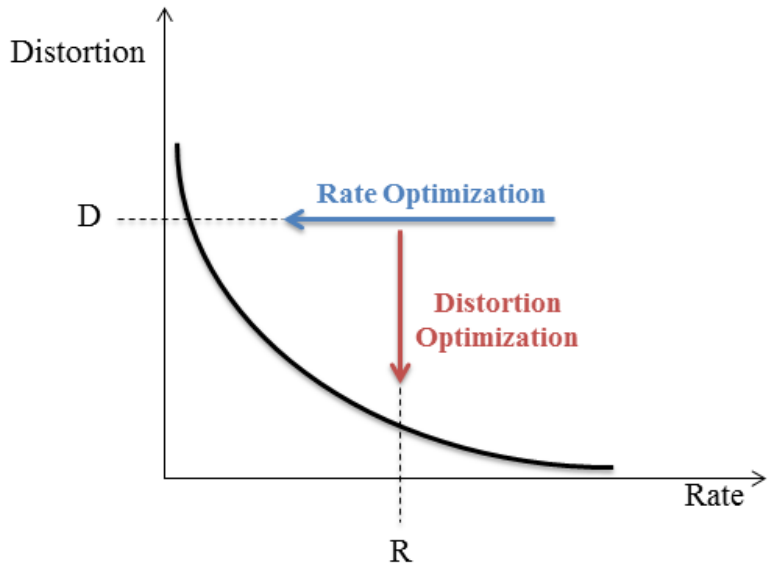


Figure 3.1: Rate-distortion optimization

Knowing that quantization consists in reducing the bit rate of the compressed video signal by accepting a loss in information (or, a certain distortion of th reconstructed signal), one of the major roles of rate control algorithms is thus to find for each transform coefficient the appropriate quantization step q under the constraint $R(q) \leq R_{\max}$. The fixed bit budget is $R_{\max}$ and $R(q)$ is the number of coding bits for the source data. If we note D the distortion measure between the original and the reconstructed samples, the RD optimization problem can be formulated as:

$$\min_{q} D(q) \quad \text{subject to} \quad R = (q) \leq R_{\max} \tag{3.4}$$

In practice, the minimization problem is reformulated using the Lagrangian multiplier

as follows to compute an optimal q per sample (frame for example):

$$\min_{q} J(q), \quad \text{where} \quad J(q) = D(q) + \lambda\, R(q) \tag{3.5}$$

where the Lagrangian rate distortion functional J minimized for a particular value of the multiplier λ. Each obtained quantization step q for a given λ corresponds to an optimal solution [43] [21].

If the sample is partitioned into N subsets (for example regions or coding units) in a way that different quantization steps q_i are associated to different subsets and an additive distortion measure $D(q_i)$ is used, the minimization problem can be written as:

$$\min_{q_{i\in[1;N]}} J(q_{i\in[1;N]}), \quad \text{where} \quad J(q_{i\in[1;N]}) = \sum_{i=1}^{N}(D(q_i) + \lambda\, R(q_i)) \tag{3.6}$$

The optimal solution of this optimization problem is a set of quantization steps $q_{i\in[1;N]}$ that minimizes the global RD performance.

3.2 Rate distortion models

3.2.1 Rate modeling

RDO technique helps finding the best representation in the rate distortion sense. This problem needs explicit models that relate the average bit rate and the distortion to the quantization parameter (QP) or the quantization step q. In video coding, several works have been done in perceptual quality, for estimating the distortion, and in rate modeling. Different rate models have been developed, some of them based on simple linear expressions, others on more complex mathematical representations.

Simple linear rate distortion model

The traditional linear model introduced in [44] was employed in the final test model of MPEG-2 (TM5) and is defined as follows:

$$R(q) = \frac{X}{q} \tag{3.7}$$

where X is the model parameter.

Quadratic model

A quadratic rate quantization model has been adopted later on. It is represented as:

$$R(q) = \frac{a}{q} + \frac{b}{q^2} \tag{3.8}$$

This so-called quadratic rate model has been used for rate control in VM8 for MPEG-4 reference encoder [45] to calculate the quantization step q. We note that by choosing the appropriate a and b, this model can realize the inverse power model of the linear representation with any $\gamma \in (1, 2)$.

In order to enhance the accurancy of the RD model, d the mean absolute difference MAD between the original frame and the reconstructed one is considered for H.264/AVC [46] and also for HEVC [47] as follows:

$$R(q) = \frac{a \times d}{q} + \frac{b \times d}{q^2} \tag{3.9}$$

Model coefficients a and b are updated after encoding each frame. Here, q is the quantization step size defined in the standard by a function of the quantization parameter QP. The accuracy of these models has been enhanced by introducing the so-called complexity of the source, using the per pixel gradient value in the R-q model in [48].Alternatively, the sum of absolute transformed differences (SAD) has been adopted in [49].

ρ Domain linear model

In a different way, the RC was improved by considering a representation in the ρ domain [50] as proposed in [51]. The proportion of the coefficients after quantization to zero, ρ, increases in a monotonic way with the growth of the quantization step, which leads to a new representation of the problem based on R-ρ relationship:

$$R(q) = \theta(1 - \rho(q)) \tag{3.10}$$

where θ is a constant. However, this model does not provide explicit relation between q and ρ. It does not lend itself to theoretical understanding of the impact of the quantization parameter on the rate.

Exponential model

In [52], an intra-only rate control scheme based on an exponential R-q model is proposed. Through experiments based on extensive testing data, the relationship is modeled as follows:

$$R(q) = \alpha e^{-\beta q} \tag{3.11}$$

where α and β are the model parameters.

Rate model under variable frame rate

In [53], a model was built considering on the first hand, the impact of frame rate$t$ on the bit rate R, under the same quantization stepsize q, and in the second hand, the impact

of q on the rate, when the video is coded at a fixed frame rate. This leads to a different representation of the RD model based on the variation of q and t, and can be written as:

$$R(q,t) = R_{max}(\frac{q}{q_{min}})^{-a}(\frac{t}{t_{max}})^{b} \tag{3.12}$$

where q_{min} and t_{max} are chosen based on the underlying application, R_{max} is the actual rate when coding a video at q_{min} and t_{max}, and a and b are the model parameters.

The R-λ model

The most recent rate distortion model in the HEVC reference software is the R-λ model expressed as follows:

$$\lambda = \alpha\ R^{\beta} \tag{3.13}$$

where α and β are the model parameters [54]. We note that this model defines a relationship between the rate in bits per pixel, R, and the Lagrange parameter λ which is used in RDO to decide the coding mode. Using this R-λ model, λ is generated first at frame and/or coding tree unit (CTU) level, and then used to compute the QP.

The R-λ has been proposed to characterize the relationship between R and λ and also between λ and QP. Previous investigations in [55] have put stress on the importance of the Lagrangian parameter and have shown that the QP and $\ln(\lambda)$ are in good linear relationship:

$$QP = 4.2005 \times \ln(\lambda) + 13.7122 \tag{3.14}$$

Research done in [56] has shown that the relationships between λ and QP (or q) depends on the type frame (I, P or B) and its hierarchical level represented by the QP factor p. The model is defined as follows:

$$\lambda = p \times 2^{\frac{\mathrm{QP}-12}{3}} \tag{3.15}$$

For example for intra-frames $p = 0.57 \times (1 - \max(0, \min(0.5, 0.05 \times N_B)))$ where N_b is the number of B frames in the GOP.

3.2.2 Distortion modeling

For visual quality, a distortion model is usually developed to help predicting the relationship between the quality degradation D and the quantization step q. In fact, as the used distortion metrics vary from one work to another, different D-q models have been proposed.

For example the sum of squared errors between the original source and the predicted one (SSE) is used in [57]. The model is defined as:

$$\mathrm{SSE} = a\ q^{b} \tag{3.16}$$

where SSE is the sum of squared errors between the original source and the predicted one,

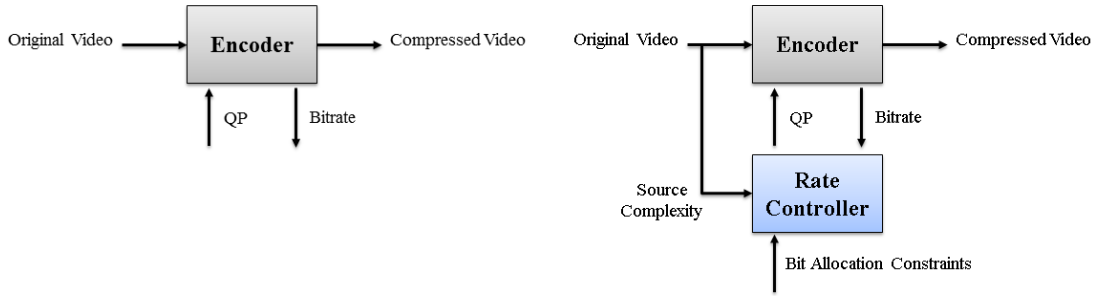
a and b are the model parameters.

In other contributions such as [58] and [59], the mean square error (MSE) is modeled as follows:

$$\text{MSE} = \frac{q^2}{12} \tag{3.17}$$

3.3 Rate control techniques

Rate control is a necessary part of the encoder, and has been widely applied in standards. The objective of RC is to achieve a target bit rate as close as possible to a given constant over time while ensuring minimum quality distortion. Rate control usually incorporates RDO, for better coding efficiency. It is a way to deal with varying bit rate and keep a good visual quality of the decoded video.



(a) Encoder without rate controller (b) Encoder with rate controller

Figure 3.2: Comparison of encoding schemes with and without rate controller

As shown in Fig.3.2(a), if no RC is performed, the encoder takes as input the original video and a fixed QP. In that case decoding quality will be fairly constant depending on the video content while output bitrate will be varying depending on the input complexity. When introducing a rate controller (Fig.3.2(b)), the QPs are computed by the controller considering different constraints and the source complexity. The output bitrate will be constant. In this process, there are two major steps: bit allocation and QP calculation [60].

3.3.1 Bit allocation

Before transmission, all rate fluctuations must be effectively controlled, since the actual network bandwidth and storage capacity are limited. Bit allocation problem arises in many situations where the rate constraints are driven by the total bit budget and transmission delay.

Budget-constrained allocation

Networks and storage media can operate in constant bit rate (CBR) or variable bit rate (VBR). In both cases, the encoding of the video sequence must be adjusted to meet the

network bandwidth and storage media capacity requirements such that the number of bits used for encoding all the N frames of the sequence, $\sum_{i=1}^{N} r_i$, does not exceed network limitations R_{max}.

Delay-constrained allocation

Some applications can introduce delay constraints. In fact, when a sequence is streamed, each coding unit is subject to a delay constraint. The video encoder ensures that the rate selection per frame is such that no frame arrives too late at the decoder.

Buffer-constrained allocation

In both CBR and VBR networks, data is stored in buffers at the encoder and decoder sides. Bit allocation does not let buffer occupancy $B(i)$ exceed physical buffer storage limitation B_{max} at time i.

3.3.2 Quantization parameter calculation

After DCT transformation, the residual signal is quantized to form a final estimate. It is then important to chose in this second step the appropriate quantization step q, the one that optimizes the RD performance of the encoder. This step consists in using allocated budget, R-q models and input data complexity to calculate the appropriate quantization step q or parameter QP. Once q computed the unit is encoded and rate control process parameters are updated. This approach is detailed in the next section for High Efficiency Video Coding standard.

3.4 Conclusion

This chapter is a brief introduction of rate control principles. It starts with a global description of the rate-distortion problem to end with an appropriate rate control scheme for video coding.

Each described model and method has been commanded by a standard during the development, e.g. TM5 for MPEG-2, VM8 for MPEG-4, TMN8 for H.263 and quadratic model for H.264/AVC. All these algorithms are studied in the next chapter and different HEVC rate control algorithms are detailed, evaluated and compared in Section 4.2.

Chapter 4

Rate control in video coding

Contents

In video coding systems, rate control is a non-normative tool that has being studied and incorporated in video codecs. Rate control modules have been developed to adjust the output bit rate and ensure high visual quality of the decoded video in constrained network conditions. Furthermore, for each standardized encoder a rate control algorithm has been proposed and studied taking into account the new features introduced by each codec. Thus, it is important to study different proposed approaches and evaluate the HEVC controller before introducing our ROI constraint.

This chapter summarizes classic rate control algorithms by putting the stress on the evolution of their bit allocation processes and rate distortion models. It ends with a detailed description of HEVC rate control algorithms. HEVC controllers are tested and their performance are evaluated to motivate the choices we made during our researches.

4.1 Evolution of rate control schemes

In video coding, controllers have been designed to achieve the main goals of high coding efficiency and accurate matching of the target rate. The classical algorithms or models are the TM5 in MPEG-2, the VM8 in MPEG-4 and TMN8 in H.263.

Rate control in recent standards such as H.264/AVC and HEVC are more complex than these classical algorithms. In fact, statistics of the current frame are not available for the rate control. This is because the quantization parameters are involved in both rate control and RDO, while, they are only involved in rate control in MPEG-2, MPEG-4, and H.263. Consequently, a new procedure has been developed for H.264 to solve this so-called "chicken and egg" dilemma [46]. The same controller has been extended to HEVC and then replaced by a recent rate control algorithm based on an hyperbolic R-λ model [54]. The accuracy of these models has been enhanced in different propositions.

4.1.1 Classical rate control schemes

Rate control of TM5

TM5 is the final test model of MPEG-2. Its rate controller consists of three steps at three operating layers to adapt the MB quantization parameter for controlling the bit rate [61]:

- Target bit allocation at GOP and frame level : This step first allocates a bit budget to the GOP, based on the target rate. Then, a number of bits is allocated per picture considering several factors: frame type (I,P or B), buffer fullness, and picture complexity.
- Rate control at MB level : Within a picture, the bit budget allocated in the previous step is divided to its MBs. Consequently, the quantization step q_j of the j^{th} MB is derived from the target bit rate R_{max}, the frame rate f and the virtual buffer fullness V_j when encoding the current block:

$$q_j = \frac{V_j \times f}{2\ R_{max}} \tag{4.1}$$

- Adaptive quantization : The last step consists in modulating the quantization step considering the MB complexity. In fact, as human eyes are not sensitive to quantization noise for active areas, q is increased in this regions and reduced for smooth areas.

Rate control in VM8

The rate control algorithm in MPEG-4 verification model VM8 follows almost the same bit allocation scheme as in MPEG-2 test model TM5 [45]. However, a new rate model has been introduced to improve QP optimization over different MBs. In fact, a target bit

rate is allocated per frame considering frame rate, the target budget and the complexity of the previous frame. Then, once the budget is divided between MBs, the quantization parameter QP is computed using the quadratic model given by Equation (3.8) and clipped from 1 to 31. After encoding each frame, the model parameters are updated.

This algorithm presents some limitations as it considers statistical information of previous frame, without any consideration of the real complexity of the current frame. Moreover, it skips the next frame when the buffer fullness riches 80%.

Rate control in TMN8

In TMN8, the controller includes two major steps. First, a bit budget is allocated per picture considering the maximum rate, the frame rate, the buffer status and the skip frame threshold. Here again, a number of frames are skipped considering a fixed threshold. Second, an adaptive computation of the quantization step q per MB is performed [62]. The controller uses the following RD model:

$$R(q) = \begin{cases} \frac{1}{2}\log(2e^2\ \frac{\delta^2}{q^2}), & \frac{\delta^2}{q^2} > \frac{1}{2e} \\ \frac{e}{\ln 2}\ \frac{\delta^2}{q^2} & , \frac{\delta^2}{q^2} \leq \frac{1}{2e} \end{cases} \tag{4.2}$$

Contrary to VM8, this model considers statistics of the current frame as it depends on the standard deviation δ of the residue in the current MB.

4.1.2 Rate control in H.264/AVC

GOP level rate control

In this level, a total number of bits is allocated for the current GOP and a QP value is initialized. Using the number of bits of the GOP, an initial bit budget is allocated to the j^{th} frame $T_g(j)$. It is computed using sequence frame rate f, available bandwidth $R(j)$ and the number of frames per GOP N as follows:

$$T_g(j) = \begin{cases} \frac{R(1)}{f}N + T_{g-1}(N) & , \quad j = 1 \\ T_g(j-1) - T'_g(j-1) + \frac{R(j)-R(j-1)}{f}(N-j+1), & j = 2, 3, ..., N \end{cases} \tag{4.3}$$

where $T_{g-1}(N)$ is the budget allocated to the last frame of the previous GOP, $T'_g(j)$ is the real encoding bits of the j^{th} frame.

Moreover, an initial quantization parameter is initialized per GOP considering QPs of all the frames of the previous GOP and the GOP length.

Frame level rate control

At frame level, previously computed bit budget and QP are adapted to frame type (reference or non-reference). For frames not used for reference a QP is computed through a linear

interpolation of QPs of the previously decoded frames. However, if the frame is used as reference, the picture budget is adjusted according to the buffer occupancy and picture complexity.

The final frame bit budget $T_f(j)$ calculated in (4.6) is a weighted sum of a first target bit budget $\widehat{T}_f(j)$ which is based on the total budget $T_g(j)$, and a second target bit budget $\widetilde{T}_f(j)$ based on the status of buffer occupancy, target buffer level, and initial buffer level.

$$\widehat{T}_f(j) = \frac{w_r(j-1) \times T_g(j)}{w_r(j-1) \times N_g^{ref} + w_n(j-1) \times N_g^{nref}} \tag{4.4}$$

where $w_r(j)$ and $w_n(j)$ are the average complexity weights, N_g^{ref} and N_g^{nref} are respectively the number of left reference frames and number of left non-reference frames.

$$\widetilde{T}_f(j) = \frac{R(j)}{f} + \mu \; V_f(j) \tag{4.5}$$

where $V_f(j)$ is the virtual buffer capacity when encoding the j^{th} frame and μ is model parameter, set to 0.5 when there is no non-stored picture and to 0.25 otherwise. Therefore, using (4.4) and (4.5), we get

$$T_f(j) = \gamma \; \widehat{T}(j) + (1-\gamma)\widetilde{T}_f(j) \tag{4.6}$$

This frame budget is then bounded to maintain the quality of the decoded frame.

Basic unit level rate control

For the i^{th} MB, the number of bits left per frame $T_i'(j)$ is computed then weighted according to Equation (4.7) to obtain the MB bit budget $T_i(j)$:

$$T_i(j) = T_i'(j) \times W \tag{4.7}$$

The current distortion $d_i(j)$ is computed using the MAD of the i^{th} MB of the last stored frame $d_i(j-L-1)$ and the following prediction model:

$$d_i(j) = c_1 \; d_i(j-L-1) + c_2 \tag{4.8}$$

where c_1 and c_2 are the model parameters. The quadratic model (3.9) is then used to compute a QP. Finally, the QP is bounded by 0 and 51.

4.2 Rate Control in HEVC

This section will basically describe the rate control schemes in the HEVC standard since they are the starting point of our research. It is important to evaluate the key elements

of the controller before introducing the ROI constraint. Comparative tests are made to evaluate the performance of existing RD models in HEVC.

In the HEVC reference software two different RC algorithms have been proposed. The first one is based on a quadratic rate-distortion model and the mean absolute difference (MAD) between the original and the reconstructed signal [63] [64]. In the second algorithm, an R-λ model that takes into account the hierarchical coding structure has been adopted [54]. This model, initially introduced in version 10 of the reference software (HM.10) has been improved in a more recent version (HM.13). Adaptive bit allocation at frame level has been introduced in [65] by considering variable weights for each hierarchical level, that depend on the video content characteristics. Then, in [66], the intra frame rate control has been modified by enabling bit allocation and QP computing at CTU level. All these features have been used in our work to perform an ROI-based rate control.

4.2.1 General scheme

The different rate control algorithms proposed for HEVC have the same scheme which is illustrated in Fig. 4.1.

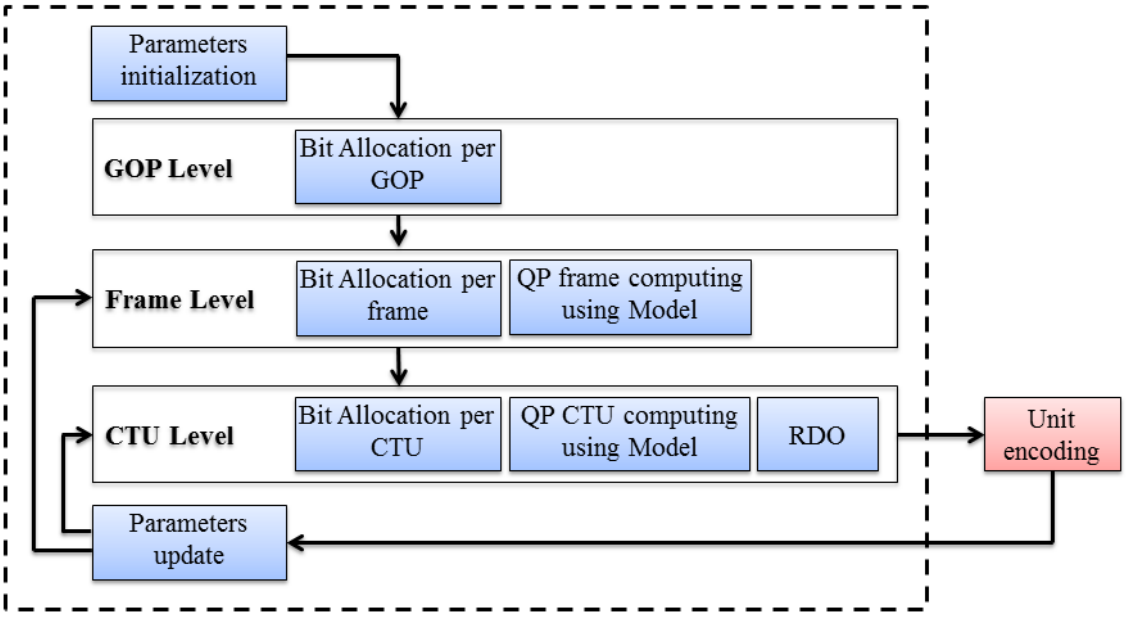


Figure 4.1: Rate control scheme for HEVC

As it can be seen, the controller operates at three main levels: GOP, frame and CTU. It performs the bit allocation at the three levels to obtain a target bit rate per unit, and, then compute an optimal QP using rate distortion models. For all propositions the global process can be described as follows:

GOP level

The input parameters are the global target bit rate (budget constraint in *bpp*), the sequence frame rate in fps, the GOP size in number of frames and the virtual buffer occupancy. The rate control algorithm uses then all these input parameters to allocate an average number of bits per GOP.

Frame level

Considering the average number of allocated bits per GOP, a target bit rate is fixed for the current frame. For B-frames, the bit allocation can introduce equal, hierarchical or adaptive weights, while for I-frames the initial budget is refined using a predefined multiplication weight. Then, the rate quantization (R-Q) model is used to compute the frame QP.

CTU level

At CTU level, the process is divided into three main steps. First, the required number of allocated bits for the CTU is computed using the frame budget, the cost of the coded CTUs of the same frame and the complexity of the CTUs. The complexity is measured using the MAD [54] or the sum of absolute transformed differences (SATD) [66]. Second, the budget is used in the R-q model to compute a QP for each CTU. The QP variation is clipped in a pre-defined range. Finally, the last step consists in finding the optimized mode decision [67] using RDO and the obtained QP. The unit is then coded and all the parameters are updated.

The particularity of each method lies in the R-Q model chosen for QP computing and the parameters used for bit allocations. The two rate control algorithms implemented in HM are detailed and compared in the following section.

4.2.2 Quadratic URQ model

Based on the quadratic R-Q model proposed for H.264/AVC and described in Section 4.1, a unified R-Q model was proposed in [63] for HEVC. It is called quadratic pixel-based unified rate-quantization (URQ) model. This model considered the new feature that the size of prediction unit varies so the bit allocation must be accordance with the number of pixels. It performs bit allocation at GOP, frame and unit levels. The units can be CTUs or group of CTUs. This algorithm has been introduced in HM.5 and improved in the future versions of the encoder [64]. The refinement of this rate control algorithm has been implemented on HM.6 and evaluated under diverse conditions.

GOP level rate control

The total bit budget in a GOP, $T_g(j)$ should be distributed according to various parameters. The bit budget allocated for the j^{th} frame of the GOP is calculated with the following

formula:

$$T_g(j) = \begin{cases} \frac{R(j)}{f} \times N_g - V_g(j), & \text{if } j = 1 \\ T_g(j-1) + \frac{R(j)-R(j-1)}{f} \times (N_g - j + 1) - T'_g(j-1), & \text{if } j = 2, 3, ..., N_g \end{cases} \tag{4.9}$$

where the available bandwidth $R(j)$ changes with a VBR encoding, N_g is the number of frames in a GOP, $V_g(j)$ is the occupancy of the virtual buffer, $T'_g(j-1)$ is the effectively generated bits of the last decoded frame in the GOP and f is the frame rate.

Frame level

At the frame level, three cases need to be considered when making bit allocation and QP computing: first frame of the GOP, reference frame and non-reference frame.

- Strategy for the first frame of the GOP:

  The QP of the initial frame of the sequence is set by the value of the initial bit rate per pixel r which is computed as follows:

  $$r = \frac{R(1)}{f \times N_f} \tag{4.10}$$

  where N_f indicates the number of pixels in the frame. The QP is then computed referring to the following table.

Condition	QP
$0.7780 < r$	12
$0.3200 < r \leq 0.7780$	17
$0.1220 < r \leq 0.3200$	22
$0.0469 < r \leq 0.1220$	27
$0.0213 < r \leq 0.0469$	32
$0.0102 < r \leq 0.0213$	37
$0.0049 < r \leq 0.0102$	42
$0.0024 < r \leq 0.0049$	47
$r \leq 0.0024$	51

Table 4.1: Initial frame QP

  The QP of the first frame of each GOP is the mean of the QPs of reference frames of the last decoded GOP. To keep smoothness, the QP value is clipped, so that it is within ± 2 of the QP of the last frame in the previous GOP.

- Strategy for reference frames:

For reference frames, the QP computing is different, as the QP is derived from a pixel-based URQ model. Frame complexity of this model is represented by the MAD. As in H.264, it is predicted using a linear model as follows:

$$\widetilde{c}(j) = a_1 \times c(j-1-M) + a_2 \tag{4.11}$$

where M is the number of non-reference frames between consecutive reference frames, $\widetilde{c}(j)$ and $c(j)$ denote the predicted complexity of the j^{th} frame and the real complexity of the last decoded reference frame computed using formula (1.7) and (a_1, a_2) are the model parameters.

As in H.264/AVC, to compute the frame budget, first $\widehat{T}_f(j)$ is calculated referring to Equation (4.4) where $w_r(j)$ and $w_n(j)$ are the average weighting factors proposed in [64]. Second, the occupancy bit budget $\tilde{T}_f(j)$ is computed using Equation (4.5). It takes into account the status of buffer occupancy $V_f(j)$, the target buffer level, and the initial buffer level. μ is model parameter set to 0.25 for random access configuration and to 0.5 for low delay configuration.

The final bit budget $T_f(j)$ is calculated as in Equation (4.6) using a weighted sum of a target bit budget $\widehat{T}_f(j)$ and an occupancy bit budget $\widetilde{T}_f(j)$. The weight γ is set to 0.6 for random access case and to 0.9 for low delay case.

Finally, the quantization step $q(j)$ of the j^{th} frame is computed using the quadratic model bellow, the QP is obtained using Equation (2.1). Then, it is clipped as done for the first frame.

$$\frac{T_f(j)}{N_f} = \alpha \frac{\tilde{c}(j)}{q(j)} + \beta \frac{\tilde{c}(j)}{q(j)^2} \tag{4.12}$$

- Strategy for non-reference frames:

  QP values of non-reference frames are derived without using buffer status. The QP of the $(j+1)^{th}$ frame is computed as follows:

$$QP(j+1) = \begin{cases} \frac{QP(j)+QP(j+2)+2}{2}, & \text{if } QP(j) \neq QP(j+2) \\ QP(j)+2, & \text{otherwise} \end{cases} \tag{4.13}$$

CTU level

At the CTU level, bit allocation is performed per unit. The initial bit budget of the first unit is exactly the allocated budget for the frame $T_f(j)$, while for the i^{th} CTU,

$$T_{init}(i) = T_{init}(i-1) - T'(i-1) \tag{4.14}$$

where $T'(i-1)$ represents real encoding bits of the $(i-1)^{th}$ unit in the frame.

Here again the final bit budget $T(i)$ allocated for a unit of order i is a weighted sum derived from two budgets. The first one $\widehat{T}(i)$ takes into the account left budget and is computed as follows:

$$\widehat{T}(i) = \frac{T_{init}(i)}{N_f(i)} N_u \tag{4.15}$$

where $N_f(i)$ and N_u denote respectively the number of pixels left in the frame after encoding the i^{th} unit and the total number of pixels in the unit.

The second bit budget depends on the buffer status $\widetilde{T}(i)$ and is computed as follows:

$$\widetilde{T}(i) = \frac{T_f(j) \times N_u}{N_f} - \frac{V(i)}{N_u(i)} \tag{4.16}$$

where $V(i)$ is the virtual buffer occupancy and $N_u(i)$ is the number of remaining units in the frame when encoding the i^{th} CTU. Thus, using (4.15) and (4.16),

$$T(i) = 0.5\ \widehat{T}(i) + 0.5\ \widetilde{T}(i) \tag{4.17}$$

This allocated budget and collocated MAD value in the previous reference frame for the unit are used to compute the quantization step as done in Equation (4.12), with N_f replaced by N_u.

4.2.3 Hyperbolic R-λ model

The latest contribution was based on an R-λ model and proposes a different bit allocation process [54]. The first version of the algorithm was implemented in HM.8. Improvements have been introduced in the later versions of HM.

GOP level

At GOP level, bit allocation takes into account the target bit rate $R_{\max}$, the frame rate f and the number of frames in a GOP N_g. The target number of bits in a GOP is determined by:

$$T_g = N_g \left(\frac{R_{\max}}{f} + \frac{(\frac{R_{\max}}{f}) \times N'_s - T'_s}{S_w} \right) \tag{4.18}$$

where the smoothing window S_w is equal to 40, N'_s is the number of pictures already encoded and T'_s is the bit cost of these pictures. The target bit rate $\frac{R_{\max}}{f}$ and the current buffer status represented by the second item $(\frac{R_{\max}}{f}) \times N'_s - T'_s$ are jointly considered in this allocation.

Frame level

In this contribution, both inter and intra picture bit allocation are supported in HEVC rate control algorithm. The allocated budget is first initialized considering picture weights and the left budget in the GOP. Then, it is refined and used for QP computing. In this algorithm, all I-frames belong to the same level. Thus, the same factor is used to refine their allocated budget, while the cost of inter pictures is determined according to different weights w_f for possible hierarchical levels [54].

- Frame budget initialization:

  The initialization step consists in calculating T_f. This bit budget is allocated per frame, using T_g computed in (4.18) and the bit cost of already coded pictures in the current GOP, T'_g,

  $$T_f = \frac{T_g - T'_g}{\sum_{i \geq f} w_i} w_f \tag{4.19}$$

- Weighted bit allocation for inter pictures:

  There are three main ways of bit allocation for inter coded frames: equal, hierarchical and adaptive allocation.

  Equal and hierarchical bit allocations have been introduced in HM.10. Equal bit allocation method considers the same weight for all inter pictures of the sequence, while hierarchical bit allocation consists in giving a predetermined weight to each frame B or P referring to its level in the GOP and the target bit rate. In a later version of HEVC test model (HM.13), adaptive bit allocation has been added to improve the model performance [65]. Using adaptive bit allocation importance weights are updated for each GOP considering the Lagrangian parameter λ computed as in Equation (3.13).

  In Table 4.2, we compare the global performance of the controller using equal, hierarchical and adaptive bit allocations. We compute the RD performance of the hierarchical method then of the adaptive one compared to equal bit allocation. The comparison is made with low delay configuration and using test sequences of class E with video-conference content [68].

	Hierarchical bit allocation			Adaptive bit allocation		
	Y	U	V	Y	U	V
Class E	-6,7%	-12,1%	-12,3%	-8,3%	-16,0%	-16,0%
Enc Time	101%			108%		
Dec Time	99%			110%		

Table 4.2: RD performance of R-λ algorithm using hierarchical and adaptive bit allocation, compared to equal bit allocation

Results show that the hierarchical and adaptive methods are slightly better then the equal bit allocation. Furthermore, the adaptive allocation gives the best performance, with 1.6% of gain compared to the hierarchical one.

- Budget refinement for intra pictures

  In the R-λ controller implemented in HM.10, the refinement is done considering a weight W that depends on the number of bits per pixel as specified in Table 4.3 [54].

Bit rate R	$R > 0.2$	$0.2 \geq R > 0.1$	$0.1 \geq R$
W	5	7	10

Table 4.3: Intra bit allocation refinement weights

If T_f is the non refined budget computed in Equation (4.19), the final allocated budget per picture T_p is then:

$$T_p = W \times T_f, \tag{4.20}$$

In HM.13, intra picture bit allocation has been improved by replacing the old refinement method by:

$$T_p = a \times \left(\frac{w_I}{T_f}\right)^b \times T_f + 0.5, \tag{4.21}$$

where $a = 0.25$, $b = 0.5582$ and w_I is the complexity measure of the frame as defined in Equation (4.32).

- Frame QP computing:

  Once the bit budget is computed, the R-λ model is used for λ and thus for the QP computing.

$$\lambda = \alpha \left(\frac{T_f}{N_f}\right)^\beta \tag{4.22}$$

To keep quality consistency, the determined λ and QP are clipped in a narrow range. They are guaranteed that:

$$\begin{aligned} &2^{-1}\lambda_l \leq \lambda \leq 2\lambda_l \\ &2^{\frac{-10}{3}}\lambda_p \leq \lambda \leq 2^{\frac{10}{3}}\lambda_p \\ &QP_l - 3 \leq QP \leq QP_l + 3 \\ &QP_p - 10 \leq QP \leq QP_p + 10 \end{aligned} \tag{4.23}$$

where (λ_l, QP_l) and (λ_p, QP_p) denote respectively λ and QP values of last decoded frame of the same hierarchical level and λ and QP values of last decoded picture.

Once the QP computed and the frame encoded, the real encoding bit cost T' and real λ value are used for model parameters (α and β) update.

$$\lambda' = \alpha \left(\frac{T'}{N_f} \right)^{\beta} \tag{4.24}$$

$$\alpha' = \alpha[1 + \delta_\alpha(\ln(\lambda) - \ln(\lambda'))] \tag{4.25}$$

$$\beta' = \beta + \delta_\beta(\ln(\lambda) - \ln(\lambda')) \ln \left(\frac{T'}{N_f} \right) \tag{4.26}$$

CTU level

Rate control at CTU level can be enabled. In HM.10, bit allocation at CTU level was performed for only inter-coded frames. All the units of intra frames have the same QP obtained at frame level. In HM.13, to better control the rate allocation of intra-coded frames rate control at CTU level was introduced.

- QP computing for inter-coded units

  The target bit budget per CTU is then determined by:

$$T_u = \frac{T_f - T_h - T'_f}{\sum_{i \geq u} w_i} w_u \tag{4.27}$$

  where T_h is the estimated bits of all headers according to previous coded picture of the same hierarchical level, T'_f is the coded bits of the current frame and w_i is the weight of the i^{th} CTU.

  The unit weight was at the beginning estimated by the prediction error (in form of MAD) of the last decoded picture of the same level as described in (1.7). However, in new versions of the reference software (such as HM.13) [69], the CTUs weight for B-frames has been modified. It depends on the model parameters α_u and β_u at CTU level, the λ of the picture and the number of pixels N:

$$w_B = N \left(\frac{\lambda}{\alpha_u} \right)^{\frac{1}{\beta_u}} \tag{4.28}$$

  Once the allocated budget per CTU is computed, the R-λ model described in (4.22) is used to compute λ and QP per CTU. T_f and N_f are respectively replaced by T_u and N_u. The model parameters are updated after encoding each CTU using the same process as in Equations (4.24) (4.25) (4.26).

  To keep quality smoothness over a frame, the determined λ and QP are clipped as

follows:

$$\begin{aligned} &2^{\frac{-1}{3}}\lambda_l \leq \lambda \leq 2^{\frac{1}{3}}\lambda_l \\ &2^{\frac{-2}{3}}\lambda_p \leq \lambda \leq 2^{\frac{2}{3}}\lambda_p \\ &QP_l - 1 \leq QP \leq QP_l + 1 \\ &QP_p - 2 \leq QP \leq QP_p + 2 \end{aligned} \tag{4.29}$$

where (λ_l, QP_l) and (λ_p, QP_p) denote respectively λ and QP values of last decoded CTU and λ and QP values of current picture.

- QP computing for intra-coded units

  The R-λ model has been modified to better control the rate allocation of intra-coded frames.

$$\lambda_u = \alpha \left(\frac{w_I}{\left(\frac{T_u}{N_u}\right)} \right)^{\beta} \tag{4.30}$$

  For an intra-coded CTU, λ_u depends on model parameters at frame level. The parameters α and β remain constant for the entire frame, however the number of allocated bits per pixel $\left(\frac{T_u}{N_u}\right)$ is computed per CTU. The complexity measure w_I for this model is calculated by deriving the sum of absolute Hadamard transformed difference (SATD) as described in [66]:

$$\mathrm{SATD} = \sum_{\mathrm{k}=0}^{7}\sum_{\ell=0}^{7} |\mathrm{h}_{\mathrm{k}\ell}| \tag{4.31}$$

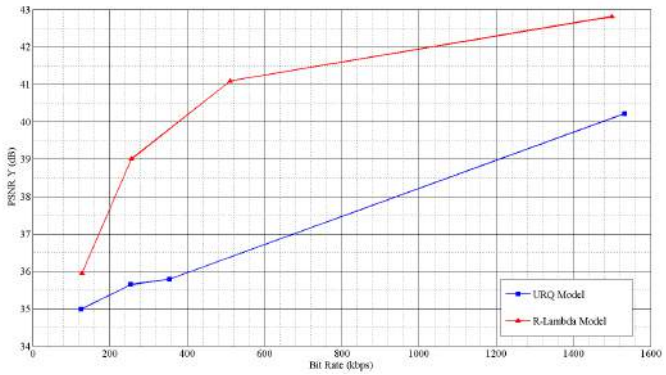
  where h_{kl} are the coefficients obtained after applying the Hadamard transform to the original 8×8 block. The weight w_I of a CTU is defined as the sum of SATD calculated for all 8×8 blocks within the CTU (N_b is the number of 8×8 units in the CTU).

$$w_I = \sum_{j=0}^{N_b-1} \mathrm{SATD(j)} \tag{4.32}$$

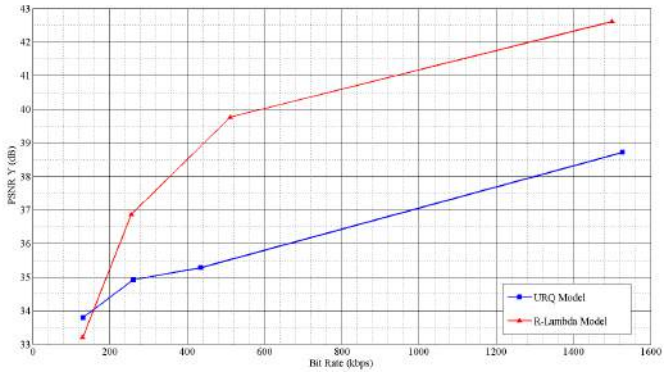
4.2.4 Comparison between URQ based and R-λ based controllers

According to experimental results given in [54], R-λ method has better RD performance and reduced bit rate error than earlier rate control algorithms in HM. We made comparative tests using our test data set (Class E sequences with video-conference content) to choose the appropriate model for our work. The obtained results confirm the document conclusions. They show that the global RD performance is improved using the R-λ model. Referring to

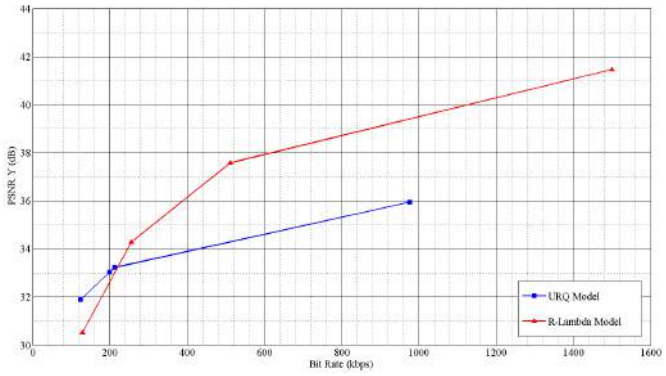
Fig.4.2 the gain goes from -22.6% to -79.6% for Class E sequences [68], using a low delay configuration with an intra period equal to 60.



(a) Johnny (BD - Rate Gain -79.6 %)



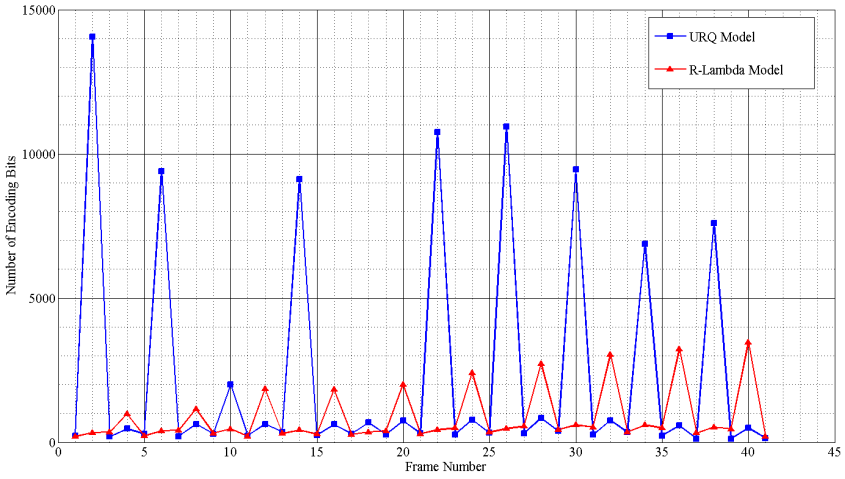
(b) KristenAndSara (BD - Rate Gain -60.8 %)



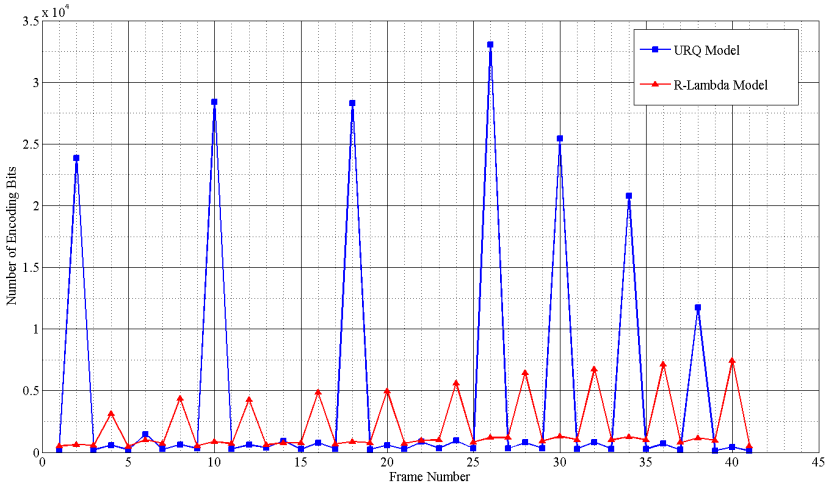
(c) FourPeople (BD - Rate Gain -22.6 %)

Figure 4.2: R-D performances of R-λ algorithm, compared URQ model

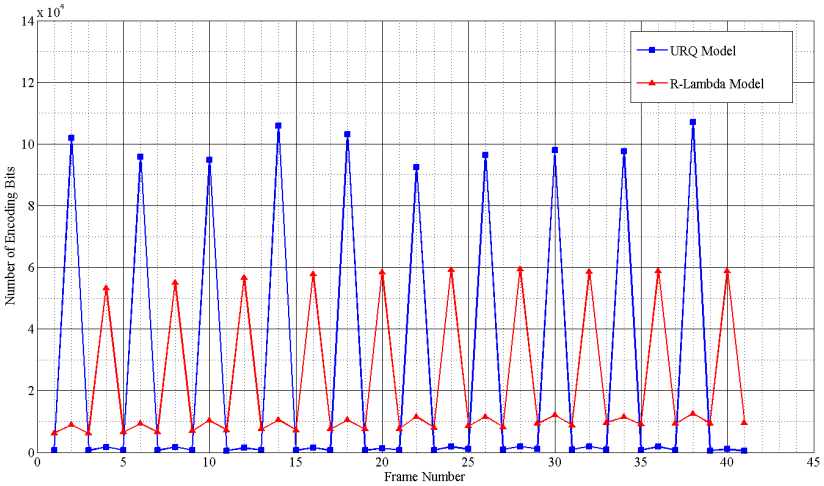
Furthermore, Fig.4.3 shows some per-frame bit cost comparing the R-λ model and the old URQ model. For example, for the test sequence "Johnny" at different bit rates, the R-λ model gives a better bit distribution over GOPs and a smoother repartition of the bit budget at frame level.



(a) 128 kbps



(b) 256 kbps



(c) 1.5 Mbps

Figure 4.3: Comparison of bit fluctuation per frame of R-λ and URQ models for sequence Johnny

4.3 Conclusion

In this chapter, we studied the evolution of different rate control algorithms such as TM5 for MPEG-2, VM8 for MPEG-4, TMN8 for H.263 and the quadratic model for H.264/AVC in Section 4.1. Although HEVC still adopts the traditional hybrid coding framework, in which many rate control algorithms have been proposed before, yet these rate control schemes cannot work well if they are applied directly to HEVC without considering the new coding tools. Thus, new rate control models have been proposed for HEVC and presented in Section 4.2.

In the next chapter, we study different rate-distortion models from the literature. Then, we propose new models based on signal characteristics that can be used to perform efficient bit allocation over regions. Moreover, existing controllers evaluated in this chapter are exploited to perfom ROI-based rate control in Part III of this thesis.

Chapter 5

Rate-Distortion models for HEVC

Contents

Accurate rate-distortion modeling at different levels of the encoder (frame and CTU levels) plays an important role in optimal bit allocation for HEVC coding. Due to the different characteristics of frames, as well as the evolution of compression techniques, several analytic RD models have been proposed and validated for bit allocation at frame level

in hybrid video coding. These models can be appropriate for ROI-based rate control for HEVC.

In this chapter, we study some models from the literature. The first section introduces two useful RD models studied in [70] for H.264/AVC: $1/R$ model and exponential model. Then, we propose an extended version of the exponential model for intra-coded frames at CTU level. Proposed solution and performed experiments are detailed for model validation for HEVC. In a second part of the section, we provide a study on RD modeling for HEVC considering appropriate probabilistic models for the transform coefficients. We derive models based on the signal characteristics for both intra- and inter-coded frames and at low and high bit rates. Comparatives tests conclude this chapter to show that a good fitting of the transform coefficient distribution give efficient RD models.

5.1 Validation of models from the literature for HEVC

5.1.1 Glossary of models used at frame level

$1/R$ model

In [71], an analytical framework for frame-level dependent bit allocation is proposed. Two different RD models for both inter and intra frames have been designed. In our work, we use the intra frame RD function that takes into account the average gradient G of the frame, its encoding bit rate R and the corresponding distortion D measured by the mean square error (MSE) between the original frame and the reconstructed one:

$$D = \frac{a\ G}{R - c\ G} - b. \tag{5.1}$$

where (a, b, c) are model parameters given in [72]. This model makes sens when $(R - c\,G) > 0$. In this work, we evaluated the fitting performances of this model at CTU level when the previous condition is respected.

Exponential models

According to the classical R-D theory [73], at frame level and for high bit rate, the relationship between R and D can be expressed as following:

$$D = \alpha\sigma^2 2^{-2R} \tag{5.2}$$

where α and σ^2 are, respectively, the PDF shape factor and the variance of the residual DCT coefficients. D is measured as the MSE between the original frame and the reconstructed one. However, such type of analytic formula may not be accurate for hybrid video coders and may be replaced by a similar R-D model studied in the team past work [70] or spline based models introduced in [74] [75].

The RD model in Equation (5.2) has been adapted for optimal bit allocation in hybrid video coding. Again both I-frames and P-frames have been considered, but, we focus on intra-frame rate allocation only and employ the following RD function:

$$D = \alpha\, \sigma^2\, 2^{-\beta\, R}, \tag{5.3}$$

where σ^2 is the variance of the coded I-frame and (α, β) are model parameters estimated as explained later. This model has shown better fitting performances at frame level than the previous ones [70]. Moreover, it has been proved in [76] that this exponential model can be used, at high bit rate, for sparse sources. Thus, we propose to use it at CTU level, adopt it to HEVC intra-coded units and compare its performance with model in Equation (5.1).

5.1.2 Proposed extension at CTU level

To make comparative tests between model (5.1) and model (5.3), both RD functions are extended at CTU level. The considered distortion D_i is the MSE between the i-th original CTU and the reconstructed one and R_i is the corresponding bit rate.

For (5.1), we use the parameters computed in [72] and the average gradient G of each coded CTU. Then, we verify if this function describes the relationship between R and D of each CTU. Moreover, the model in Equation (5.3) is modified to perform bit allocation at CTU level. For each CTU of index i, the variance σ_i^2 is computed, the distortion D_i and bit rate R_i are measured. Then, a fitting algorithm is used to generate the parameters (α_i, β_i) and model the required relationship between the rate and the distortion at CTU level:

$$\widetilde{D}_i = \alpha_i\, \sigma_i^2\, 2^{-\beta_i\, R_i}, \tag{5.4}$$

where $\widetilde{D}_i$ is the estimated distortion.

5.1.3 Validation process

Implementation and test condition

Our experiments consist in comparing the two RD models described before at CTU level. We used all the test sequences from class A to E [68]. We encode two frames per sequence with an all intra configuration and a CTU size that goes from 16x16, 32x32 to 64x64. We use independently decodable CTUs. Each CTU corresponds to one slice and in-loop filtering is deactivated across slice boundaries. Consequently, only spatial dependencies inside a CTU are considered while encoding the unit.

Moreover, we estimated the model parameters (α_i, β_i) of Equation (5.4) and the variance σ_i^2 of each CTU by encoding the sequences using HEVC test model (HM.16) [77]. The rate control disabled, in order to manually fix different quantization parameters: $1, 2, 3, \ldots, 50$.

For each CTU, we recorded the values of D_i and R_i produced at the encoder output. We estimated the model parameters and compute the new distortion.

Finally, a fitting function is used to verify if each model well describes the relationship between CTU encoding bit rate R_i and the corresponding distortion $\widetilde{D}_i$.

Evaluation metric

As said before our experimental analysis consists in comparing the proposed RD model with the one in [72]. The validation was performed on the basis of the ρ metric, defined as:

$$\rho = 1 - \frac{\sum_i (X_i - \hat{X}_i)^2}{\sum_i (X_i - \bar{X})^2} \tag{5.5}$$

where X_i and $\hat{X}_i$ are the real and the estimated values of one data point, and $\bar{X}$ is the mean of all data points. ρ was designed to quantitatively measure the degree of deviation from a given model [78]. The closer the value of ρ is to 1, the more accurate is the model.

5.1.4 Validation results

Table 5.1 shows the ρ values associated to the RD function of intra coded CTUs given in both $1/R$ model traduced by equation (5.1) and Exp model defined by the RD Equation (5.3). For all the tested sequences, the proposed Exp model shows superior fitting performance, giving ρ values very close to 1 and higher than ρ values given by $1/R$ model.

The fitting performance are presented for different CTU size and at high bit rate coding. We also notice that the fitting is better for bigger CTUs in most of the sequences; the average of ρ is equal to 0.8459 when the CTU size is 64x64 and it is reduced to 0.7652 when the CTU size is equal to 16x16. This is due to the fact that statistical models are less efficient when the data set is reduced.

A second table is given to summarize the fitting performance of the two models at lower bit rates. Table 5.2 shows that ρ values decrease for both RD models at low bit rate. However, Exp model is still having better fitting performance than the $1/R$ model for QPs from 26 to 50. Moreover, as explained previously, to have a valid estimation of the distortion using Equation (5.1) the real bit rate R should be higher than a certain limit $(R - c\,G) > 0$. Experiments have shown that $1/R$ model is not valid for many CTUs. This condition cannot be verified for 60% of the cases.

To better evaluate the fitting performances, Fig. 5.1 reports the RD curve per CTU for sequences from different classes. The RD curves represented in the figures are given by:

- HEVC encoding: the black line represent the real output values of the encoder R_i and D_i of a CTU of index i.
- $1/R$ model: the blue line report real bit rates R_i and distortion generated using (5.1).

class	*Sequence*	CTU 64x64		CTU 32x32		CTU 16x16	
		Exp model	$1/R$ model	Exp model	$1/R$ model	Exp model	$1/R$ model
A	Nebuta Festival	0.9269	0.5836	0.8878	0.5410	0.8106	0.4955
	People On-Street	0.8897	0.7096	0.8140	0.6340	0.7158	0.4637
	Steam LocomotiveTrain	0.8845	0.5429	0.8347	0.5248	0.7376	0.5113
	Traffic	0.8580	0.6669	0.8000	0.5473	0.7168	0.3830
Average class A		**0.8898**	**0.6258**	**0.8341**	**0.5618**	**0.7452**	**0.4634**
B	Basketball Drive	0.9148	0.5894	0.8974	0.5398	0.8387	0.3039
	BQTerrace	0.8341	0.5385	0.7931	0.4683	0.7121	0.4208
	Cactus	0.8590	0.4858	0.8828	0.4278	0.8290	0.3009
	Kimono1	0.8378	0.7037	0.7891	0.6008	0.7145	0.3571
	ParkScene	0.8885	0.4841	0.8790	0.3991	0.8263	0.2765
Average class B		**0.8668**	**0.5603**	**0.8483**	**0.4872**	**0.7841**	**0.3318**
C	Basketball Drill	0.9118	0.6155	0.8843	0.3854	0.8125	0.2509
	BQMall	0.6183	0.5396	0.7762	0.4595	0.7724	0.3632
	PartyScene	0.8760	0.5617	0.9117	0.4191	0.8244	0.3198
	RaceHorses	0.8841	0.5894	0.8286	0.4236	0.7168	0.3114
Average class C		**0.8226**	**0.5766**	**0.8502**	**0.4219**	**0.7815**	**0.3113**
D	Basketball Pass	0.8678	0.6635	0.7916	0.4846	0.7126	0.3105
	Blowing Bubbles	0.9480	0.4467	0.9026	0.2840	0.8213	0.2402
	BQSquare	0.9407	0.6234	0.9014	0.4836	0.7808	0.3821
	RaceHorses	0.9194	0.5825	0.8613	0.4319	0.7447	0.3057
Average class D		**0.9190**	**0.5790**	**0.8642**	**0.4210**	**0.7649**	**0.3096**
E	Four People	0.8046	0.6936	0.8126	0.6124	0.7426	0.4491
	Johnny	0.7061	0.7129	0.8035	0.5361	0.7489	0.3757
	Kristen &Sara	0.6840	0.5102	0.7832	0.4681	0.7594	0.3164
Average class E		**0.7316**	**0.6389**	**0.7998**	**0.5389**	**0.7503**	**0.3804**
Average all sequences		**0.8459**	**0.5961**	**0.8393**	**0.4861**	**0.7652**	**0.3593**

Table 5.1: ρ values of the R-D functions at CTU-level for QP $\in [1, 25]$

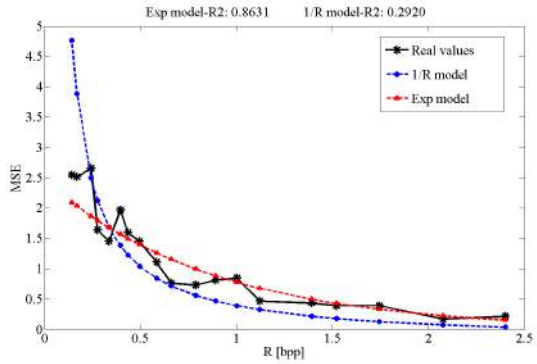
class	QP $\in [26, 39]$		QP $\in [40, 50]$	
	Exp model	$1/R$ model	Exp model	$1/R$ model
A	0.7589	0.5490	0.5282	0.4434
B	0.6150	0.4152	0.4652	0.4941
C	0.7829	0.4088	0.5626	0.1469
D	0.7842	0.3457	0.5819	0.2577
E	0.4686	0.5305	0.4071	0.0992

Table 5.2: ρ values of the R-D functions at CTU-level for different bit rate levels and for CTU size equal to 64x64

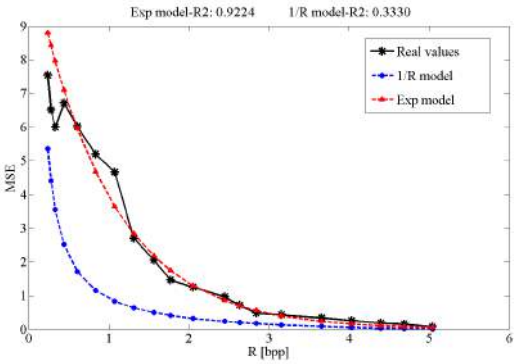
- Exp model: the red line describes the relationship between real bit rates R_i and estimated distortion $\widetilde{D}_i$ of Equation (5.4).

These results show that the rate distortion relationship achieved with the proposed exponential method is close to real values given by HEVC encoder. The results prove that the proposed RD model with appropriately chosen parameters can accurately describe the relationship between the encoded bits and the corresponding distortion for independently

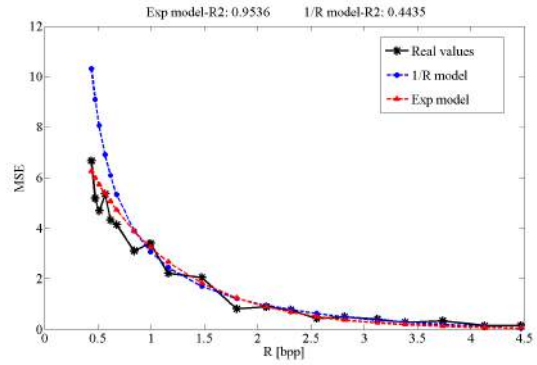
decodable CTUs.



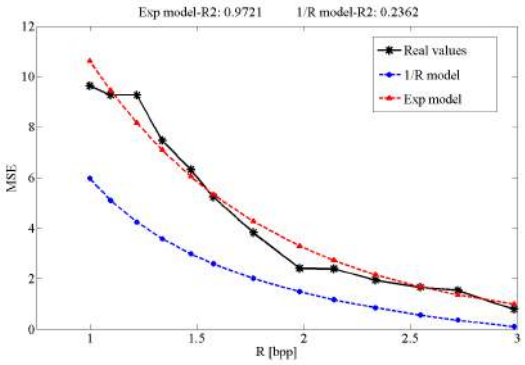
(a) PeopleOnStreet CTU 3 (Class A)



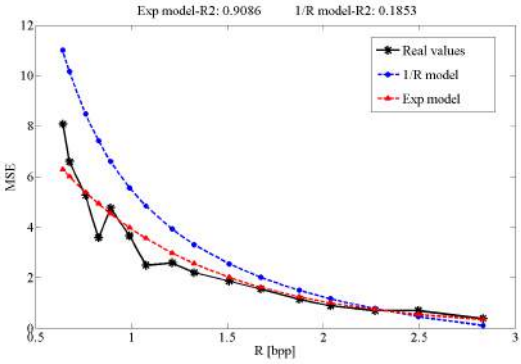
(b) Cactus CTU 4 (Class B)



(c) BQMall CTU 4 (Class C)



(d) BasketBallPass CTU 5 (Class D)



(e) FourPeople CTU 3 (Class E)

Figure 5.1: Model fitting at CTU level for QP $\in [1, 25]$ and CTU size equal to 64x64

5.2 Study on statistical distribution of HEVC transform coefficients

5.2.1 Probabilistic distributions

It is important to have a clean and precise probability model that can sufficiently describe images. Different statistical distributions are studied and parameters of the probability density function are estimated by maximizing the likelihood of the data under the model. Normal, Laplacian, Generalized Gaussian (GG) and Bernoulli-GG (BGG) distributions have been studied and their parameters have been estimated to find the best fitting. The GG probability density function (PDF) is given by [79],

$$\forall \xi \in \mathbb{R}, f(\xi) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} e^{(\frac{|\xi|}{\alpha})^{\beta}} \tag{5.6}$$

where $\Gamma(.)$ is the gamma function,i.e, $\Gamma(z) = \int_0^{\inf} e^{-t}t^{z-1}dt, z > 0$. Here α models the width of the PDF peak (standard deviation), while β is inversely proportional to the decreasing rate of the peak. Sometimes, α is referred to as the scale parameter while β is called the shape parameter.

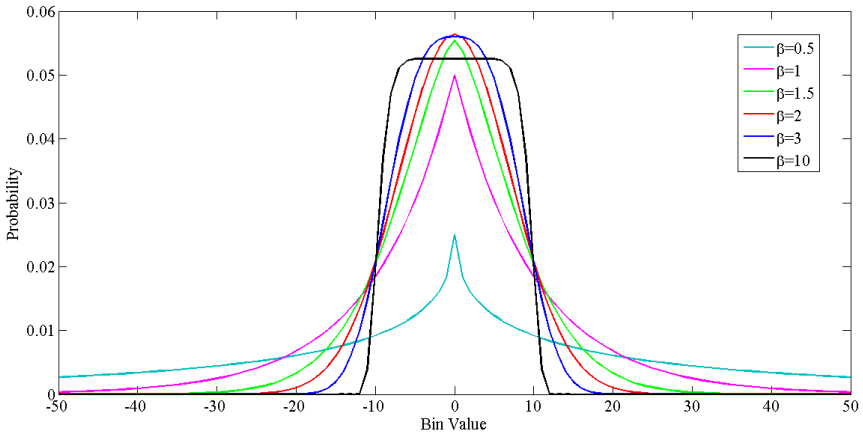


Figure 5.2: Probability density function for different β

The GG model contains the Normal and Laplacian PDFs as special cases, using respectively $\beta = 2$ and $\beta = 1$. As represented in Fig.5.2, this family allows for tails that are either heavier than normal (when $\beta < 2$) or lighter than normal (when $\beta > 2$). When $\beta \to \inf$, the density converges to a uniform density.

In addition, the differential entropy of this distribution can be written as follows:

$$h_{\beta}(\omega) = \ln\left(\frac{2\Gamma(\frac{1}{\beta})}{\beta\omega^{1/\beta}}\right) + \frac{1}{\beta} \tag{5.7}$$

When the data to be modeled is sparser a Bernoulli-GG can be adopted. Considering a

mixture parameter $\epsilon \in [0, 1]$ the BGG distribution is defined as a combination of a Dirac distribution δ (i.e., point mass at 0) and a GG distribution f:

$$\forall \xi \in \mathbb{R}, g(\xi) = (1 - \epsilon)\delta(\xi) + \epsilon f(\xi) \tag{5.8}$$

5.2.2 HEVC transform coefficient distribution

The transform designed for HEVC needed to be efficient in both software and hardware versions of HEVC. Thus, an integer approximation of DCT and DST transforms are implemented in the encoder and a clipping is added to deal with hardware limitations. Consequently, the obtained coefficients' distribution is discrete but can be fitted by a continuous distribution. To select the appropriate probabilistic models, we carried out a study on transform coefficients distribution X of HEVC. In fact, we studied histograms of non-quantized transform coefficient per transform level at frame level and per unit at CTU level. The distribution has been evaluated for both intra- and inter-coded frames. The impact of TU partitioning, prediction type and QP are then evaluated.

Impact of transform level

The transform unit (TU) is a square region of size 4×4, 8×8, 16×16 or 32×32 luma samples/pixels defined by a quadtree partitioning of a leaf CU as represented in Fig.2.3 of Chapter 2. The quadtree partitioning of the CU into one or more TUs is known as a residual quadtree (RQT). In general, each TU is associated with a partitioning depth and a transform matrix. Thus, we studied 4 transform levels ; level 0, leve 1, level 2 and level 3 that corresponds respectively to transform size 32×32, 16×16, 8×8 and 4×4.

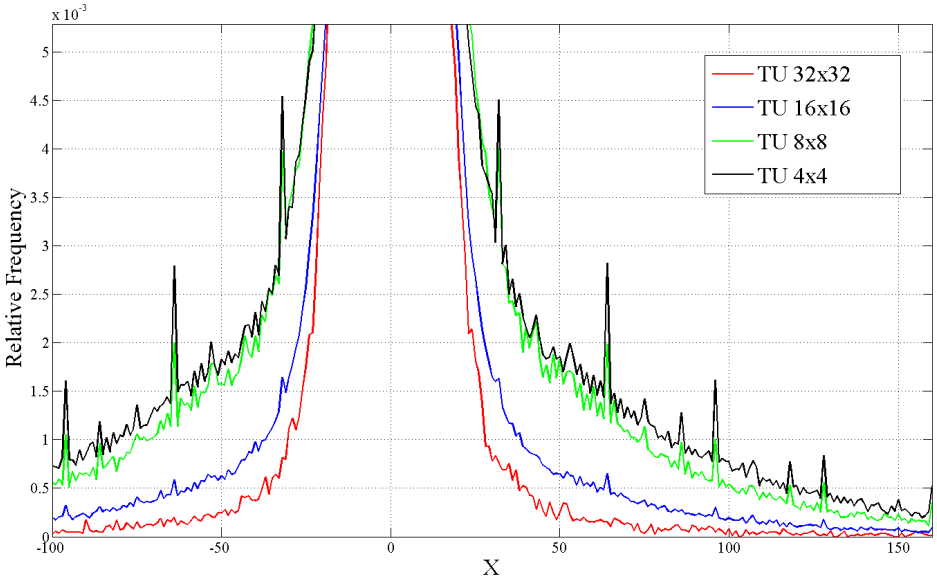
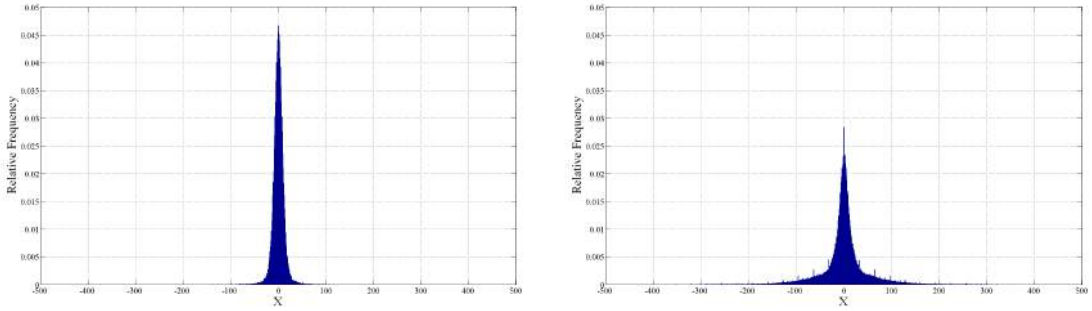


Figure 5.3: Comparison of distributions of different level of transform

Fig.5.3 represents the histograms of transform coefficients for different sizes of TU. It is shown that the smaller is the size of the TU, the shorter and the wider is the coefficient distribution. In fact, big units (for example of size 32×32) corresponds to smooth regions, so residuals are around 0 which explain the fact that the distribution is tall and narrow.

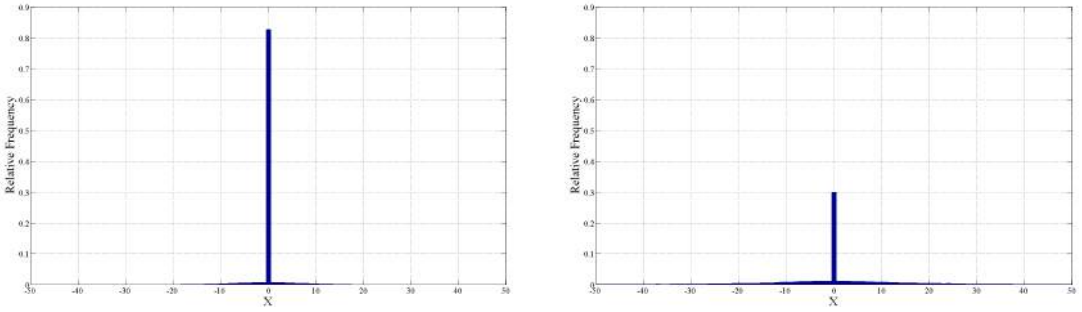
Impact of prediction type

In HEVC, both the intra and the inter coded pictures use predictive coding. However, for intra pictures, only a spatial prediction using neighboring pixels of the previously encoded unit is used for a given CTU to be encoded. The inter prediction uses pixels from previously encoded pictures and has much greater prediction ability compared to the intra prediction.



(a) Level 0 (b) Level 3

Figure 5.4: Transform coefficients' histograms for different transform levels of an I-frame at QP=22



(a) Level 0 (b) Level 3

Figure 5.5: Transform coefficients' histograms for different transform levels of a B-frame at QP=22

In fact, Fig.5.4 and Fig.5.5 shows histograms of transform coefficients at 2 different levels for both I-frames and B-frames. These normalized histograms are obtained after encoding the sequence "BaseketBallDrive" at a $QP = 22$. We notice that they have different characteristics when a given video source is intra-coded versus inter-coded. In fact, for I-frames the distribution of transformed coefficients have the shape of a Generalized Gaussian distribution. For B-frames it is more like a BGG distribution as it is a combination of a dirac and a Gaussian. This is due to the nature of intra and inter prediction residual.

Thanks to temporal redundancy between successive frames, inter prediction tends to reduce considerably the cost of the residual. Transform coefficients are then concentrated in 0. This analysis can be confirmed by evaluating the percentage of zero coefficients per unit.

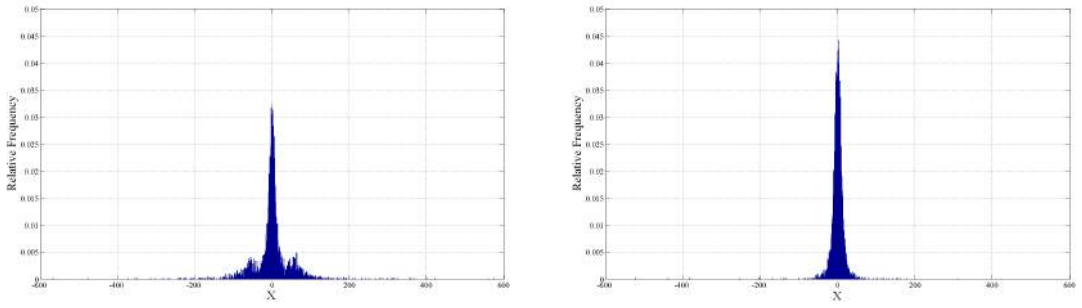
Experiments have shown that the percentage of zero coefficients per unit is important in inter coded frames. Referring to Table.5.3 the zero coefficients probability can go up to 13.92% for "ParkScene" coded in inter prediction mode. However, it is only 1.3% when encoding the sequence in intra mode. Consequently, for I-frame the percentage of zero coefficient is not very relevant. $(1-\epsilon)$ is very small so the transform coefficient distribution can be represented by a GG. While, for B- and P-frames $(1-\epsilon)$ is bigger so the transform coefficient distribution should be represented by a BGG as described in Equation (5.8).

Sequence	Intra prediction	Inter prediction
"BasketBallDrive"	2.56%	9.27%
"BQTerrace"	4.42%	13.59%
"Cactus"	1.12%	9.93%
"Kimono1"	2.17%	11.60%
"ParkScene"	1.30%	13.92%

Table 5.3: Percentage of zero coefficients for different prediction types at QP=1

Impact of quantization parameter

The study of the impact of the QP on transform coefficients has been done at CTU level. The quantization parameter selection has an indirect effect on the non-quantized DCT coefficient distribution via the predictive coding. The reconstructed picture distortion is proportional to the quantization level used while encoding it. It represents less details and smoother texture. Considering the fact that inter and intra prediction is done referring to compressed units of the picture, the prediction produce a smaller residue when the reference picture is of low quality (i.e. with a bigger QP).



(a) QP=10 (b) QP=27

Figure 5.6: Impact of QP on transform coefficients' histograms of and intra coded CTUs

Fig.5.6 confirms our hypothesis. It illustrates two histograms of DCT-transformed coefficients at two different quantization levels. It shows that DCT coefficients concentrated around zero statistically when the QP is bigger. Moreover, the tail of the DCT coefficient

distribution is lighter than when the reference is encoded with a smaller QP. Thus, in our experiments, we managed to reduce the impact of quantization in transform coefficients distribution by setting the QP to 1.

5.2.3 Transform coefficients modeling

The modeling of the distribution of transformed coefficients in HEVC are performed using previously detailed distributions; Normal, Laplacian, GG and BGG. The transformed coefficients are derived after mode decision. Modeling are performed in two steps. Firstly, the coefficients from the whole frame are collected. Secondly, they are grouped per CTU. Then modeling is performed on each group. The model representing the best fitting is considered and its parameters are estimated.

Fitting evaluation metric

All studied distributions are tested for transform coefficient modeling. To evaluate fitting performance and select the appropriate probabilistic distribution, we use a root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{(\hat{y_i} - y_i)^2}{(y_i)^2}} \tag{5.9}$$

The distribution giving the smallest RMS is the one that better models the transform coefficient distribution. This evaluation has been done at both frame level and CTU level.

Generalized Gaussian fitting for intra-coded units

As explained before, intra-coded units have a short and wide distribution. Experiments have shown that the percentage of 0 coefficients does not exceed 5% of the distribution (Table.5.3). To select the appropriate probabilistic model that fits the coefficient distribution, Laplacian, Normal and Generalized Gaussian models are tested.

Fig.5.7 shows an example of modelling the distribution of the DCT coeffcients from the first frame of the sequence "BasketBallDrive" under an all intra configuration using the three distributions. We notice that the GG distribution gives the best fitting approximation as it suits the data well and gives the smallest RMS comparing to Normal and Laplacian distributions. In fact, experiments show that a good PDF approximation for HEVC transform coefficients of an intra-coded unit can be achieved by adaptively varying two parameters (α, β) of the Generalized Gaussian density defined in Equation (5.6), the GG distribution is used in our work when intra prediction is performed. This is useful to estimate the distribution parameters per CTU of I-frames and perform rate distortion optimization per unit. To evaluate the GG fitting, ClassB sequences has been coded in all intra mode.

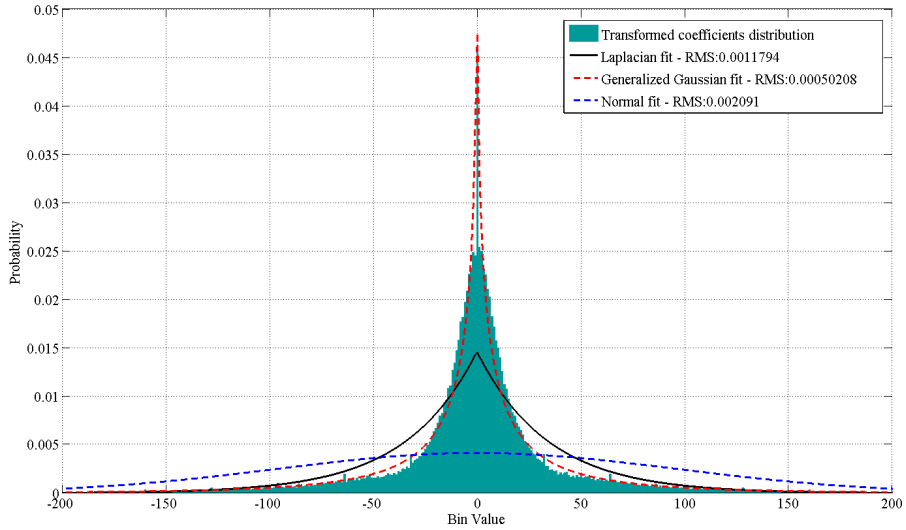
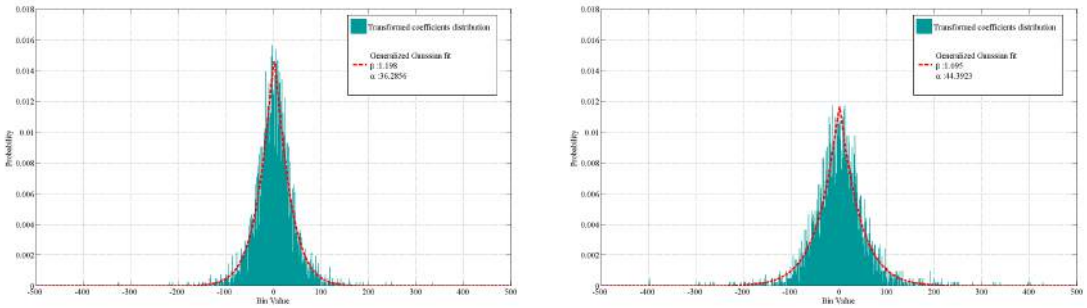


Figure 5.7: Example of transform coefficients histogram of "BasketBallDrive" intra-coded frame fitted with Normal, Laplacian and GG densities

Models	"BasketBallDrive"	"BQTerrace"	"Cactus"	"Kimono1"	"ParkScene"
Normal	16	22	9.86	14	11
Laplacian	11	18	6.43	7.77	7.10
GG	6.93	15	6.34	7.54	6.45

Table 5.4: RMSE $\times 10^4$ of tested distributions

Table 5.4 gives the obtained RMS after fitting different probabilistic models to all transform coefficients obtained after encoding all units of all ClassB sequences in intra mode. We can see that the GG distribution has the smallest error for all sequences. Consequently, it gives the best fitting. If we represent the obtained modeling of these transform coefficients using the GG distribution as in Fig.5.8, we notice that the GG model gives a good representation of non-quantized transform coefficients for different textures.



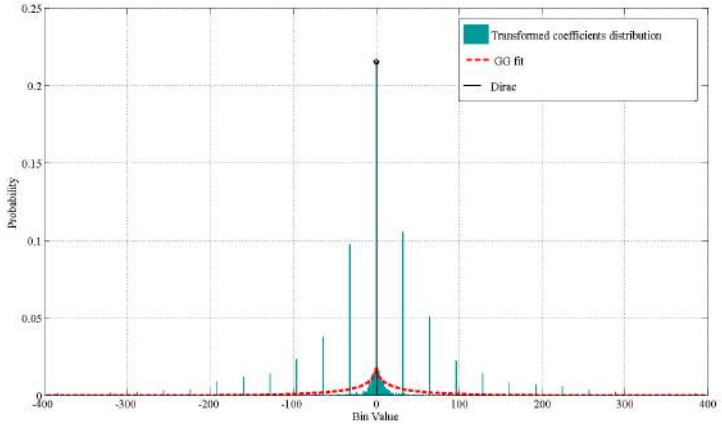
(a) "ParkScene" (b) "Cactus"

Figure 5.8: GG fitting of residual of different intra-coded CTUs

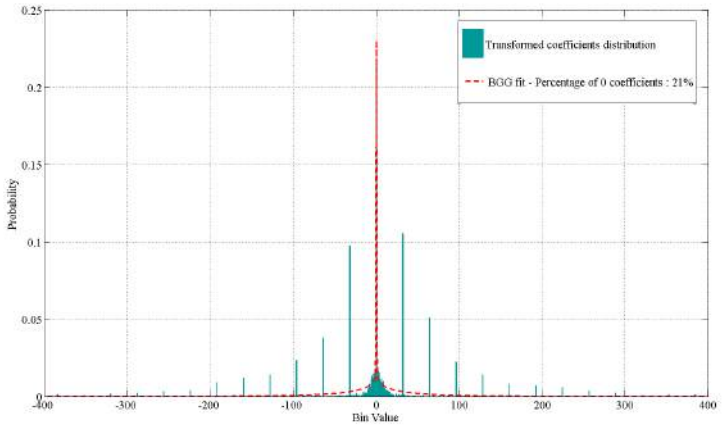
Bernouilli Generalized Gaussian fitting for inter-coded units

Inter-coded units represent a transform coefficient distribution tall and narrow with an important number zero coefficients. In that case, the percentage of 0 coefficients exceeds 5% of the distribution. Thus, the best fitting of this data is a BGG model. The final model is then a sum of a Dirac and a GG as explained in Equation (5.8).

The fitted distribution has been obtained after encoding in inter mode a CTU using a quantization parameter equal to 1 (Here again we need to minimize the quantization effect when performing motion estimation referring to the appropriate decoded unit). Fig.5.9(a) shows the fitted Dirac and GG distributions for respectively zero coefficients and non-zero coefficients. The combination of both distributions gives a BGG model that perfectly fits the studied data as represented in the example given in Fig.5.9(b).



(a) Dirac + GG



(b) Final BGG

Figure 5.9: BGG fitting of residual of different inter-coded CTUs

However, as show in Fig.5.9, the histogram represents some regular pics that appears because a shifting and clipping step is preformed by HEVC during the transform process. The residual obtained after inter prediction is more sensitive to this shifting.

5.3 Proposed operational rate-distortion modeling for HEVC

At this stage of our researches, we derive operational rate-distortion functions at high and low bit rates considering signal characteristics. This has been done in the literature such as in [80], where a simplified rate model for texture has been derived based on a Laplacian PDF model. In our work, new rate and distortion models are derived [76] for both inter and intra coded units and using a more accurate fitting based on a BGG distribution.

5.3.1 Rate distortion models parameters

For inter- and intra-coded units and at low and high resolutions, asymptotic expressions of the distortion for a p^{th}-order moment error measure and close approximations of the entropy are provided considering the source characteristics. Consequently, for an accurate representation of the rate and the distortion, parameters of the source distribution are introduced as shown in [76]. These RD models are adapted to HEVC coding and simplified versions are proposed in our work. Here is a summery of used parameters and notations:

- p : the order of the moment error measure is equal to 2 is our cas as MSE is used as a distortion metric.
- ζ : the offset parameter indicating the shift of the reconstruction is equal to 0 as a midpoint reconstruction is performed in HEVC.
- ν : is equal to 1/4 in our experiments.
- ϵ : the mixture parameter defined in Equation (5.8). It is fixed at 1 when using a GG distribution in intra mode. However, it should be computed per CTU in inter model to estimate the appropriate parameters of the BGG distribution. It corresponds to the percentage of non zero coefficients.
- H_ϵ : the entropy of a Bernoulli can be written as $H_\epsilon = -\epsilon \ln \epsilon - (1-\epsilon)\ln(1-\epsilon)$ it is equal to 0 when $\epsilon = 1$ (Intra case).
- μ_p : is written as $\mu_p = \frac{\Gamma(\frac{p+1}{\beta})}{\Gamma(\frac{1}{\beta})}$ and for $p = 2$, $\mu_2 = \frac{\Gamma(\frac{3}{\beta})}{\Gamma(\frac{1}{\beta})}$.
- β : is the exponent parameter of the model (shape) defined in Equation (5.6) . It should be estimated in our experiments.
- α : is a model parameter representing the scale defined in Equation (5.6). It should be estimated in our experiments.
- ω : the scaling factor is $\omega = \alpha^{-\beta}$.
- q : represents the quantization step-size.
- $\bar{q}$: is the normalized quantization step-size $\bar{q} = q/\alpha$.

- $\tilde{q}$: is defined as $\tilde{q} = (2^{-\beta}\bar{q}^{\beta} + 1 - 1/\beta)^{1/\beta} = ((\frac{q}{2\alpha})^{\beta} + 1 - 1/\beta)^{1/\beta}$.

5.3.2 Proposed rate distortion models for intra-coded units

For frames of type I, the GG model is used to fit the transform coefficient distribution per CTU and (α, β) are estimated. These parameters are introduced in the proposed RD model for both low and high resolutions.

Low bit rate

- The R-q model at low bit rate is a function of the normalized quantization stepsize $\bar{q}$ and distribution model parameters. It can be represented as follows :

$$\bar{d}_{p,\zeta}(\bar{q}) = \mu_p - \frac{\bar{q}^{p+1}e^{-\bar{q}^{\beta}/2^{\beta}}}{2^{p+1}\Gamma(\frac{1}{\beta})\tilde{q}^{\beta}} \times (1 - (1+2\zeta)^{p} + \frac{p}{\beta\tilde{q}^{\beta}} \times (1 + (1+2\zeta)^{p-1})) \tag{5.10}$$

  Considering midpoint reconstruction in HEVC quantization ζ is set to 0 and MSE for error measure p is set to 2, the equation can be then simplified:

$$\bar{d}_{2,0}(\bar{q}) = \mu_2 - \frac{\bar{q}^{3}e^{-\bar{q}^{\beta}/2^{\beta}}}{2\beta\Gamma(\frac{1}{\beta})\tilde{q}^{2\beta}} \tag{5.11}$$

  The final distortion model $\bar{d}(q)$ at low bit rate is a function of the quantization stepsize and can be written as follows:

$$\bar{d}(q) = \mu_2 - \frac{(\frac{q}{\alpha})^{3}e^{-(\frac{q}{2\alpha})^{\beta}}}{2\beta\Gamma(\frac{1}{\beta})\tilde{q}^{2\beta}} \tag{5.12}$$

  knowing that $\tilde{q}^{\beta} = 2^{-\beta}\bar{q}^{\beta} + 1 - 1/\beta = (\frac{q}{2\alpha})^{\beta} + 1 - 1/\beta$, we get

$$\bar{d}(q) = \frac{\Gamma(\frac{3}{\beta})}{\Gamma(\frac{1}{\beta})} - \frac{(\frac{q}{\alpha})^{3}e^{-(\frac{q}{2\alpha})^{\beta}}}{2\beta\Gamma(\frac{1}{\beta})((\frac{q}{2\alpha})^{\beta} + 1 - 1/\beta)^{2}} \tag{5.13}$$

  It is important to note that $\bar{d}(q)$ should be positive.

- The R-q model at low bit rate is:

$$R_p(\epsilon, \bar{q}) = \epsilon\frac{\beta\bar{q}^{2\beta-p}}{2^{\beta-p+1}p} \tag{5.14}$$

  It also depends on the normalized quantization step and GG distribution parameters. Considering $p = 2$ and $\epsilon = 1$, the equation can be then simplified:

$$R_2(1, \bar{q}) = \frac{\beta\bar{q}^{2\beta-2}}{2^{\beta}} \tag{5.15}$$

Finally,

$$R(q) = \frac{\beta(\frac{q}{\alpha})^{2\beta-2}}{2^{\beta}} \tag{5.16}$$

High bit rate

- The D-q model at high bit rate is:

$$\bar{d}_{p,\zeta}(\bar{q}) = \frac{\nu \bar{q}^{p}}{p+1} \tag{5.17}$$

For the chosen configuration of high efficiency video coding, we replace p, ζ and ϵ by respectively 2, 0 and 1, we get:

$$\bar{d}_{2,0}(\bar{q}) = \frac{\bar{q}^{2}}{12} \tag{5.18}$$

Finally as the normalized quantization stepsize is $\bar{q} = \frac{q}{\alpha}$,

$$\bar{d}_{(}q) = \frac{1}{12} \times (\frac{q}{\alpha})^{2} \tag{5.19}$$

- The proposed R-D model at high bit rate is a logarithmic function of $\bar{d}$:

$$R_p(\epsilon, \bar{d}) = H_\epsilon + \epsilon \left(h_\beta(1) - h_p(1) - \frac{1}{p} \ln(\frac{p\bar{d}}{\epsilon}) \right) \tag{5.20}$$

As shown before, the mixture parameter $\epsilon = 1$ as we are considering a GG distribution. Consequently, the entropy $H_\epsilon = 0$.

$$R_2(1, \bar{d}) = \left(h_\beta(1) - h_2(1) - \frac{1}{2} \ln(2\bar{d}) \right) \tag{5.21}$$

To obtain the needed R-q model, we replace $\bar{d}$ from Equation (5.19). Finally, the rate model can be written at high bit rate as following:

$$R(q) = \left(h_\beta(1) - h_2(1) - \frac{1}{2} \ln \left(\frac{1}{6} (\frac{q}{\alpha})^2 \right) \right) \tag{5.22}$$

5.3.3 Proposed rate distortion models for inter-coded units

For P and B frames, the BGG distribution is used, three parameters need to be estimated $(\alpha, \beta, \epsilon)$. The same distortion models as intra coded frames are used at both low bit rate (Equation (5.13)) and high bit rate (Equation (5.19)). However, different rate models are introduced taking into account the mixture parameter ϵ.

Low bit rate

The R-q model at low bit rate defined in Equation (5.14) is used and can be simplified as follows:

$$R(\epsilon, q) = \epsilon \frac{\beta(\frac{q}{\alpha})^{2\beta-2}}{2^{\beta}} \tag{5.23}$$

High bit rate

The R-q model at high bit rate is derived from Equation (5.20) by replacing the distortion $\bar{d}$ by Equation (5.19). Thus, it can be written as:

$$R(\epsilon, q) = H_{\epsilon} + \epsilon \left(h_{\beta}(1) - h_2(1) - \frac{1}{2} \ln \left(\frac{1}{6\epsilon} (\frac{q}{\alpha})^2 \right) \right) \tag{5.24}$$

5.3.4 Optimization problems and algorithms

Cost function

The previously obtained rate and distortion models are used to compute optimal QPs of all CTUs of the sequence. The optimization process uses the Lagragian cost J of each frame. It corresponds to:

$$J(q) = \sum_{i=1}^{N} d(q_i) + \lambda r(q_i) \tag{5.25}$$

where, N is the number of CTUs per frame, $q = q_{i \in [1,N]}$ is the vector of quantization steps of all CTUs of a frame, $d(q_i)$ and $r(q_i)$ denote respectively the distortion in MSE and the rate in bpp of the i^{th} CTU. However, in practice, the cost per CTU should respect bit rate condition. Thus, depending on the q_i value, the distortion $d(q_i)$ and the rate $r(q_i)$ used are:

- for intra-coded frame : Equations (5.13) and (5.16) at low bit rate or Equations (5.19) and (5.22) at high bit rate.
- for inter-coded frame : Equations (5.13) and (5.23) at low bit rate or Equations (5.19) and (5.24) at high bit rate.

To compute optimal q per frame, we investigated different convex optimization approaches. We started with a non optimal approach based on the gradient descent algorithm that minimizes the cost J without considering particular constraints on the QP value. Then, we used interior-point methods to properly minimize the cost J considering required constraints on the QP value, mainly to limit the QPs in a limited range and to reduce the QP variation inside a frame [81].

Minimization without constraints on the QPs

- Minimization problem :

  The problem considered in our work uses a convex and differentiable Lagrangian cost function $J : \mathbb{R}^n \to \mathbb{R}$. The unconstrained minimization problem can be written as follows,

$$\underset{q}{\text{minimize}} \quad J(q) \tag{5.26}$$

- Gradient descent :

  An unconstrained minimization can be solved by the Gradient descent algorithm [82]. Using the gradient $\nabla J(q)$ at location q points toward direction where the function increases, this method finds the minimum of our cost J. It starts from an initial point q_0, then iteratively takes a step along the steepest descent direction $-\nabla J(q)$ that can be scaled by a step-size α, until convergence. The algorithm used in our experiments can be described as following:

Algorithm 1 Gradient descent algorithm

Input: Strating point q , a function J, stepsize α, tolerance θ
Output: A q vector minimizing J
1: **Repeat**
2: $q \leftarrow q - \alpha \nabla J(q)$
3: **Until** $\Delta J < \theta$

Minimization with constraints on the QPs

- Minimization problem :

  In a practical case, our optimization problem should consider particular constraints. First, in HEVC coding, the quantization parameter (QP) cannot exceed a certain range fixed by the encoder (QP $\in [0, 51]$) and r and d are positive values. Second, for a smooth and regular quality of the sequence it is possible to introduce a constraint to limit spatial and temporal variation of QP i.e. between CTUs of the same frame and between successive pictures. Furthermore, to evaluate obtained results it is important to test the algorithm at different rate levels. Thus, we tested different λ values that corresponds to the initialized QP. To obtain the appropriate quantization step q with respect to all these constraints, the problem can be represented as follows:

$$\begin{aligned} \underset{q}{\text{minimize}} \quad & J(q) \\ \text{subject to} \quad & q_{min} \leq q_i \leq q_{max},\ i = 1, \ldots, N \\ & |q_i - q_{mean}| \leq L,\ i = 1, \ldots, N \end{aligned} \tag{5.27}$$

In this constrained optimization problem L corresponds to the limit of q variation. In the I frames, we have chosen a QP variation limited to ± 2 comparing to the mean QP of the whole frame. While, in the B frames, the variation between successive frames is limited to ± 10 compared with an I picture and ± 3 compared to a B picture.

A simple extension of the unconstrained minimization method does work well. The idea is to solve a sequence of unconstrained minimization problems, modify the last point found and use it as a starting point for the next iteration. Thus, in our experiments, we first use the gradient descent algorithm which is an unconstrained minimization approach and we just perform a clipping of the obtained$q$ values. We also tested constrained minimization methods such as interior-point algorithm for faster convergence [81].

- Proposed algorithm based on Gradient descent :

  Given the cost function $J(q)$ defined in (5.25). We want to find its minimum using the Gradient descent algorithm and considering particular cases and required restrictions. In fact, considering that in the practical case q values are defined in a limited set, a clipping of q values is added at each iteration of the Gradient descent algorithm. λ is fixed during the full process. However, the cost J is updated at each iteration considering appropriate RD models at high or low bit rate. The number of iterations have been chosen empirically. Experiments have shown that 10000 iterations could be enough as in many cases the algorithm converges before.

Algorithm 2 Modified gradient descent algorithm

Input: An initial quantization step q_{int} , stepsize α, tolerance θ , clipping range q_{min} and q_{max} , Number of iterations M
Output: A q vector minimizing J
1: Compute λ and q considering initial Q_{int}
2: $J(q) \leftarrow \sum_{i=1}^{N} d(q_i) + \lambda r(q_i)$
3: **Repeat**
4: $q \leftarrow q - \alpha \nabla J(q)$
5: $q \leftarrow \text{clip}(q, q_{min}, q_{max})$
6: Update cost J(q)
7: **Until** $\Delta J < \theta$ for M iterations in frame
8: Compute q_{mean}
9: $q \leftarrow \text{clip}(q, q_{mean} - L, q_{mean} + L)$

- Interior-point algorithm :

  The studied optimization problem includes inequality constraints. The Interior-point method formulate the inequality constrained problem as an equality constrained problem. Then, it solves the problem in Equation (5.27) by applying Newton's method to a sequence of equally constrained problems [81]. The Matlab function "fmincon" with algorithm option "Interior-point" is used in our experiments.

5.4 Experimental results

5.4.1 Experimental setting

In this part of the work, the data set used for experiments are Class B sequences represented in Fig.5.10. They are tested to analysis impact of different factors in transform coefficient distribution and to evaluate proposed model performance.

(a) "BaseketBallDrive" (b) "BQTerrace" (c) "Cactus"

(d) "Kimono1" (e) "ParkScene"

Figure 5.10: Class B sequences (1920x1080)

In fact, test sequences are encoded using HM.16 reference software. Non-quantized coefficients are generated per transform level and per CTU. First, we evaluate the performance of our model in intra-coded units with CTU size equal to 64×64. All-intra configuration is used to encode the sequence with a $Q = 1$. We estimate α and β parameters of the GG distribution per CTU. Second, we evaluate the performance of our model in inter-coded units. Low-delay configuration is used with a $Q = 1$ to encode test sequences. Hierarchical levels are not considered and each frame take as reference the last decoded frame. The BGG distribution parameters $(\alpha, \beta, \epsilon)$ are estimated per CTU to be used in our RD models.

For each configuration, we run the optimization algorithm and analyze the obtained quantization maps. Then, we introduce the optimized QP map to encode the sequence with the corresponding configuration. Finally, we make comparative tests between proposed model and R-λ model by evaluating their RD performance.

5.4.2 Gradient descent algorithm behavior

Gradient descent algorithm is performed to compute appropriate QP per unit. Now we study the behavior of the proposed optimization algorithm and the evaluate the obtained QP selection. At each iteration, the algorithm find the QP map that reduces the cost

J and improve RD performance. In other words, at each iteration RD performance increases until we reach an optimum value. Fig.5.11 represents RD curves at 4 iterations of the optimization algorithm (Iteration 1, 10, 100 and 500). Seven rate levels are tested ($Q_{int} \in \{1, 10, 15, 20, 25, 30, 35\}$) to plot the RD curve. The figure shows and improvement in RD performance from one iteration to another. If we compute Bjontegaard metric between the RD curve at iteration 1 and the one obtained at iteration 500, we notice an increase in PSNR of 6 dB and a bit rate saving of 20%.

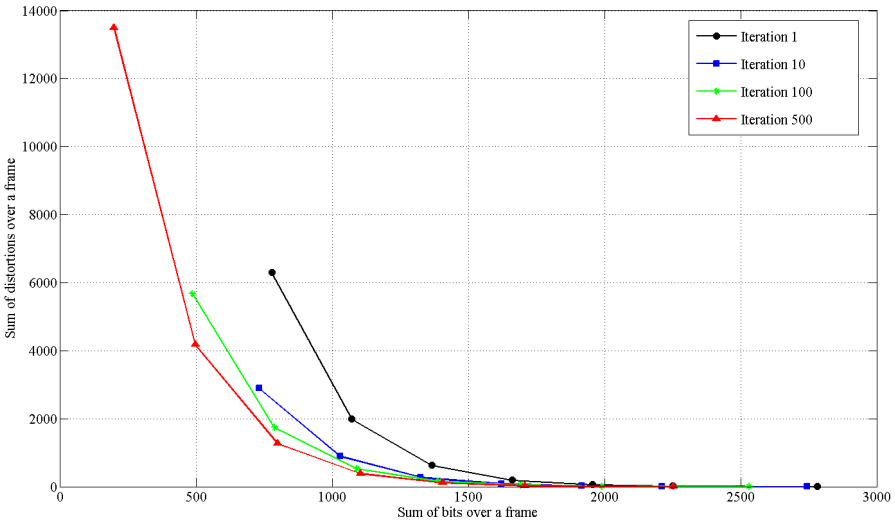
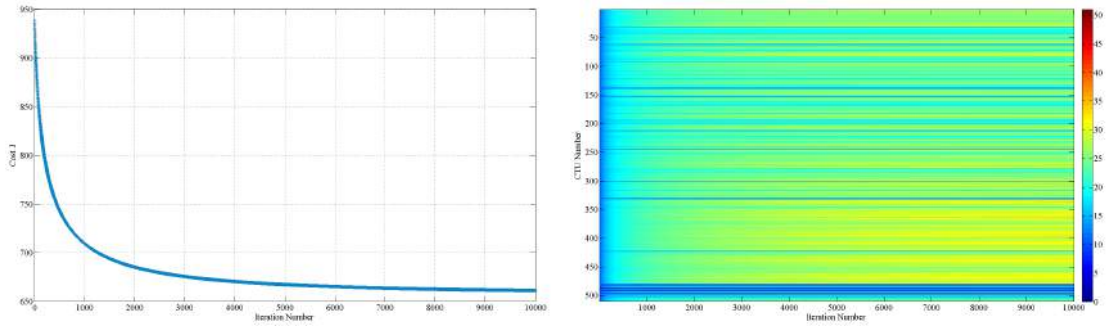


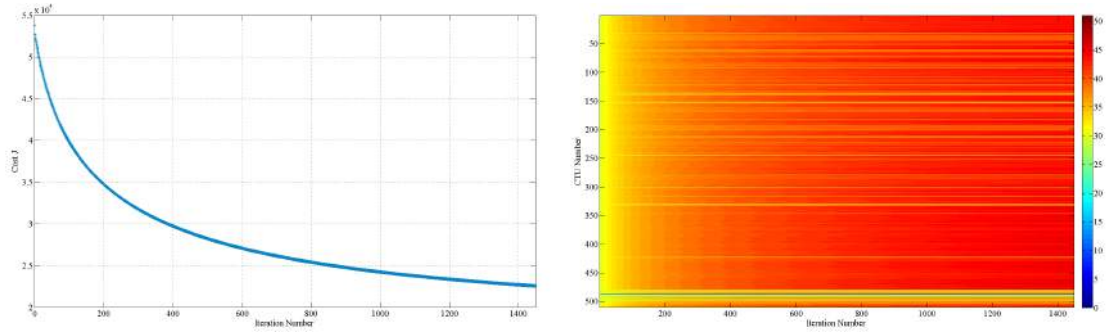
Figure 5.11: Improvement of RD performance at each iteration of proposed gradient descent algorithm - Example of the first frame of "BasketBallDrive" sequence

It is also possible to evaluate the evolution of the cost J when optimizing QPs per frame. In Fig.5.12 and Fig.5.13, we plot the cost value per iteration (a) and the list of QPs per iteration (b). We notice that depending on rate level, the minim cost can be reached after 10000 iteration 5.12(a) or 1448 iterations 5.13(a). In both cases the algorithm converges to a minim cost and give a new QP map for the frame.



(a) J evolution for $Q_{int} = 10$ (b) QP evolution for $Q_{int} = 10$

Figure 5.12: Evolution of frame cost J and QP of all CTUs over gradient descent algorithm iterations at low bit rate



(a) J evolution for $Q_{int} = 30$ (b) QP evolution for $Q_{int} = 30$

Figure 5.13: Evolution of frame cost J and QP of all CTUs over gradient descent algorithm iterations at high bit rate

In Fig.5.12(b) and Fig.5.13(b) we represent the QP evolution per iteration. When we start with an overall $Q_{int} = 10$ as shown in the first figure, λ is so equal to 0.41 and after 10000 iterations the mean of the overall obtained QPs is equal to 25. While, if we start with an over $Q_{int} = 30$ 5.13(b), the fixed value of $\lambda = 48.31$ and after 1448 iterations the mean of the overall obtained QPs is equal to 42.

5.4.3 Optimal QP selection

The proposed model gives different QP maps than the one given by R-λ model. In this section, we evaluate the QP repartition over different frames of "BasketBallDrive" sequence coded in intra and inter mode. As the QP computing is done at CTU level, we start by showing in Fig.5.14 the used CTU partitioning. CTU size is equal to 64×64 and the frame resolution is 1920×1080. Thus, the obtained QP map is a matrix of 30×17 values.

Figure 5.14: CTU partitioning of "BasketBallDrive" sequence

All-intra configuration

In intra case, the proposed RD model parameters (α, β) models local characteristics (texture) of the frame as they are computed by CTU and spatial dependencies between units as they are generated by fitting transform coefficient after intra prediction. This results in a texture based QP map. From Fig.5.15, we notice that the smallest QPs are assigned to very textured regions such as the walls (QP values from 5 to 20), medium values are selected for regions with regular texture for example the floor (QP values around 30), while the highest QPs are used for encoding smooth regions such as player t-shirt (up to 41). The obtained QP range goes from 5 to 41. It can be reduced to have smother quality over a frame by introducing a constraint when performing QP optimization.

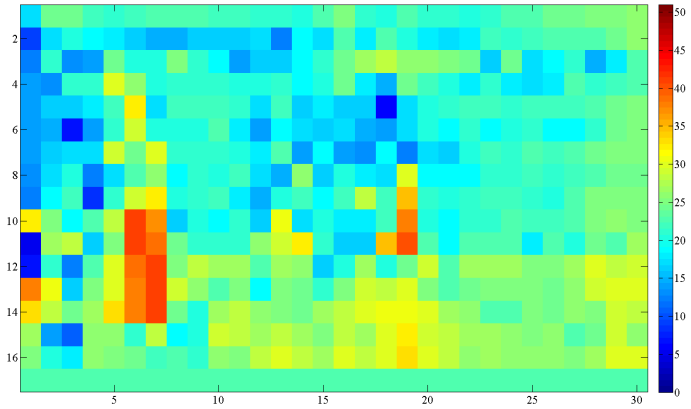
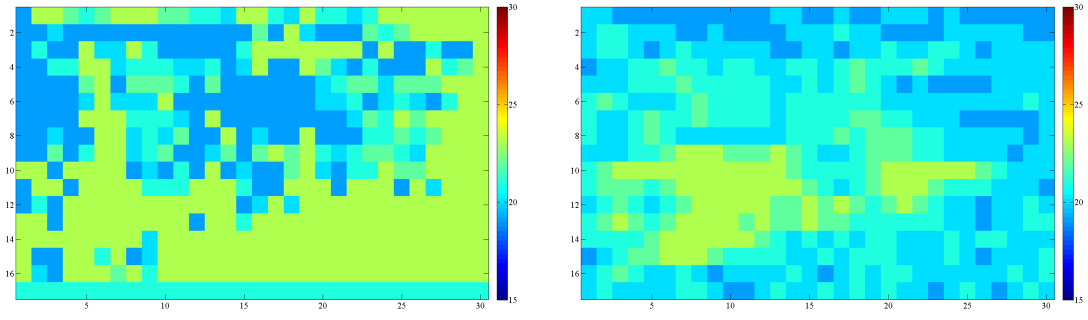


Figure 5.15: Optimal QP map using proposed unconstrained model of an intra-coded frame of "BasketballDrive" sequence

Fig.5.16 shows in (a) the QP map obtained using our model after performing a constrained rate distortion optimization. While, (b) shows the QP repartition given by the R-λ model. We notice that our method gives a texture based repartition which is not the case with the R-λ model used in HEVC.



(a) Proposed constrained model

(b) R-λ model

Figure 5.16: Comparison of obtained QP maps using proposed constrained model and R-λ model of an intra-coded frame of "BasketballDrive" sequence

Low-delay configuration

In inter case, the proposed RD model parameters $(\alpha, \beta, \epsilon)$ are generated by fitting transform coefficient after inter prediction. Consequently, they model temporal dependencies between successive frames. Fig.5.18 shows that the optimization process gives smaller QP values to moving objects (moving player) where a motion estimation should be performed and higher QPs to regions where no movement is noticed over time (background). An important QP range is obtained using our model ($Q \in [13; 44]$).

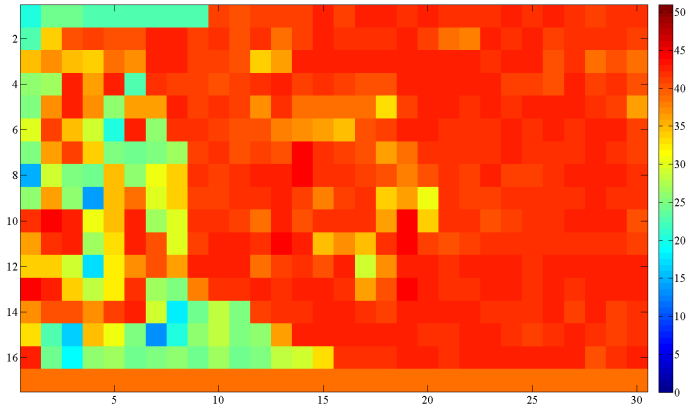
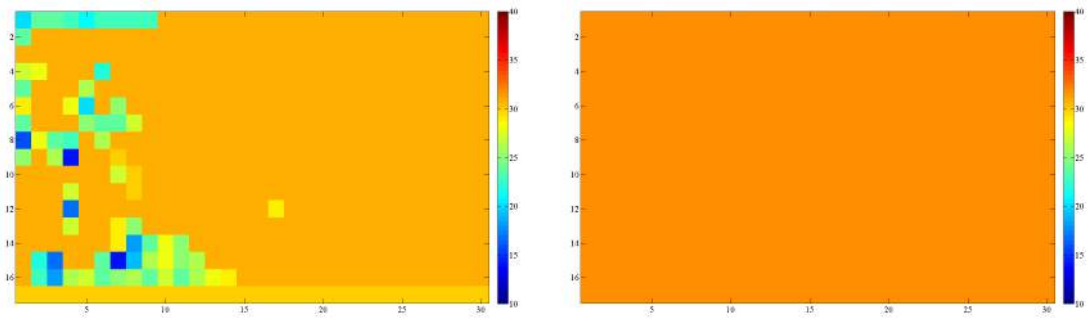


Figure 5.17: Optimal QP map using proposed unconstrained model of an inter-coded frame of "BasketballDrive" sequence

Here again a range constraint was introduced to limit QP variation between successive frames. This results in the reduction of the QP range in the given example to $[14; 32]$ (Fig.5.18(a)). Our model is more relevant as small QPs are used to encode CTUs in the moving edges. However, in the R-λ model foreground and background CTUs are coded using the same QP value (b).



(a) Proposed constrained model (b) R-λ model

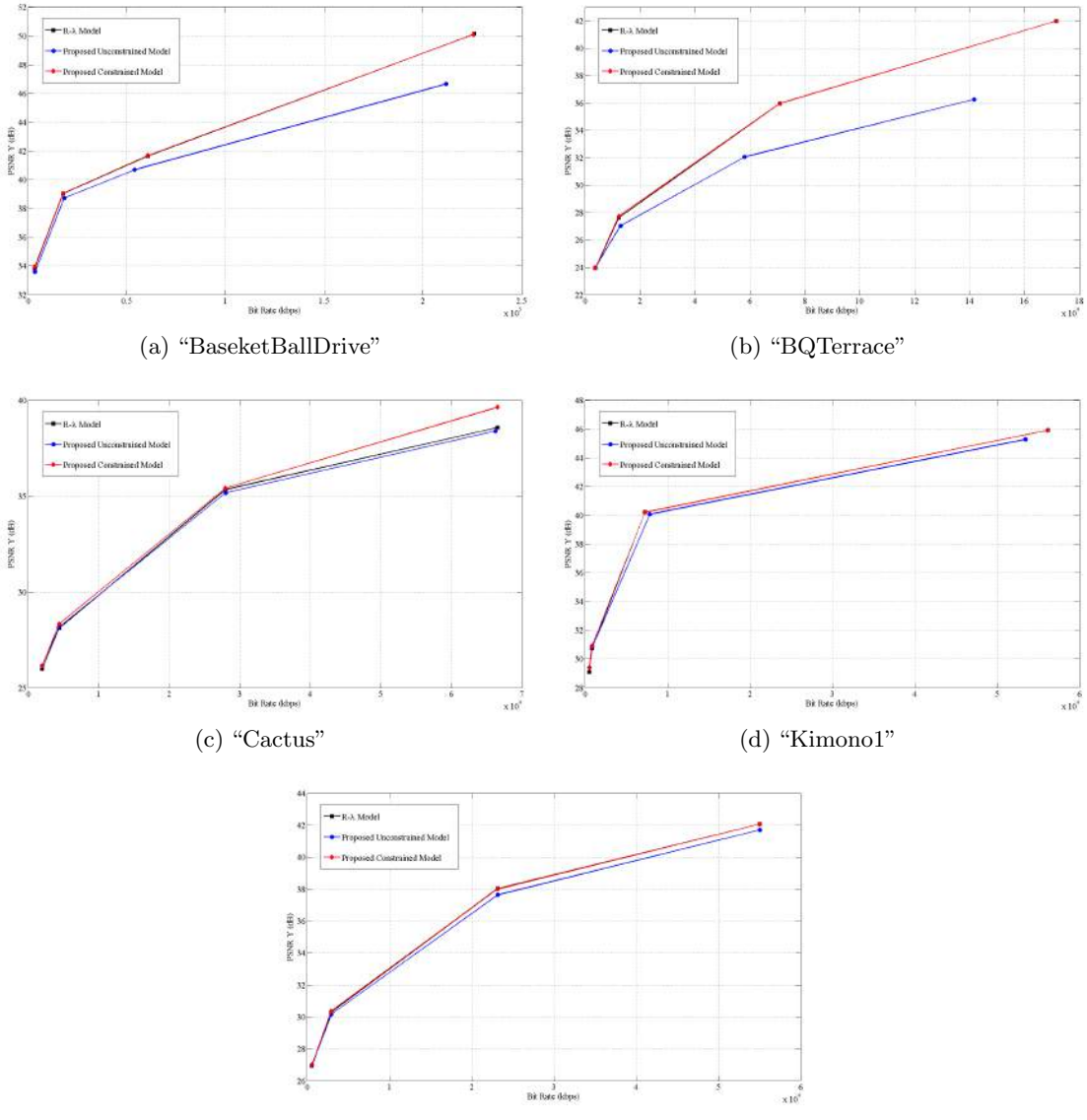
Figure 5.18: Comparison of obtained QP maps using proposed constrained model and R-λ model of an inter-coded frame of "BasketballDrive" sequence

5.4.4 Comparison of RD performance of the proposed model and R-λ model

The proposed RD model has been compared to the R-λ model. The optimized QP map is used in HEVC to encode the sequence in both all-intra and low-delay configurations.

All-intra configuration

In intra case, we notice from Fig.5.19 that clipping the optimized QP map is important to get better RD performance. The quality fluctuation over CTUs of the same frame may affect the encoding process and reduce the global quality.



(a) "BaseketBallDrive"

(b) "BQTerrace"

(c) "Cactus"

(d) "Kimono1"

(e) "ParkScene"

Figure 5.19: Comparison of RD performance of R-λ model and proposed model in all-intra mode

As shown before the proposed constrained model gives a different QP partitioning than the R-λ model. The obtained QP map helps improve slightly the RD performance. The bit rate gain goes from -1.16% to -4.75% as shown in Table.5.5.

Sequence	Percentage of bit rate gain	PSNR gain (dB)
"BasketBallDrive"	-1.38%	0.04
"BQTerrace"	-1.48%	0.07
"Cactus"	-4.75%	0.18
"Kimono1"	-1.70%	0.04
"ParkScene"	-1.16%	0.04

Table 5.5: RD performance of the proposed model compared to R-λ in all-intra mode

The importance of the clipping can be explained by the fact that neighboring CTUs are dependent and a big quality difference may affect the intra prediction and consequently costs a lot.

Low-delay configuration

Using a low-delay configuration, only the first frame is coded in intra mode. All following frames are B-pictures. From Table.5.6, we notice an important bit rate gain comparing to R-λ model. Depending on the encoded sequence our unconstrained model have a bit rate gain from -45.10% up to -88.79%. In fact, our model gives a better bit repartition over CTUs of the same frame but also over successive frames.

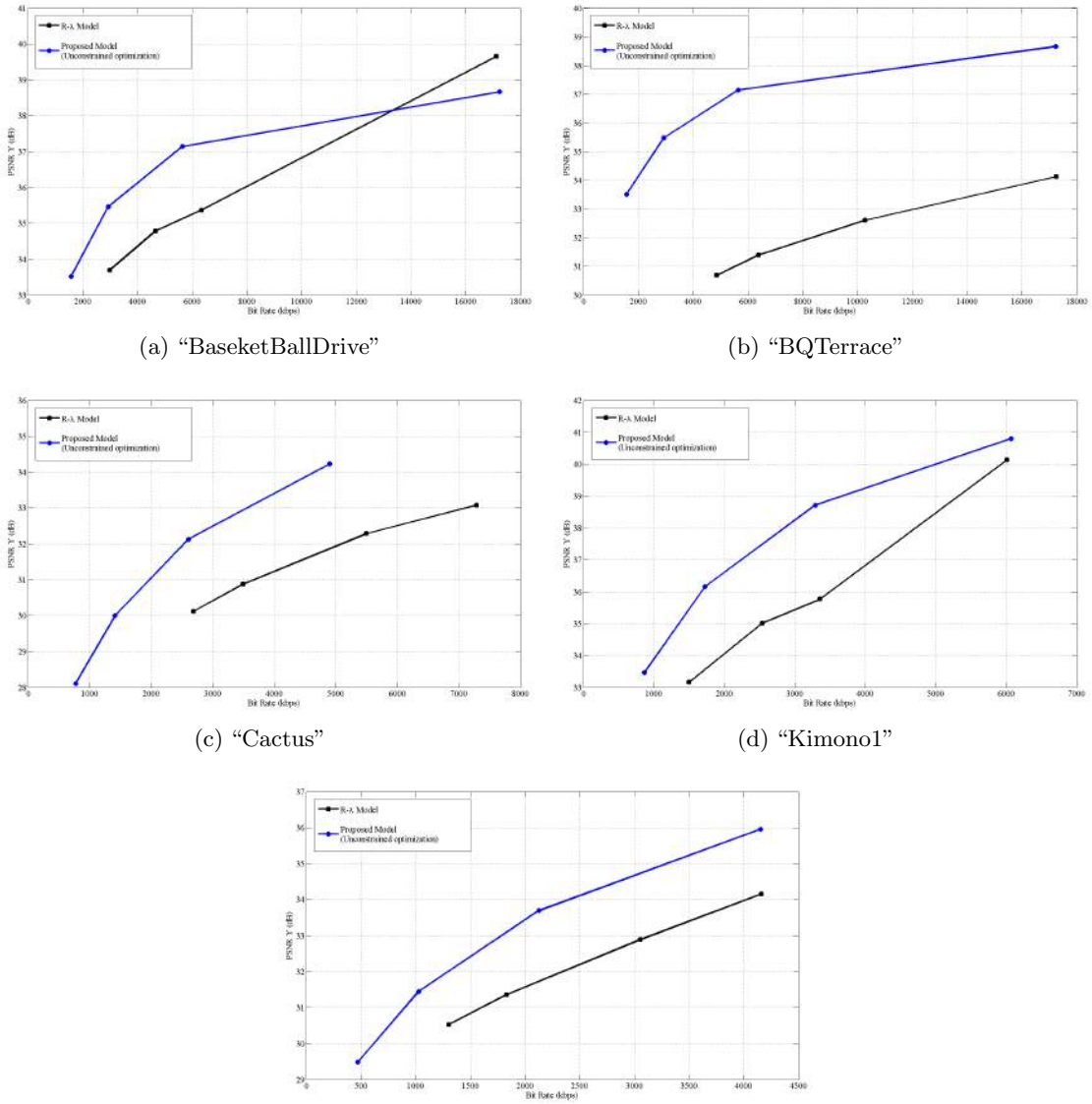
Introducing a constraint in the QP range reduces the RD performance in some cases. In fact, the selection is then limited and foreground and background units may have too close QP values. In that case, the selection in not anymore optimal.

Sequence	Unconstrained optimization		Constrained optimization	
	Percentage of bit rate gain	PSNR gain (dB)	Percentage of bit rate gain	PSNR gain (dB)
"BasketBallDrive"	-55.00%	1.48	-36.51%	0.58
"BQTerrace"	-88.79%	5.60	-2.82%	0.03
"Cactus"	-49.17%	2.25	-52.88%	2.49
"Kimono1"	-51.28%	2.48	-44.42%	2.32
"ParkScene"	-45.10%	1.89	-40.05%	1.60

Table 5.6: RD performance of the proposed model compared to R-λ in low-delay mode

Furthermore, the proposed model can reach low bit rates that the R-λ model cannot respect. Fig.5.20 represents RD curves of all tested sequences obtained using proposed unconstrained model and R-λ model in HEVC. It shows that for each sequences 4 bit rate points are tested and in many cases the R-λ model is not able to respect the budget constraint. The R-λ given QP selection is not optimal and our algorithm show considerable rate and quality gain.

Furthermore, experiments have shown an improvement in encoding quality. For the same bit budget, an increase in PSNR is measured for all decoded sequences (Table.5.6). We notice that considering the proposed model we obtain a better bit partitioning than the



(a) "BaseketBallDrive" (b) "BQTerrace"

(c) "Cactus" (d) "Kimono1"

(e) "ParkScene"

Figure 5.20: Comparison of RD performance of R-λ model and proposed model in low-delay mode

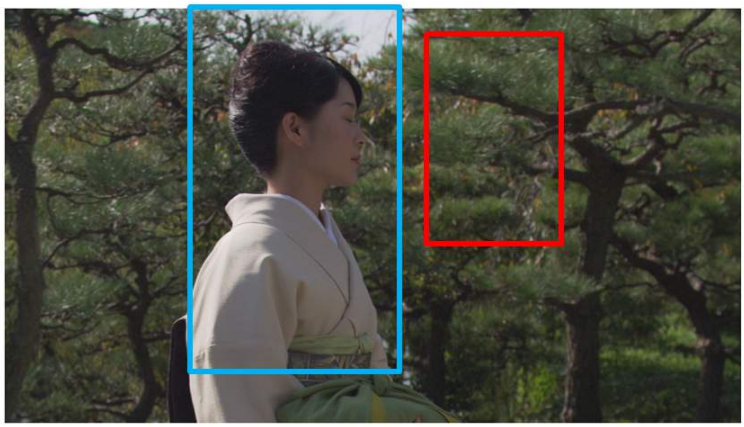
R-λ model. Fig.5.21 shows that for the same bit rate we better code the moving objects such as the ball (red square) and the numbers in players' t-shirts (blue square). In Fig.5.22, the sequence is decoded at 3Mbps, the eye is correctly decoded using our model (closed eye in the blue square) and the trees have better texture (red square).

(a) Original frame

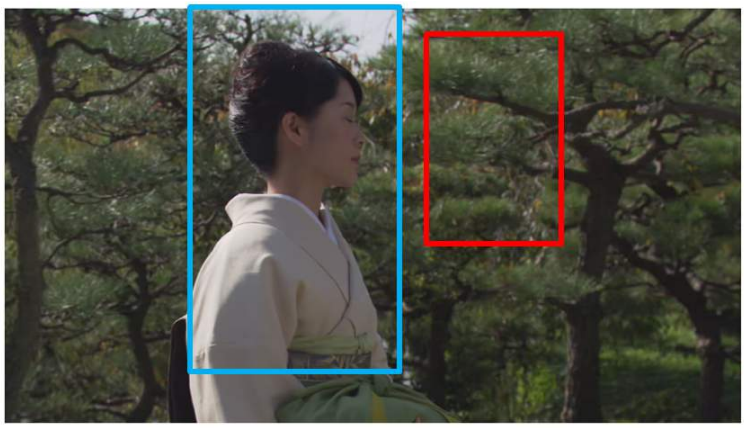
(b) Proposed model

(c) R-λ model

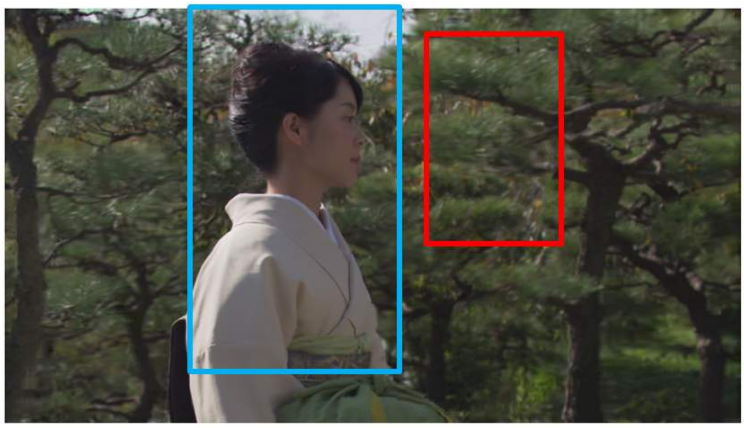
Figure 5.21: Comparison of subjective encoding quality of "BasketballDrive" frame using R-λ and proposed models at 6Mbps



(a) Original frame



(b) Proposed model



(c) R-λ model

Figure 5.22: Comparison of subjective encoding quality of "Kimono" frame using R-λ and proposed models at 3Mbps

5.5 Conclusion

In this work, we studied different content-based RD models. In a first part (Section 5.1), we adapted appropriate models to CTU level bit allocation by approximating their parameters by considering HEVC at independently decodable CTUs of type I. The obtained results demonstrate that the proposed exponential RD model allows us to accurately describe the spatial dependencies over an intra coded CTU. The proposed model can be useful for optimal bit rate allocation at CTU and frame level to perform efficient rate control for HEVC.

In a second part (sections 5.2, 5.3 and 5.4), the rate-quantization equations proposed in [76] are adapted for high efficiency video coding (HEVC). A simplified GG model is used for intra-coded frames while a BGG model is appropriate for inter-coded frames. They are commonly used probabilistic models. However, they have never been used to evaluate HEVC transformed residual and they well fit the transform coefficient distribution. Novelty in our work consists in estimating model parameters based on BGG and GG distributions, evaluating obtained values considering different configurations, and, proposing RD models for QP computing at CTU level for HEVC.

Part III

ROI-based rate control

Chapter 6

ROI-based rate control: State of the Art

Contents

Region-of-Interest (ROI) based perceptual video coding has been established for a long time. The main idea is straightforward: enhance visual quality by improving fidelity of human interested regions. Since the standardization of H.264/AVC, several improvements and variations to the original algorithms have been done. Several ROI-based coding strategies have been proposed providing effective rate controlled image compression over regions. Given a particular ROI of a fixed or varying shape, proposed schemes are able to compress a given image by a required ratio.

This chapter presents a state-of-the-art review on the available schemes for rate control in ROI-based video coding. In the first section, we introduce different ROI detection and tracking techniques. Then, we describe in the second section the ROI-based rate control algorithms implemented in earlier standards such as H.264/AVC.

6.1 ROI detection and tracking

Many algorithms have been proposed for automatic ROI detection. They can be classified into two categories: bottom-up methods assume that human eyes skirt rapidly across the entire image and select small areas, while top-down methods suppose that people pay more attention to areas corresponding to semantic objects of the image [83]. Top-down approaches mainly consist in generating a saliency map taking into account the importance of semantic objects such as text, faces, eyes, etc.

Generally, researches on object detection and tracking have focused on the pixel domain approach since it provides powerful tracking capability. In pixel domain the ROI can be detected using different models, such as visual attention models, object detection models and face detection models.

6.1.1 Visual attention models

Visual attention models refer to the ability of a human to concentrate his attention on a specific region of the image. This involves selection of the sensory information by the primary visual cortex in the brain by using a number of characteristic, such as intensity, color, size and orientation in space. Actually, the visual attention models simulate the behavior of the Human Visual System (HVS), and in turn enable to detect the ROI within the image [84] as represented in Fig.6.1. It often represents a semantic object such as a human face, a flower, a car, a text, etc.

Figure 6.1: Example of concentrating the attention on a specific region of an image

6.1.2 Movement detection

Detecting, classifying as well as tracking objects and human motions are important tools in image processing used in security systems. Several approaches for moving-object detection

and tracking have been proposed during the last decades using image processing techniques, because computer vision lets us to manipulated videos to extract useful information contained in the coded stream. These algorithms consist in estimating the location of each object in each frame and keeping track of it. In [85], the proposed system detects movements of objects and persons based on video sequence processing. A movement trace is estimated referring to motion vectors to determine whether tracking should be carried or not. In [86], the proposed moving-object detection method combines both temporal variance of the pixel intensities as temporal thresholding approach with background modeling. Then, tracking is performed by combining motion and appearance information.

6.1.3 Object and Face detection

Automatic object detection methods locate objects in an image and extract the regions containing them (the extracted regions are ROIs). The detection is performed using particular features of the object. In fact, having a good feature-based representation of objects increases the effectiveness of the detector. The ROI detection is especially useful for medical applications, video surveillance systems, etc.

The face detection is a specific case of object detection. In object-class detection, the task is to find the locations and sizes of all objects in an image that belong to a given class. One of the earliest works in face detection is a real-time system developed in [87] to emphasize the face region. The proposed method is based on a shape recognition algorithm. The system is able to detect and track human faces considering skin color segmentation and contour evaluation. Face detection can be combined with silent features (color, intensity and orientation) as done in [88] to improve ROI detection accuracy.

Furthermore, Viola and Jones object detector [89] is a famous and successful tool, widely used for face detection. For specific applications, such as video-conference or supervision systems, this algorithm is appropriate as it has shown strong power in detecting faces, while for other applications, some improvement has been introduced to Viola and Jones algorithm by introducing new feature images. This framework used a set of Haar-like features in which each characteristic was described by a template. OpenCV library has included different implementations of Viola and Jones object detector algorithm [90].

In our work, we focus on face detection as we are studying videoconferencing systems. We need a simple and rapid method to detect the ROI and perform ROI-based bit allocation at real time. Consequently, we use OpenCV library for face detection as shown in Fig.6.2. We do not aim at making a perfect segmentation of the face at pixel level. Our algorithm requires a classification of CTUs in different regions. Thus, once the faces are detected a binary mask is generated to register if each CTU of the frame belongs to the ROI or not.

(a) Johnny (b) KristenAndSara

(c) FourPeople

Figure 6.2: Face detection using OpenCV library

6.2 ROI-based rate control for H.264

With rapid demands for ROI in applications like videoconferencing, video surveillance and telemedecine, ROI-based rate control has gained increasing attention from researchers. Previously described ROI detection and tracking techniques make ROI-based video coding possible. In a ROI video coding scheme, smaller quantization parameter is used to represent the ROI with lower distortion which could significantly contribute to the subjective quality of the ROI and the overall video. Different controllers have been proposed for different situations and implemented in the H.264/AVC reference software. This review of available schemes helps us chose appropriate model to compare with.

6.2.1 ROI quality adjustable rate control scheme

In [91], a ROI quality adjustable rate control algorithm has been proposed. Bit allocation is initially done according to user's interest level and available budget. The proposed quadratic RD model defined in (3.8) considers the bit rate constraint and possible quality levels to define a QP margin. A number of bits is then allocated for each region and the QP is refined. Finally, the quadratic *R-q* model is used to assign a QP per region. MBs of the same region get the same QP.

The particularity of this method consists in reordering MBs. In fact, ROI is processed first, then the non-ROI areas. This approach cannot be adapted to HEVC coding as it processes units in raster scan order.

6.2.2 ROI-based rate control for traffic video-surveillance

In [92], a ROI-based rate control was designed for traffic surveillance systems. A fast ROI extraction method for the real time video compression is used to generate the ROI map. A linear function has expressed the relation between the bit-stream length and the quantization step (3.7). This model helps to predict the frame level bit allocation and the region level QP determination. In this work, the model is applied for each macroblock. Thus, a QP is computed for each macroblock.

This method is based on linear rate quantization model which is not the case in HEVC.

6.2.3 ROI-based rate control for low bit rate applications

In [93], a complete ROI-based controller is proposed. The scheme includes five steps, starting with region dividing using the RD characteristics of each MB. Macroblocks with similar characteristics are classified in the same basic unit and an overall bit allocation is performed using two linear models: a rate quantization (R-q) model and and a distortion quantization (D-q) model. A QP is computed for each basic unit. Finally, RDO is performed for each MB and models' parameters are updated as done in previous propositions.

Here again linear rate distortion models are used for QP computing which not adapted to HEVC coding.

6.2.4 ROI-based rate control scheme with flexible quality on demand

In [94], the same quadratic model described in (3.8) is used. Faces are considered as ROIs. However, new features are introduced comparing to previously detailed proposition. First, human psycho-visual clues are used to compute a saliency map for each frame, which is used for rate control. A quality factor is defined and the bit budget is allocated for ROI and non-ROI separately. Finally, the quadratic model is used to assign a QP for each region considering a clipping range for smooth visual quality along the temporal direction and across region boundaries.

The RC algorithm proposed in [94] is the most appropriate for HEVC. It is possible to adapt it to the HEVC controller, as it uses a quadratic model for QP computing which is not the case in [92] and keeps processing blocks in encoding order, which is not the case in [91]. Consequently, we implemented the proposed ROI-based controller in [94] in HM.9 and compared it to our algorithm. A detailed description of these algorithms is given in the next section.

6.3 Conclusion

In literature, different ROI detection algorithms have been proposed as shown in Section 6.1. they have been afterworld used to improve perceptual quality of important areas in a video sequence.

The above-mentioned algorithms in Section 6.2 provide a bit rate repartition that takes into account the high priority of the ROI. They have been developed considering linear and quadratic models and implemented in the H.264/AVC software. In next chapters, we propose a new ROI-based rate control scheme for HEVC characterized by several features. We used Viola and Jones algorithm for face detection and we compare obtained performance with ROI-based rate control approach described in Section 6.2.4.

Chapter 7

ROI-based rate control for HEVC

Contents

This chapter presents novel rate control scheme designed for HEVC, and aimed at enhancing the quality of ROIs for a videoconferencing system. The proposed approach has been compared to the reference controller implemented in HM.10 and to a ROI-based rate control algorithm initially proposed for H.264/AVC that we adopted to HEVC and implemented in HM.9.

In the first section of this chapter, we motivate the introduced modifications in the quadratic model proposed in [94] and described in Section 6.2.4 to adapt it to HEVC. In the second section, we explain the proposed ROI-based R-λ model. Main features of the proposed algorithm and different versions are detailed. The chapter ends with experimental results related to all proposed algorithms and interpretation of obtained results.

7.1 ROI-based quadratic model

The ROI-based controller proposed in [94] for H.264/AVC standard consists in estimating the bit count per region by using a quadratic RD model. We adopted this algorithm to the quadratic URQ controller presented in HEVC contributions [63] [64] and implemented in HM.5 and later versions. The proposed algorithm has been enhanced with several features. In fact, bit allocation is performed per region. Then, quadratic URQ mode is used to compute a QP per CTU as done in H.264/AVC. Finally, QP is adjusted and the unit is encoded.

7.1.1 Bit allocation per region

At frame level, separate bit allocation per region is performed. First, the initial budget fixed by the network is divided into two parts using a quality factor K assigned by users or control systems. Target bit counts T_r and T_n are initialized to ROI and non-ROI referring to Equation (7.1) , then used for bit allocation at frame and CTU level.

$$T_p = T_r + T_n \text{with, } \frac{T_r}{N_r} = K \times \frac{T_n}{N_n} \tag{7.1}$$

where N_r and N_n denote respectively number of pixels of ROI and number of pixels of non-ROI. The final target bit left budget $\widehat{T}_r(i)$ for the i^{th} CTU from the ROI is based on the remaining bits in ROI $(T_r - T'_r)$, the number of pixels in the current CTU $N(i)$ and the number of pixels left in ROI:

$$\widehat{T}_r(i) = \frac{(T_r - T'_r) \times N(i)}{\sum_{\substack{j \in I_r \\ j>i}} N(j)} \tag{7.2}$$

The final target bit occupancy $\widetilde{T}_r(i)$ for CTU from the ROI is computed using the initialized bit count in ROI and ROI virtual buffer occupancy $V_r(i)$:

$$\widetilde{T}_r(i) = T_r - \frac{V_r(i)}{U_r(i)} \tag{7.3}$$

where $U_r(i)$ is the number of units left in ROI after encoding CTU of index i.

The final bit budget is a weighted average of the target bit left and the target bit occupancy:

$$T_r(i) = \beta \times \widehat{T}_r(i) \times (1 - \beta) \times \widetilde{T}_r(i) \tag{7.4}$$

where β is the weight defined in [64]. Depending on the application needs this paramater can give more weight to the target bit left or the target bit occupancy. The same process is done for CTUs of the rest of the frame.

7.1.2 Quadratic model for QP determination

The strategy for intra pictures and non-reference frames is kept as described in HM document [64] Howver, for referenced B-frames the ROI-based URQ model is used at CTU level. In this case, the final bit target $T_r^f(i)$ is refined as follows:

$$T_r^f(i) = T_r(i) \times \frac{w_B(i)}{\sum_{\substack{j \in I_r \\ j \geq i}} w_B(j)} \tag{7.5}$$

where $w_B(i)$ is the MAD of the current CTU as expressed in Equation (1.7). After estimating this target bit count for the considered CTU, the preliminary QP value is determined as in [64] by the quadratic model introduced in Chapter 3 by Equation (3.8).

The obtained QP using the quadratic RD model is then modified by considering the smoothness issues over the temporal and spatial domains. The four constraints proposed in [94] are respected. All QPs are then clipped between 0 and 51 as proposed in the URQ reference controller implemented in HM.9.

7.2 ROI-based R-λ model

The proposed approach is based on the R-λ model presented in HEVC. The relationship between R and λ represented in Chapter 3 by Equation (3.13) is used to compute QP of the frame and each CTU of the image. As shown in Section 4.2, this model gives better performance than the quadratic one. Our contribution proposes a ROI-based rate control algorithm where bit allocation at CTU level depends on the number of bits allocated per region and on the weights of CTUs of the same region [95] [96].

In this section, we describe the proposed approach that has been implemented in HM.10 and how we adapted it to a later version of HEVC test model 13 (HM.13). We focus on the two main steps of the rate control: the bit allocation at both frame and CTU levels and the computation of QP by the proposed model for both I and B frames.

7.2.1 Proposed ROI-based scheme

Fig. 7.1 shows the proposed ROI-based rate control scheme. The first step consists in detecting the faces in the scene and generating automatically a binary ROI map per frame, which will be given as input to our controller. The target bit rates allocated for the GOP and the current frame are obtained using the reference algorithm described in [54] and improved in [66].

Then, the frame budget is divided into two parts according to a fixed factor K which is the *desired* ratio between the bit rate of the ROI and the bit rate of the rest of the frame (non-ROI). At the CTU level, the binary ROI map is used to make a separate bit allocation for CTUs of different regions. The R-λ model is then applied for each CTU using the allocated bit budget for the corresponding region (ROI or non-ROI). Once the

CTU is encoded, the model parameters of the corresponding region are updated, and the next CTU is processed in a similar way.

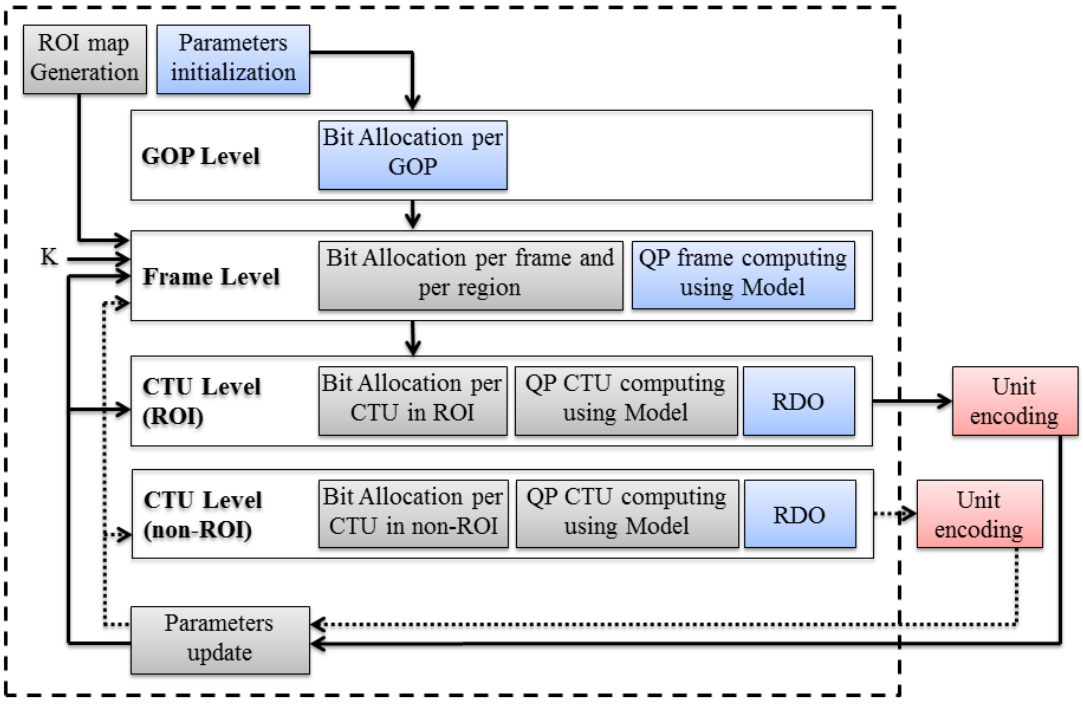


Figure 7.1: ROI-based rate control scheme for HEVC

In the first implementation of the controller (in HM.10), the described process is only used for B-frames of different hierarchical levels. Then, it was adapted to HM.13 and introduced in both I-frames and B-frames, considering some differences in CTUs' weights computing and model parameters update.

7.2.2 Main features of the proposed ROI-based controller

Region bit allocation for inter-frames

We introduce the region bit allocation at two levels; at frame level to initialize a target amount of bits for each region, and at CTU level to make independent bit allocation of CTUs of different regions. At frame level, the positive constant K is selected. It represents the desired ratio between the ROI and non-ROI bit rates:

$$R_r = K \times R_n \tag{7.6}$$

where the subscript r denotes the ROI and n the non-ROI. We assume that the current number of allocated bits per frameT_p is the sum of the number of bits of the two regions,

T_r for the ROI and T_n for the non-ROI:

$$T_p = T_r + T_n \tag{7.7}$$

$$T_n = R_n \times M \times P_n \tag{7.8}$$

where M is the total number of pixels of the frame and P_n the area of non-ROI. From Equations (7.6), (7.7) and (7.8), the non-ROI bit rate R_n is computed as follows:

$$R_n = \frac{T_p}{M\ (1 + P_r\ (K - 1))} \tag{7.9}$$

At CTU level, the bit allocation for B-frames depends on the number of bits allocated per region and on the weights of CTUs of the same region. For CTU of index i of the ROI, the allocated bits are:

$$T_r(i) = \frac{T_r - T'_r}{\sum_{j \in I_r} w_r(j)}\ w_r(i) \tag{7.10}$$

where T'_r is the effective number of bits of already encoded CTUs of the ROI, I_r is the set of indexes of ROI CTU that have not yet been coded, and $w_r(i)$ is the weight of the current CTU of the ROI computed referring to Equation (1.7). The same process is applied independently to CTUs of the rest of the frame (non-ROI). In fact, if T is the effective number of bits used to code the current CTU, the following test is performed; if the encoded CTU is in the ROI, then $T_r = T_r - T$ else $T_n = T_n - T$.

Region independent rate control models

For B-frames, once the rate of each CTU is found, the QP is computed using the R-λ model. Our proposal separates the models of the different regions. Consequently, the model parameters of CTUs from the ROI r are independent from the ones of CTUs of the non-ROI n. In fact, we have two models; in ROI, using the effective number of bits per pixel $R_r(i)$ of each unit of index $i \in I_r$,

$$\lambda_r(i) = \alpha_r R_r(i)^{\beta_r} \tag{7.11}$$

and for CTUs from the non-ROI (of index $j \in I_n$), using the effective number of bits per pixel $R_n(i)$,

$$\lambda_n(j) = \alpha_n R_n(j)^{\beta_n} \tag{7.12}$$

The model parameters are then updated separately. For the ROI, the parameters α_r and β_r are updated referring to the original rate control algorithm [54], as follows:

$$\lambda'_r = \alpha_r {R'}_r^{\beta_r} \tag{7.13}$$

$$\alpha'_r = \alpha_r + 0.1(\ln \lambda_r - \ln \lambda'_r)\alpha_r \tag{7.14}$$

$$\beta'_r = \beta_r + 0.05(\ln \lambda_r - \ln \lambda'_r) \ln R'_r, \tag{7.15}$$

where α', β' and λ' are the updated values of α, β and λ. In Equation (7.13) and Equation (7.15), R'_r is the effective number of bits per pixel after encoding the unit. The same update process is used for the CTUs of the non-ROI.

QP and λ variation

The last modification compared to the reference algorithm consists in considering new clipping ranges for λ and QP, at CTU level. As we try to make independent QP computing for each region, the QP of the current CTU depends on the QP of the last CTU of the same region and the QP of the current frame. We allow a larger QP range than in the reference algorithm, to accommodate differences in quality between the ROI and the non-ROI. We define $\Delta\mathrm{QP_p} > 2$ and $\Delta\mathrm{QP_u} > 1$ that guarantees

$$\mathrm{QP_p} - \Delta\mathrm{QP_p} \leq \mathrm{QP_u} \leq \mathrm{QP_p} + \Delta\mathrm{QP_p} \tag{7.16}$$

$$\mathrm{QP_{u'}} - \Delta\mathrm{QP_u} \leq \mathrm{QP_u} \leq \mathrm{QP_{u'}} + \Delta\mathrm{QP_u} \tag{7.17}$$

where $\mathrm{QP_u}$, $\mathrm{QP_p}$ and $\mathrm{QP_{u'}}$ are respectively the QPs of the current CTU, the current picture and the previously encoded CTU of the same region. It is also possible to consider different clipping ranges for CTUs of different regions and use asymmetric clipping.

7.2.3 Extended version of the proposed ROI-based controller

Modifications have been introduced to our initial approach taking into consideration the evolution of the controller in HEVC test model 13 (HM.13). There are two main modifications in the new proposal: ROI bit allocation for inter coded frames uses a novel complexity metric and ROI bit allocation for intra coded frames at CTU level is introduced.

Region bit allocation for inter-frames

In the new version of the controller, the weight of a CTU is computed by Equation (4.28). Thus, in our updated ROI-based controller the weight of a CTU from the ROI of index i is expressed as follows,

$$w_r(i) = N \left(\frac{\lambda_{\mathrm{Pic}}}{\alpha_r} \right)^{\beta_r} \tag{7.18}$$

where α_r and β_r are the R-λ model parameters for CTUs of the ROI and λ_{Pic} is the current picture λ. This weight is then used to compute an initial target allocated bit rate $T_r(i)$:

$$T_r(i) = \frac{T_r \; w_r(i)}{\sum_{j \in I_r} w_r(j)} \tag{7.19}$$

The target allocated bits for a CTU $\widetilde{T}_r(i)$ takes into account $T_r(i)$, the allocated budget

for the rest of CTUs of the same region, the effective number of bits of already encoded units of the ROI T'_r and a smoothing window W fixed at 4 in our simulations:

$$\widetilde{T}_r(i) = T_r(i) - \frac{\left(\sum_{\substack{j \in I_r \\ j \geq i}} T_r(j) - (T_r - T'_r)\right)}{W} + 0.5 \tag{7.20}$$

The number of bits per pixel for a CTU of the ROI is then:

$$R_r(i) = \frac{\widetilde{T}_r(i)}{N} \tag{7.21}$$

Region bit allocation for intra-frames

At frame level, the refinement of the initial number of bits is done referring to Equation (4.21) then the K factor is considered to make ROI based budget repartition as represented in Equation (7.20) and compute T_r and T_n. At CTU level, the weight of a unit is its cost and is calculated by deriving the SATD as descibed by Equations (4.31) and (4.32). This weight is used to compute an initial target allocated bits $T_r(i)$ as in Equation (7.19). Then, the number of bits left to encode the i^{th} CTU $\widetilde{T}_r(i)$ is defined as:

$$\widetilde{T}_r(i) = (T_r - T'_r) + \frac{\left((T_r - T'_r) - \sum_{\substack{j \in I_r \\ j \geq i}} T_r(j)\right)\ (L_r - i)}{W} \tag{7.22}$$

Finally, the number of bits per pixel for an intra CTU of the ROI is:

$$R_r(i) = \frac{\widetilde{T}_r(i)\ w_I(i)}{N \sum_{\substack{j \in I_r \\ j \geq i}} w_I(j)} \tag{7.23}$$

7.3 Experimental results of R-λ ROI-based rate control

7.3.1 Experimental setting

First, we implemented the URQ ROI-based model described in Section 7.1 in HM.9 [97]. Second, we implemented the proposed rate control scheme proposed in Section 7.2 on HM.10 available on [98] and described in [99]. Then, we introduced the extended version on HM.13 encoder presented in [100] and available on [69] by taking into account the evolution of the controller. We evaluated the obtained results of each ROI-based rate control method. Then, we make comparative tests to evaluate the performance of the proposed methods.

Test conditions

To compute a binary map as represented in Fig. 7.2, we used face detection method described in Section 6.1. We introduce in HEVC software the Viola and Jones object

detection algorithm [89].

(a) Johnny (ROI represents 13%)

(b) Kristen&Sara (ROI represents 14%)

(c) FourPeople (ROI represents 10%)

Figure 7.2: Test sequences and ROI maps

Since videoconferencing applications require typically low coding delay, all pictures were coded in display order. Three different configurations have been used to test the first and the second ROI-based controller: All-B, All-I and an hybrid configuration that considers GOPs of B-frames and introduces an intra picture each second. In the first and the third algorithms (HM.10 and HM.9), I-frame bit allocation at CTU level has not been yet introduced, so, all the frames were considered as B-frames except the first one (I-frame), while, for the extended version of our code in HM.13, we tested all the configurations. We tested an all intra configuration to evaluate our algorithm in I-frames and a low delay configuration where all the frames are coded as B-pictures to evaluate the ROI-based algorithm in inter pictures. The CTU size is equal to 64x64 and different bit allocation approaches at frame level are tested. So, if the frame rate is equal to 60 frame per second, the intra period is then equal to 60. Here we use low-delay hierarchical prediction structures with groups of four frames (BBBB coding structure) and a CTU size equal to 64x64.

Three HD 720p sequences from class E have been tested: “Johnny”, “Kristen&Sara”, “FourPeople” [68]. As we can see in Fig. 7.2, the selected test sequences have typical videoconferencing content and different characteristics, like number of faces and ROI size. We used different bit rates, budget partitioning per-region and QP ranges to evaluate the performance of our approach.

Implementation and performed tests

The introduced modifications have been done mainly in rate control class of the reference softwares HM.9 [97], HM.10 [98] and HM.13 [69]. A reference test “Ref” is performed using the rate control algorithm described in [54] and improved in [65]. While evaluating the

URQ model the reference used is described in [64]. These first tests give us the reference performance: the ratio between ROI bit rate and non-ROI bit rate K, the bit budget used for encoding each region, the PSNR and the structural similarity (SSIM) index [14] of each region that goes from 0 to 100. Second, we activate all modified functions: we introduce a new bit repartitioning between regions by fixing a factor K and a large QP margin. Then we perform an evaluation test of our method that we note "New".

7.3.2 Performance of ROI-based controller in HM.10

Table 7.1 summarizes the results of the performed test at 128kbps and 256kbps. Both equal and hierarchical bit allocations are tested. The table shows that introducing a K factor for bit repartitioning between regions does not impair the rate-distortion performance. We can increase the effective ratio comparing to the reference by keeping an output bit rate close to the assigned value. Moreover, the overall PSNR is practically the same as the reference encoder.

Seq	Equal bit allocation											
	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR(dB)		ΔSSIM	
	Ref	New	Ref	New	Ref	New	Ref	New	ROI	non-ROI	ROI	non-ROI
Johnny	128.01	127.89	36.48	36.04	92.76	92.07	5.82	10.41	0.76	-0.40	0.60	-0.90
	256.01	255.80	39.17	38.72	94.96	94.53	6.11	9.89	0.53	-0.46	0.30	-0.55
Kristen &Sara	128.04	128.02	33.96	33.74	92.20	91.94	3.35	5.10	0.70	-0.77	0.69	-0.43
	256.08	256.06	37.04	36.75	94.50	94.33	3.25	4.67	0.61	-0.68	0.43	-0.27
Four People	128.05	128.06	31.47	31.28	88.26	88.03	4.41	6.67	0.61	-0.33	1.22	-0.42
	256.07	256.06	34.48	34.27	92.28	92.09	4.33	6.16	0.61	-0.35	0.87	-0.33
Seq	Hierarchical bit allocation											
	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR(dB)		ΔSSIM	
	Ref	New	Ref	New	Ref	New	Ref	New	ROI	non-ROI	ROI	non-ROI
Johnny	128.96	127.73	37.15	36.64	93.46	92.74	5.47	9.27	0.65	-0.38	0.45	-0.91
	256.01	255.84	39.48	39.20	95.21	94.91	5.95	9.62	0.66	-0.34	0.37	-0.41
Kristen &Sara	128.19	128.11	34.40	34.21	92.66	92.46	2.89	4.51	0.73	-0.85	0.63	-0.34
	256.32	256.23	37.36	37.18	94.77	94.66	3.00	4.43	0.62	-0.53	0.44	-0.21
Four People	128.01	129.05	31.75	31.56	88.80	88.54	4.30	7.06	0.70	-0.34	1.23	-0.45
	256.05	257.70	34.94	34.59	92.73	92.51	4.35	6.33	0.72	-0.40	0.89	-0.36

Table 7.1: Control accuracy comparison of the reference and the proposed controller for inter frames using HM.10

Now we examine the quality of ROI and non-ROI for different ratios K. In Table 7.1, ΔPSNR ROI is the difference in quality of the ROI using the proposed controller and the reference one and ΔSSIM ROI is the difference in similarity of the ROI using the proposed controller and the reference one (and the same for non-ROI). We notice that the overall quality of the ROIs is improved using different configurations but also different target rates. The global gain in the ROI goes from 0.5 to 0.7dB in terms of PSNR and from 0.3 to 1.2 in terms of SSIM. However, as we reduce the number of allocated bits in the non-ROI, its quality decreases.

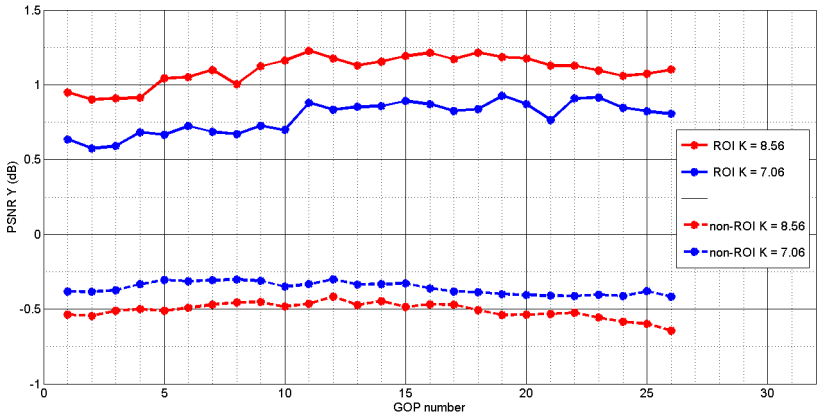


Figure 7.3: ΔPSNR ROI and non-ROI (dB) for the last 25 GOPs of FourPeople at 128kbps and using hierarchical bit allocation

In Fig. 7.3, we plot ΔPSNR of the ROI and ΔPSNR of the non-ROI per GOP. Overall, the bigger is K the better is the global quality of the ROI in the sequence and the lower is the PSNR of the non-ROI. The quality of the ROI is improved in all the GOPs (and frames) while the quality of the non-ROI is slightly decreased. The curves show that for each region the difference in quality between the proposed scheme and the reference RC [54] is more important when K is bigger. This means that our method leads to allocate more bits to the ROI by improving its quality and respecting the bit rate constraint.

(a) Reference scheme K=3.77 (b) Modified scheme K=5.89

Figure 7.4: Subjective comparison for "Johnny" coded at 100kbps

Experimental results show both advantages in objective (PSNR and SSIM) and subjective evaluation for ROI as represented in figure 7.4. We notice that using our proposed scheme we can distinguish more details in the face and less artifacts. However, our ratio can not reach relatively big values. The non-ROI does not represent noticeable deterioration, which means that the background requires a minimum coding budget to keep the balance.

7.3.3 Performance of ROI-based controller in HM.13

Intra picture ROI-based algorithm performance

Using an all intra configuration of the encoder, we tested the performance of the proposed algorithm. Three different rate points are used per sequence (640kbps, 1280kbps and 2560kbps). The budget constraint is respected and the global quality is not altered as shown in Table 7.2.

Seq	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR	ΔSSIM
	Ref	New	Ref	New	Ref	New	Ref	New	ROI (dB)	ROI
Johnny	640.00	639.99	28.78	28.90	83.61	83.70	2.60	2.90	0.35	0.94
	1280.04	1279.97	31.77	31.84	87.08	86.88	2.62	3.04	0.44	0.91
	2560.05	2559.93	34.84	35.00	91.44	91.00	2.41	2.88	0.74	1.04
Kristen &Sara	649.46	649.26	26.46	26.47	83.04	83.08	1.48	1.49	0.01	0.01
	1280.02	1280.07	29.31	29.44	87.27	87.17	1.21	1.31	0.40	0.65
	2560.02	2560.02	32.72	32.80	91.38	91.15	1.23	1.57	0.30	0.29
Four People	666.27	665.42	25.17	25.17	74.17	74.17	1.57	1.57	0.00	0.00
	1280.01	1279.88	27.10	27.09	78.99	78.80	1.40	1.32	-0.17	-0.35
	2559.98	2559.74	29.75	29.83	85.16	85.10	1.31	1.23	-0.19	-0.42

Table 7.2: Control accuracy comparison of the reference and the proposed controller for intra frames using HM.13

In intra case, units from the ROI are coded from other units of the non-ROI. Consequently, our novel bit repartition affects the non-ROI and so the ROI. the more affected units of the ROI are CTUs at the edge of the region of interest. Thus, the algorithm is working well when we have one big ROI. However, when we have multiple and small ROIs, important CTUs are more affected by the quality decrease of the non-ROI.

Inter picture ROI-based algorithm performance

A low delay B configuration is used to evaluate the performance of ROI-based allocation for B-frames. We first evaluate the global performance as done in HM.10. Results are given at 128kbps and 256kbps to compare the performance with the first version of the controller and equal, hierarchical and adaptive bit allocations are tested.

From Table 7.3, we can deduce the same conclusions as in the previous version of our controller implemented in HM.10: the bit budget constraint is respected and ROI quality is improved proportionally to the repartition factor K. Here again, we notice that, using different bit allocation approaches at frame level, the bigger is the K the better is the quality of the ROI and the lower is the quality of the non-ROI. Making a differential bit repartitioning improves the quality of the ROI. The effective K can be increased by introducing higher repartitioning factor and larger QP ranges.

Seq	Equal bit allocation											
	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR(dB)		ΔSSIM	
	Ref	New	Ref	New	Ref	New	Ref	New	ROI	non-ROI	ROI	non-ROI
Johnny	128.00	127.91	37.01	36.56	93.17	92.54	7.14	11.45	0.53	-0.69	0.39	-0.80
	256.01	255.82	39.49	39.01	95.20	94.75	6.57	11.01	0.53	-0.70	0.31	-0.56
Kristen &Sara	128.03	128.02	34.89	34.58	92.77	92.44	3.59	5.45	0.91	-0.59	0.81	-0.53
	256.05	256.01	37.75	37.46	94.84	94.60	3.15	5.21	0.89	-0.54	0.62	-0.38
Four People	128.03	128.03	32.36	32.13	90.07	89.83	4.71	7.51	0.84	-0.46	1.64	-0.48
	256.07	256.03	35.15	34.87	93.13	92.89	4.17	6.96	1.03	-0.53	1.32	-0.44

Seq	Hierarchical bit allocation											
	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR(dB)		ΔSSIM	
	Ref	New	Ref	New	Ref	New	Ref	New	ROI	non-ROI	ROI	non-ROI
Johnny	128.01	127.94	37.36	36.88	93.60	92.91	6.99	10.33	0.56	-0.75	0.43	-0.87
	256.01	256.32	39.74	39.26	95.40	94.98	6.94	10.72	0.55	-0.71	0.32	-0.54
Kristen &Sara	128.09	128.10	35.13	34.89	93.30	92.75	3.29	4.99	0.92	-0.50	0.79	-0.47
	256.10	256.08	37.91	37.65	95.01	94.80	3.19	4.92	0.92	-0.51	0.64	-0.35
Four People	128.02	128.27	32.58	32.35	90.42	90.15	5.16	7.34	0.88	-0.45	1.69	-0.52
	256.03	254.80	35.43	35.10	93.46	93.20	4.77	6.86	0.93	-0.55	1.14	-0.43

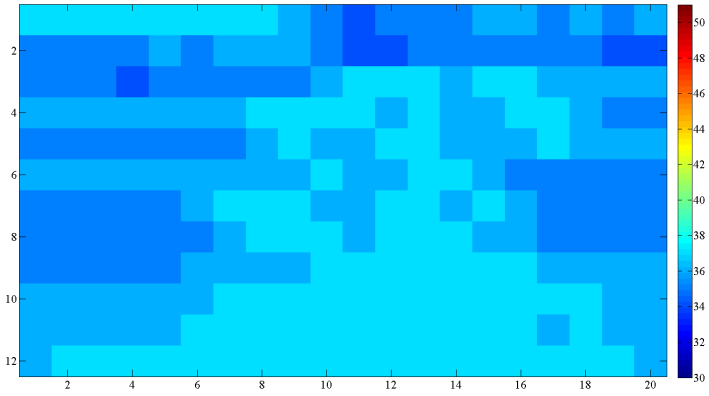
Seq	Adaptive bit allocation											
	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR(dB)		ΔSSIM	
	Ref	New	Ref	New	Ref	New	Ref	New	ROI	non-ROI	ROI	non-ROI
Johnny	128.00	127.87	37.48	37.00	93.74	93.05	6.53	9.93	0.54	-0.76	0.37	-0.86
	256.00	255.41	39.84	39.35	95.48	95.07	6.86	10.59	0.55	-0.73	0.33	-0.53
Kristen &Sara	128.05	128.09	35.21	34.96	93.14	92.84	3.19	4.87	0.94	-0.52	0.79	-0.48
	256.07	256.03	37.95	37.71	95.06	94.87	3.19	4.89	0.93	-0.50	0.63	-0.33
Four People	127.98	127.45	32.66	32.44	90.57	90.30	5.08	7.42	0.85	-0.43	1.47	-0.49
	255.98	253.94	35.50	35.17	93.55	93.28	4.85	6.93	0.90	-0.54	1.04	-0.43

Table 7.3: Control accuracy comparison of the reference and the proposed controller for inter frames using HM.13

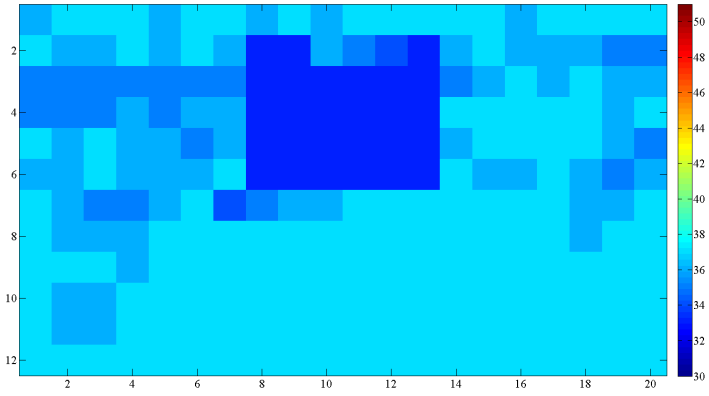
At CTU level the proposed approach gives a new QP distribution. Fig. 7.5 represents the ROI map of Johnny at CTU level, the QP partitioning at CTU level using HM.13 reference rate control algorithm and the QP partitioning when using our algorithm. The encoding of the given result is done at 128kbps. Fig. 7.5(c) shows that smaller QP values are assigned to Johnny's face ($QP = 30$), while, the rest of the frame takes bigger QPs that go from 34 to 38.

Proposed algorithm has shown improvement in ROI quality for both intra- and inter-coded frames. Thus, next step consists in evaluating an hybrid configuration that considers both I and B frames.

(a) ROI map



(b) Reference RC



(c) Proposed RC

Figure 7.5: Comparison of QP repartition at CTU level of Johnny

ROI-based algorithm performance using hybrid configuration

For a videoconferencing systems a low delay configuration is the most appropriate as we have the real time constraint. However, to reduce packet loss effect and limit error propagation,

an intra frame is introduced every second. Consequently, the final configuration of our encoder is the hybrid one. It handles a GOP of four B-frames coded in display order and an I-frame after 60 inter pictures. We choose the adaptive bit allocation at frame level as it gives the best RD performance and we tested four different rate points per sequence (128kbps, 256kbps, 512kbps and 1500kbps).

From Table 7.4, we conclude that the controller global performance is maintained and the quality of the ROI is improved. At low bit rate, we can gain up to 2dB in the ROI. Moreover, SSIM of the ROI is improved considerably when picture SSIM is smaller than 95. We can reach an improvement in the ROI quality of 3.18 dB for example. As SSIM is saturated when it gets close to 100, ΔSSIM is reduced when the picture index is higher than 95. We still in that case have noticeable improvement in ROI quality as the SSIM index goes from 0.20 to 0.92.

Seq	Bit rate (kbps)		PSNR Y (dB)		SSIM		K		ΔPSNR (dB)		ΔSSIM	
	Ref	New	Ref	New	Ref	New	Ref	New	ROI	non-ROI	ROI	non-ROI
Johnny	128.00	128.02	35.95	36.11	92.59	92.25	5.50	9.06	1.59	-0.38	1.50	-0.56
	256.00	255.90	39.00	38.90	94.97	94.74	6.60	9.97	0.69	-0.33	0.30	-0.31
	512.01	511.34	41.09	40.84	96.16	96.86	6.58	10.71	0.48	-0.41	0.20	-0.24
	1500.01	1492.79	42.81	42.62	96.96	96.86	4.88	11.70	0.68	-0.35	0.26	-0.16
Kristen &Sara	129.86	128.18	33.21	33.72	91.76	91.93	2.71	4.30	1.91	-0.07	1.83	-0.11
	256.07	256.10	36.87	36.91	94.38	94.37	3.03	4.72	1.48	-0.34	1.06	-0.21
	512.07	512.00	39.76	39.60	95.97	95.89	3.03	4.76	0.95	-0.41	0.54	-0.18
	1500.10	1496.62	42.61	42.42	97.13	97.07	2.43	4.97	0.75	-0.39	0.34	-0.13
Four People	129.57	128.05	30.52	31.15	88.60	88.54	5.30	7.43	2.03	-0.22	3.18	-0.47
	256.00	255.48	34.29	34.26	92.64	92.33	5.02	7.10	1.46	-0.39	1.86	-0.59
	511.97	509.40	37.58	37.30	95.18	94.90	4.55	6.65	1.02	-0.55	0.92	-0.43
	1499.97	1484.96	41.46	41.18	97.05	96.93	3.87	6.70	0.78	-0.47	0.37	-0.18

Table 7.4: Control accuracy comparison of the reference and the proposed controller in HM.13

Experimental results show advantages in objective PSNR, in SSIM that predicts subjective opinion with high precision and visual evaluation for ROI as represented in Fig. 7.6, Fig. 7.7, Fig. 7.8, Fig. 7.9, Fig. 7.10 and Fig. 7.11. We notice that for both intra and inter pictures and using our proposed scheme we can distinguish more details in the face and less artifacts, while the non-ROI does not present noticeable deterioration in visual quality as in videoconferencing system the background is not changing in most of the cases.

Locally the SSIM index has been evaluated and an SSIM map has been computed for each frame to prove quality improvement in the ROI. Fig. 7.12, Fig. 7.13 and Fig. 7.14 represent the SSIM index over the whole frames (SSIM values goes from 0 for high distortion to 1 for high similarity). We notice that considering the proposed method SSIM index in the faces is closer to 1 (white faces). It shows an improvement in the details of the faces of the three tested sequences.

(a) Reference RC

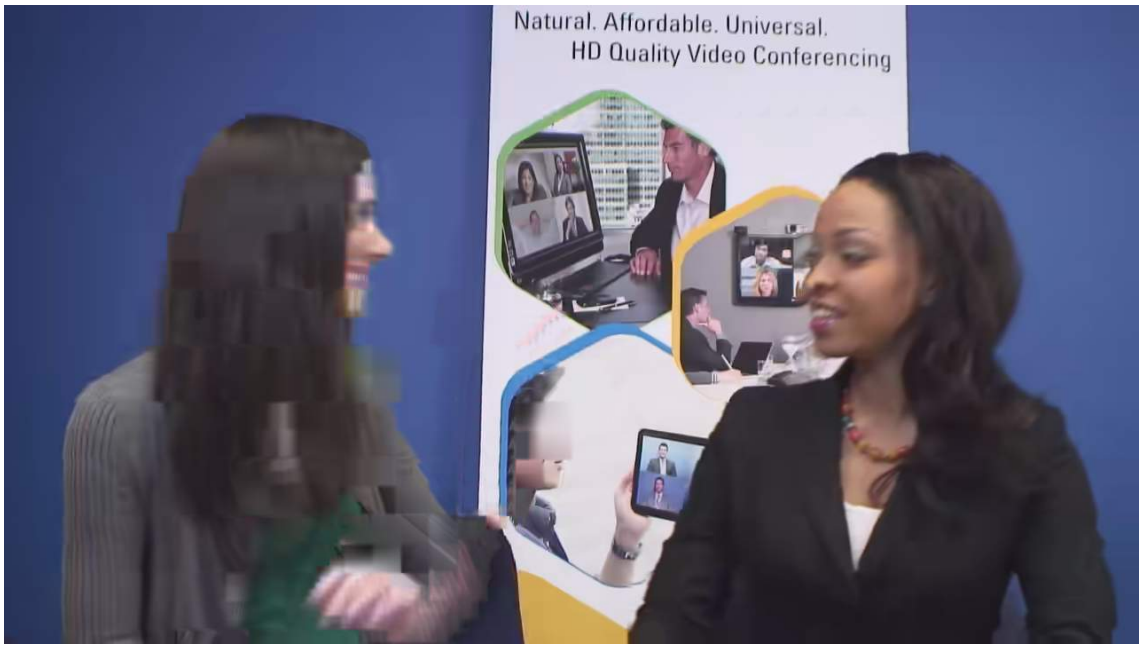
(b) Proposed RC

Figure 7.6: Subjective comparison of Johnny coded at 128kbps for an I frame

(a) Reference RC

(b) Proposed RC

Figure 7.7: Subjective comparison of Johnny coded at 128kbps for a B frame

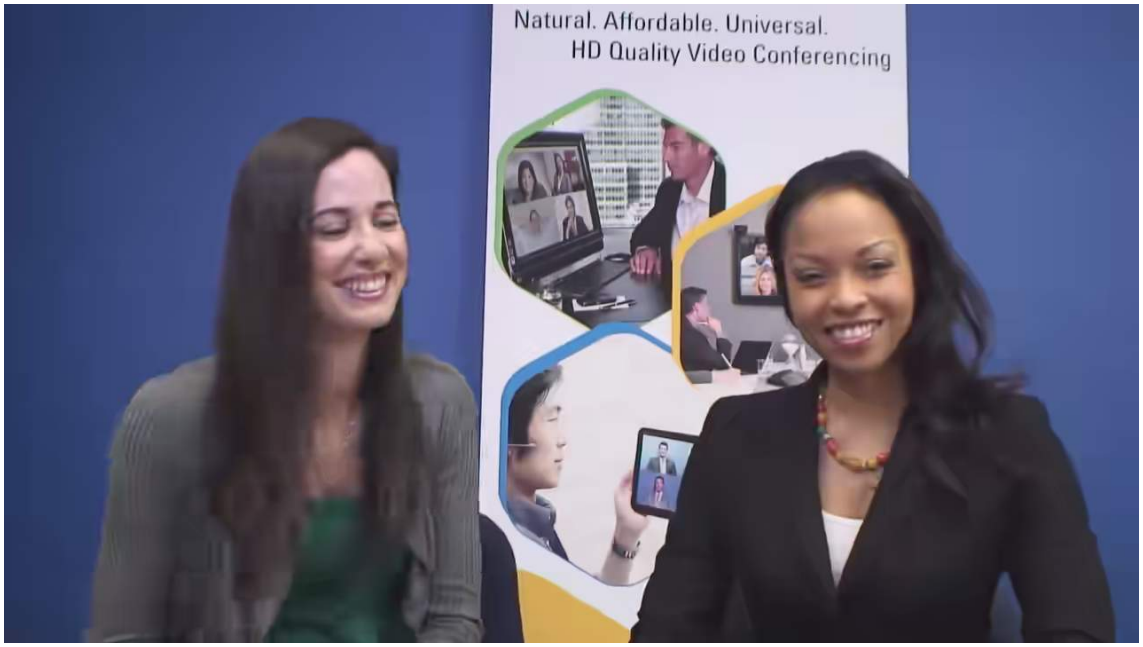


(a) Reference RC

(b) Proposed RC

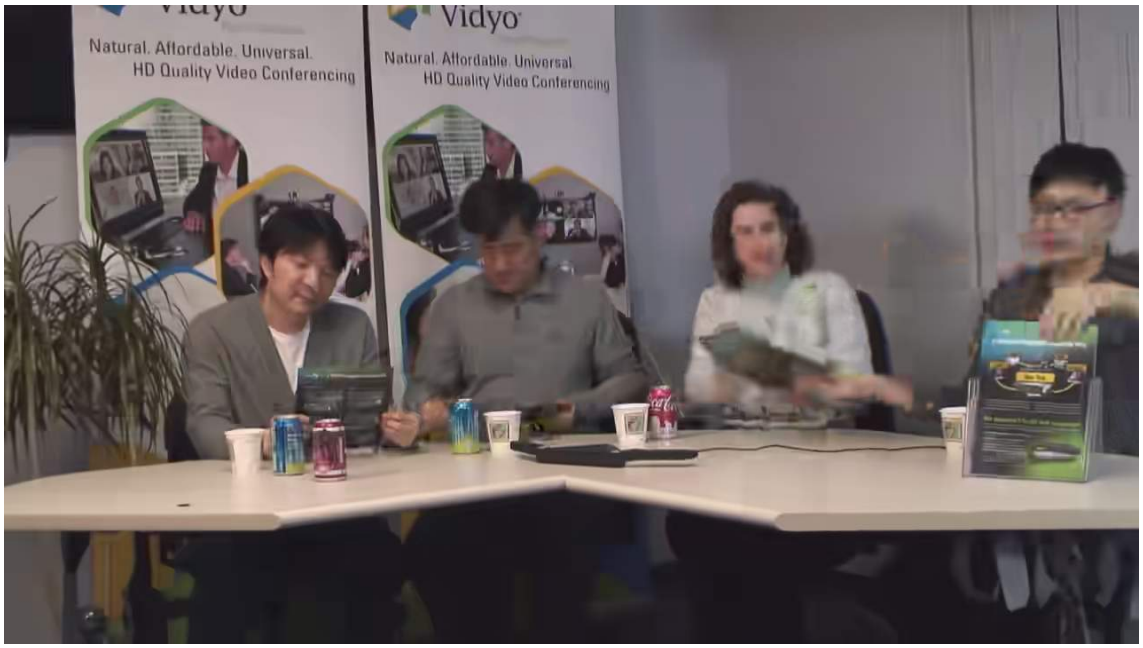
Figure 7.8: Subjective comparison of Kristen&Sara coded at 128kbps for an I frame

(a) Reference RC

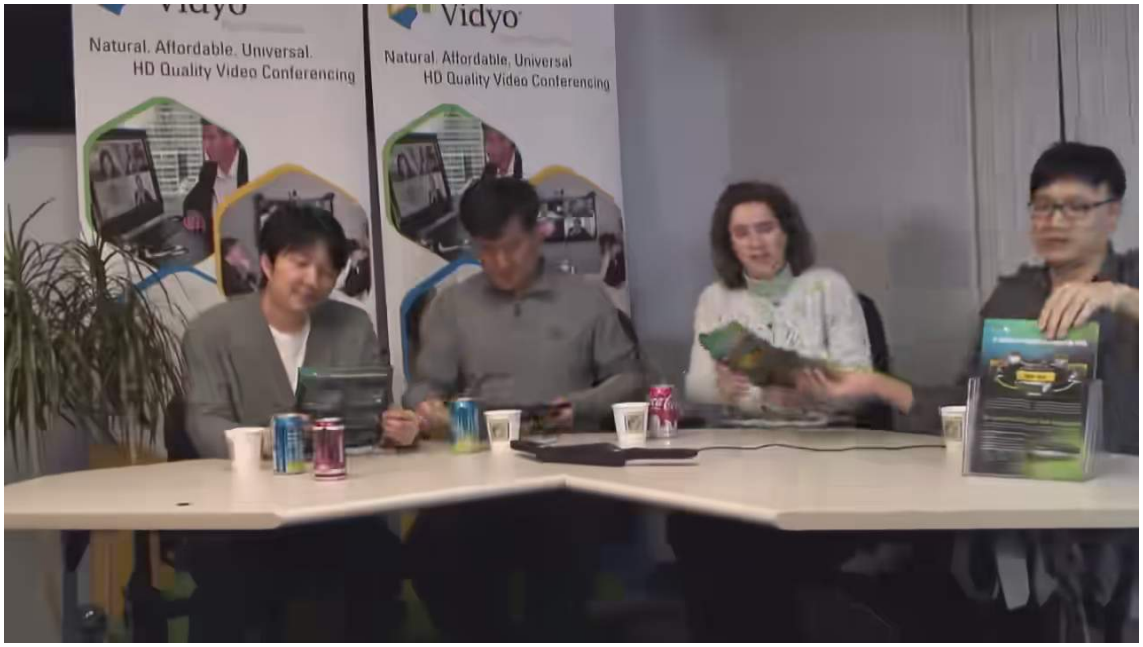


(b) Proposed RC

Figure 7.9: Subjective comparison of Kristen&Sara coded at 128kbps for a B frame

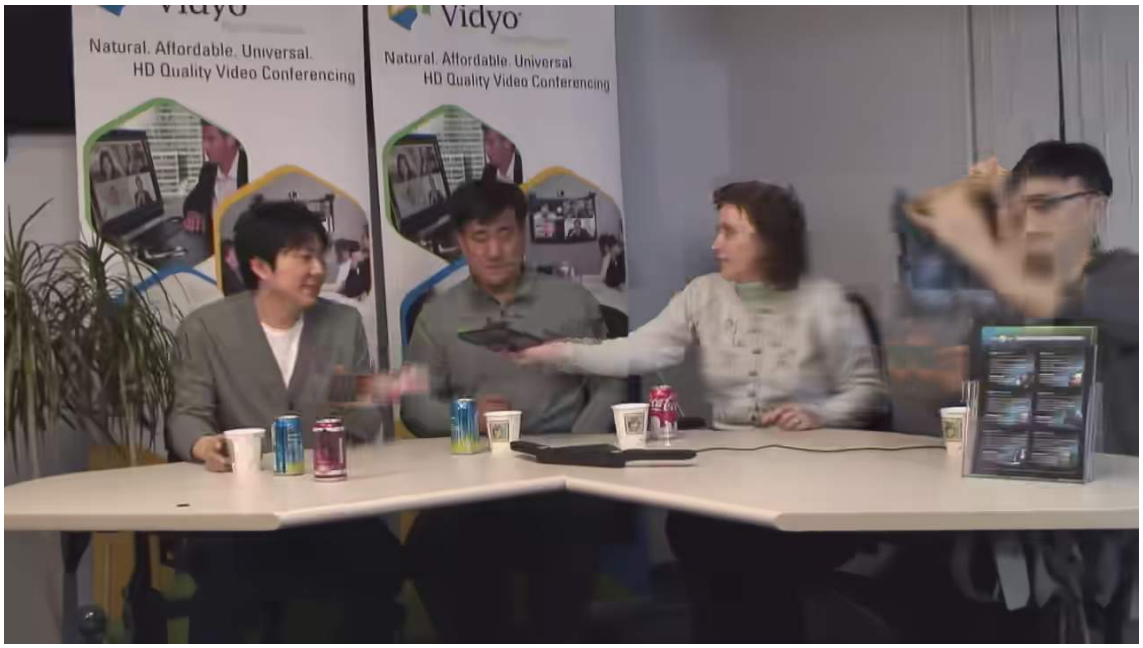


(a) Reference RC

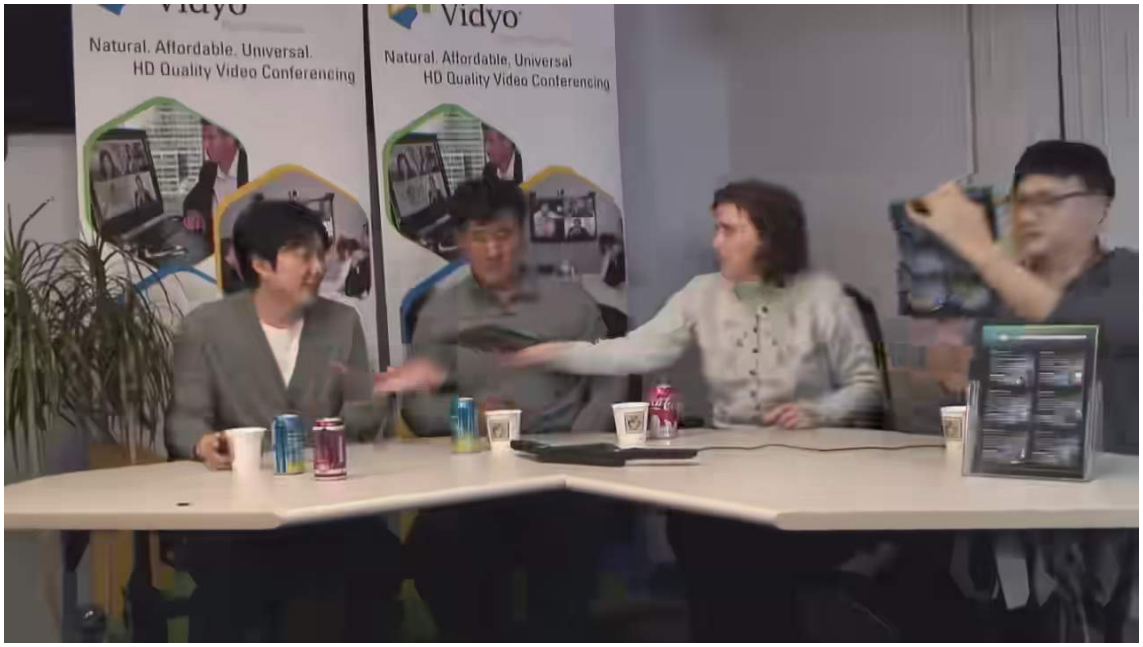


(b) Proposed RC

Figure 7.10: Subjective comparison of FourPeople coded at 128kbps for an I frame



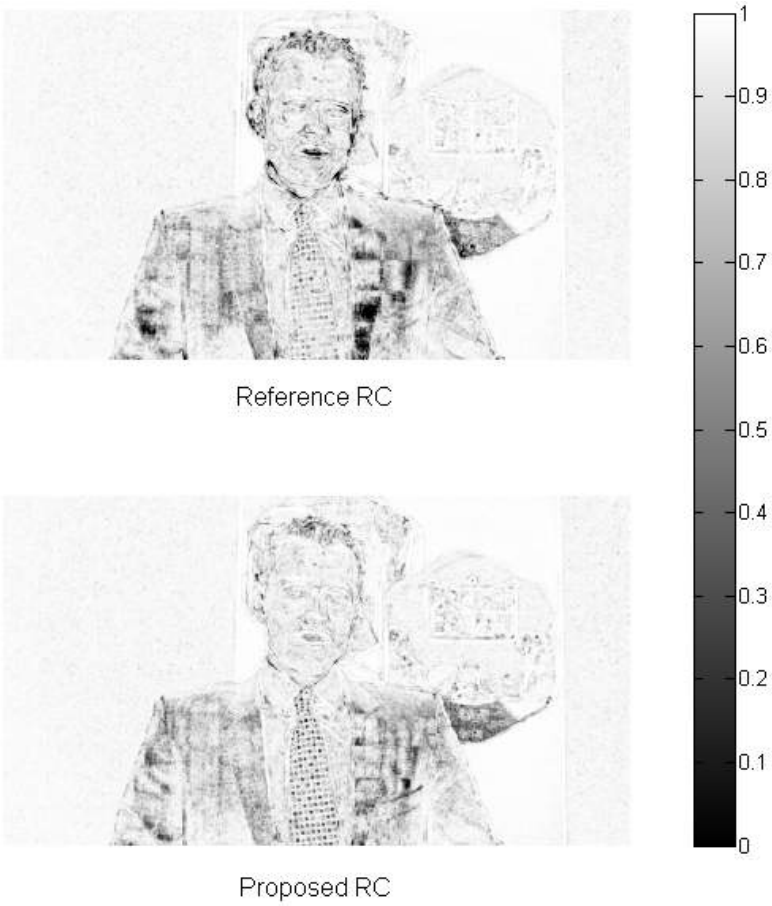
(a) Reference RC



(b) Proposed RC

Figure 7.11: Subjective comparison of FourPeople coded at 128kbps for a B frame

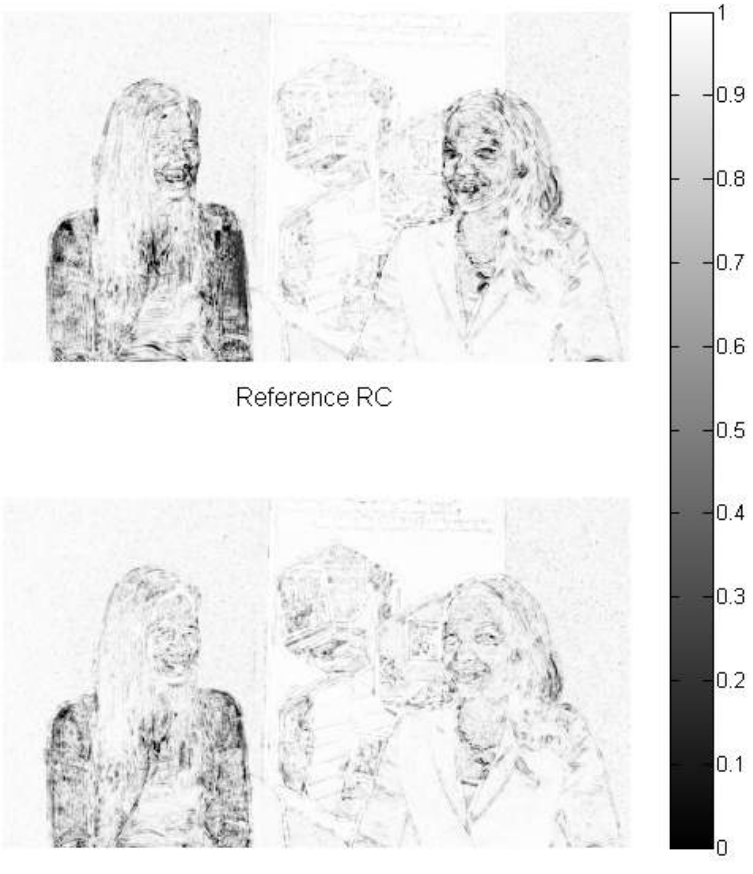
(a) Original frame



(b) SSIM maps

Figure 7.12: SSIM map comparison Johnny

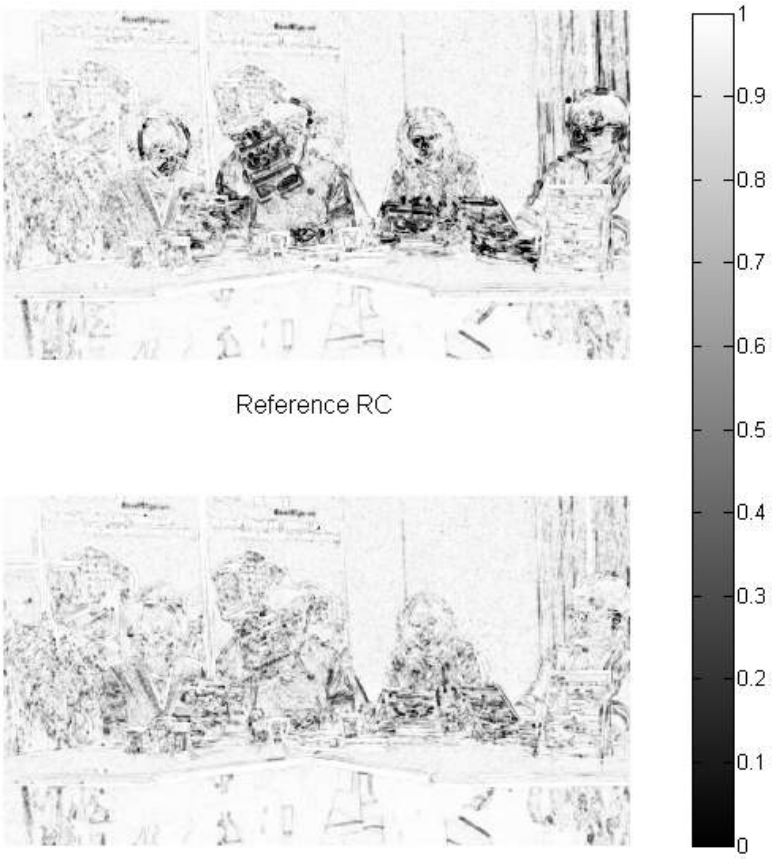
(a) Original Frame



(b) SSIM maps

Figure 7.13: SSIM map comparison Kristen&Sara

(a) Original Frame



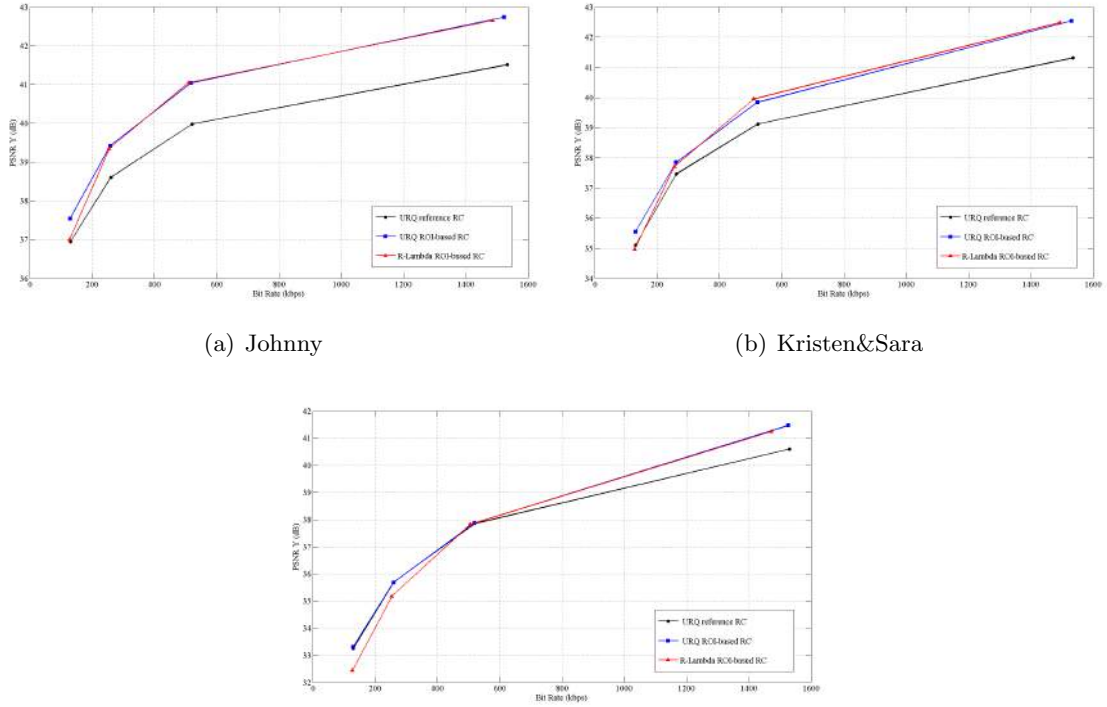
(b) SSIM maps

Figure 7.14: SSIM map comparison FourPeople

7.3.4 Comparison with quadratic model

The last experiment consists in comparing the performance of our algorithm to a stat-of-the-art approach. The used reference method is a ROI-based RC algorithm initially proposed for H.264/AVC and based on the quadratic model which we adapt to HEVC as described in Section 7.1. The performed tests in this section use a low delay configuration with all frames are coded in bidirectional mode (B-frames). We tested the three sequences at four different bit rates (128kbps, 256kbps, 512kbps and 1500kbps).

We notice from Fig.7.15 that the URQ ROI-based method implemented in HM.9 respects the budget constraint at both low bit rate and high bit rate.



(a) Johnny

(b) Kristen&Sara

(c) FourPeople

Figure 7.15: RD performance of R-λ ROI-based algorithm and URQ ROI-based model compared to URQ reference RC algorithm

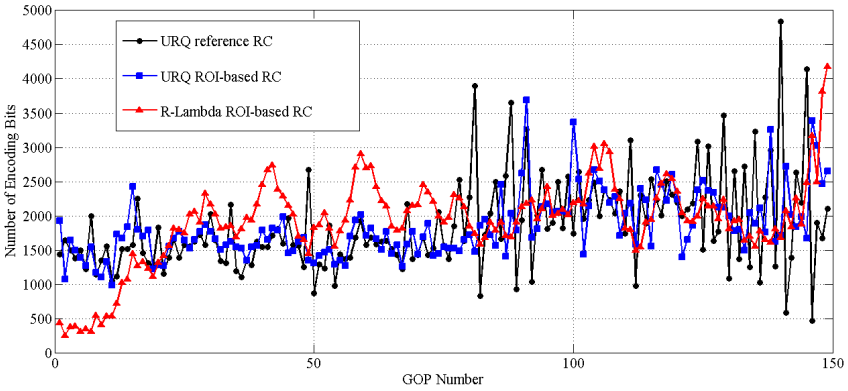
RD performance evaluation shows an important improvement in rate control performances. Introducing a K factor improves bits partitioning in different regions of the frame, which leads to an improvement in the quality of the whole sequence. The obtained RD curve is better than the reference URQ model and comparable to the one given by our R-λ algorithm implemented in HM.13. Moreover, Table 7.5 shows that URQ ROI-based method improves the quality of the ROI while using higher bit ratio K.

Fig. 7.16 shows bit distribution over GOP at low and high bit rates for Johnny sequence. We conclude that the proposed R-λ method gives a smoother bit allocation compared to

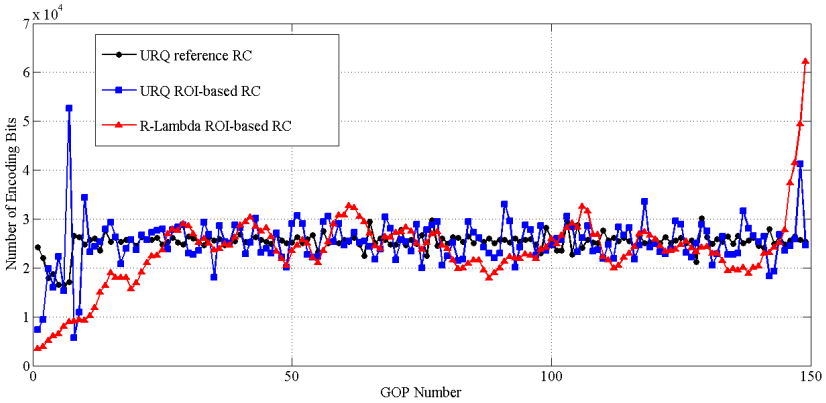
Seq		K	Bit rate (kbps)	PSNR Y (dB)	SSIM ROI	PSNR ROI (dB)	SSIM	PSNR non-ROI (dB)	SSIM non-ROI
Johnny	Ref	3.93	130.48	36.94	93.66	32.60	94.18	38.28	93.59
	New	7.90	129.19	37.54	93.59	35.25	95.21	38.04	93.39
		8.54	129.08	37.56	93.59	35.45	95.25	38.00	93.38
Kristen &Sara	Ref	2.13	130.92	35.10	93.10	31.74	93.75	36.04	93.02
	New	4.26	130.73	35.54	93.24	34.02	94.35	35.85	93.10
		4.59	130.65	35.60	93.30	34.29	94.49	35.87	93.15
Four People	Ref	4.31	129.87	33.25	91.38	29.92	82.86	33.94	92.46
	New	6.08	129.69	33.30	91.38	30.68	84.30	33.80	92.28
		6.42	129.94	33.28	91.36	30.87	84.67	33.73	92.21

Table 7.5: Rate control results using URQ model at 128kbps

the URQ methods at low bit rate with no unsettled bit picks, while at high bit rate the three algorithms gives comparable distribution over GOPs. The same conclusion is valid for all tested sequences.



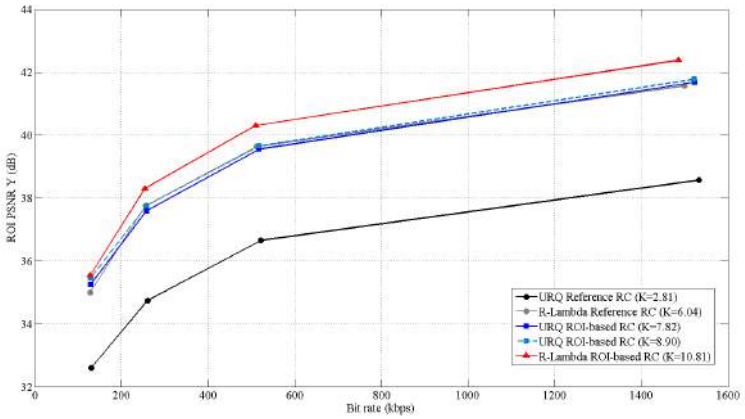
(a) 128 kbps



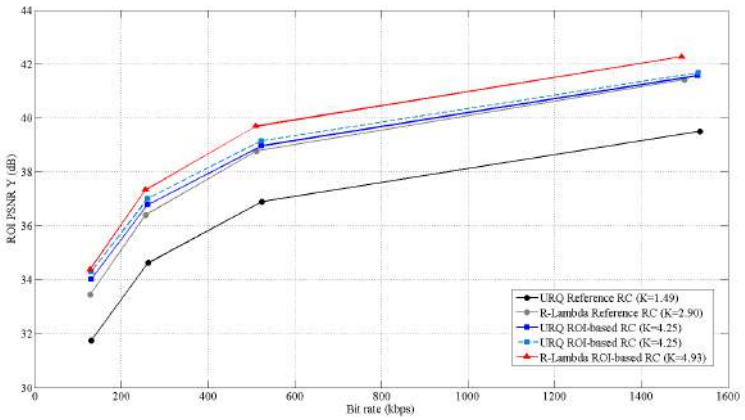
(b) 1.5 Mbps

Figure 7.16: Comparison of bit fluctuation per GOP of R-λ and URQ ROI-based models at low and high bit rate for sequence Johnny

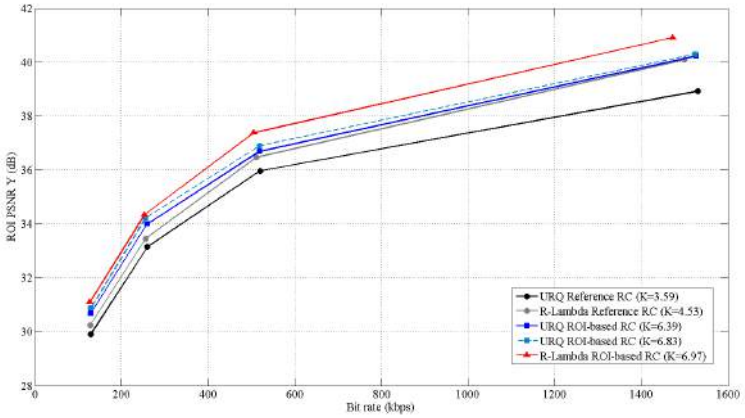
Fig. 7.17 represents RD performance of all evaluated methods. It gives the overall ROI PSNR for each bit rate. For the three tested sequences, the reference URQ controller has the worst RD performances. Once introducing the ROI, both URQ-based method and R-λ-based method show better RD performance compared with the reference.



(a) Johnny



(b) Kristen&Sara



(c) FourPeople

Figure 7.17: Comparative ROI-based RD performance of different methods

Finally, in the URQ scheme, ROI-based bit allocation is only performed for referenced frames of type B. Our algorithm (based on R-λ method) makes ROI-based allocation for all frame types, which leads to a better QP repartition over regions in the full sequence. With our algorithm we can reach higher ratios K, as shown in Fig. 7.17.

7.4 Conclusion

We have presented in this chapter ROI-based rate control methods for HEVC. In Section 7.1, a method taken from the literature has been studied adapted to HEVC and implemented in HM.9. This first algorithm that uses a ROI-based quadratic R-Q model for QP computing in referenced inter pictures has shown improvement in ROI quality comparing to the reference quadratic URQ controller. Then, in Section 7.2, we present our ROI-based rate control method. Novelty consists in using the R-λ model for computing QPs of CTUs of different regions, performing rate control in both intra- and inter-coded frames and making independent bit allocation between ROI and non-ROI. The proposed algorithm has been initially introduced in HM.10 then improved in a second version implemented in HM.13. It shows important gain in ROI quality while respecting the global bit rate constraint.

Section 7.3 of this chapter has detailed obtained results of different encoder configurations. We conclude that activating ROI-based rate control helps improve ROI quality considering differentiated bit allocation between regions and independent R-Q models. Moreover, compared to the quadratic R-Q model, the R-λ model offers higher quality increase in ROIs and better global RD performance. This work has been published in [95] and [96]. However, this approach presents some limitations. In fact, improving the ROI quality while respecting the total budget decreases the non-ROI quality. Knowing that all CTUs are dependent, in some cases the decrease of non-ROI quality may affect the quality of the ROI. Furthermore, at transport layer a loss of any unit of the frame leads to the loss of the whole frame and dependent ones. Consequently, in next chapter we introduce tiles to perform rate control over independently decodable regions and reduce error propagation.

Chapter 8

Tiling for ROI-based Rate control

Contents

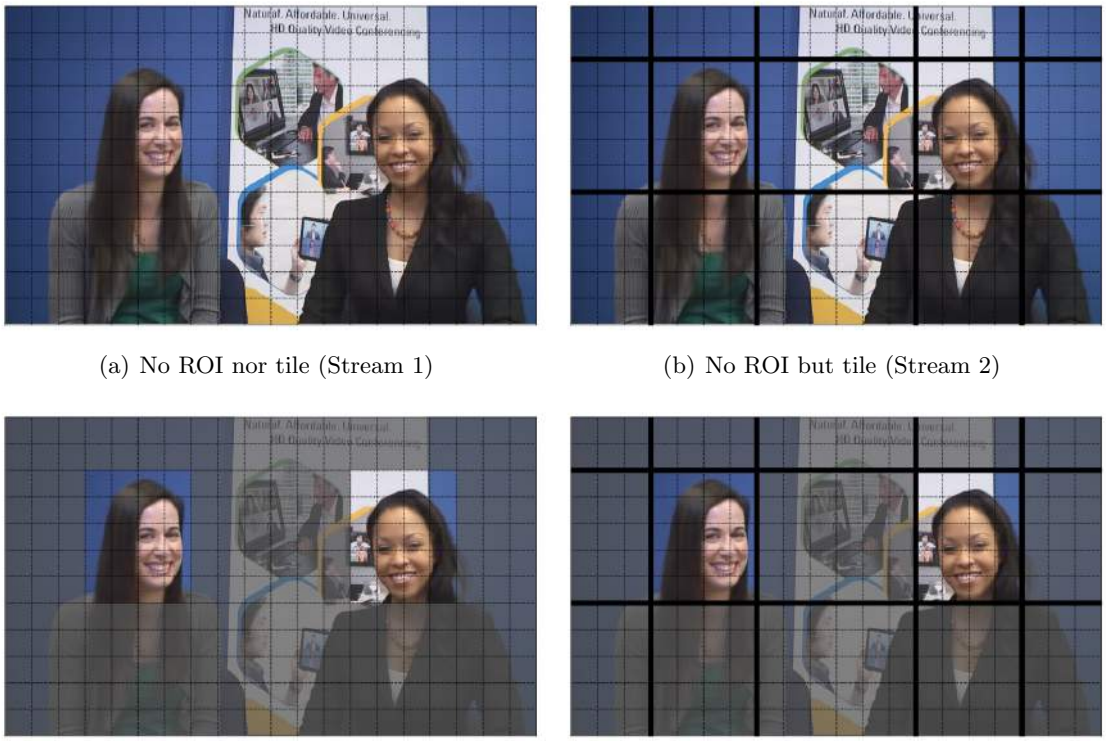
Tiling is a new feature introduced in the HEVC standard to ensure picture partitioning into independently decodable rectangular regions. As described in Chapter 2, tiles increase the capability of parallel processing and facilitate encoding based on the region-of-interest. Thus, tiles can be an interesting tool to separate ROIs from non-ROIs when encoding the frame and can ensure independent rate control over regions.

This chapter describes a new approach in ROI-based rate control that takes into account tiling for region partitioning to ensure an improvement of ROI encoding and transmission over the network. The aim is to reduce error propagation inside a frame and limit it to the affected region. The first section presents the main features of the proposed method at both video coding layer (VCL) and network abstraction layer (NAL). In the second section, performed tests are detailed and obtained results are analyzed. We end with a conclusion and an evaluation of the proposed tile- and ROI-based rate control algorithm.

8.1 Tile- and ROI-based controller for HEVC

8.1.1 Possible rate control configurations

Depending on the controller requirements, we can propose different partitioning configurations based on two features: ROIs and tiles. When ROI detection is activated, ROI-based bit allocation is performed as represented in Fig.7.1. In addition, dividing the frame in tiles gives independently decodable regions. Fig.8.1 represents possible configurations for "Kristen&Sara" sequence encoding.



(a) No ROI nor tile (Stream 1)

(b) No ROI but tile (Stream 2)

(c) ROI but no tile (Stream 3)

(d) ROI and tile (Stream 4)

Figure 8.1: Possible rate control configurations

In our work, we studied four possible configurations:

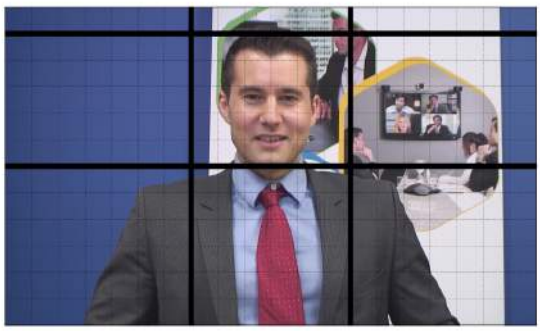
- "No ROI nor tile": Configuration (a) is used when no ROI-based processing is needed. The frame is divided into CTUs to perform bit allocation per unit. We used this configuration when testing the reference controller of HEVC represented in Fig.4.1.
- "No ROI but tile": Configuration (b) is also used when no ROI-based processing is needed. Tiles delimit independently decodable regions for parallel processing. In this case tiles are not considered when performing rate control.
- "ROI but no tile": To perform ROI-based rate control it is important to delimit the region of interest as represented in (c). More bits are allocated for the ROI and

independent R-Q models are used for QP computing. This configuration has been used in all performed tests of the previous chapter (Chapter 7). As tiling is not considered all units are dependently decodable.

- "ROI and tile": Considering both ROIs and tiles to perform bit allocation is the novelty introduced by our work and described in this chapter (used configuration is (d)). In fact, ROI detection is used to perform differentiated bit allocation over regions at VCL layer and transmit independently decodable tiles of different regions.

8.1.2 Rate control at Video coding layer

At the VCL layer, we proposed a tile partition of the different sequences of class E ("Johnny", "Kristen&Sara", "FourPeople") to identify faces as in the example represented in Fig.8.2. Tiles containing faces are classified as ROIs. The tested sequences has different characteristics for example one tile is lying within the ROI for "Johnny", two tiles for "Kristen&Sara" and four tiles for "FourPeople".



(a) Johnny (9 tiles)

(b) Kristen&Sara (15 tiles)

(c) FourPeople (24 tiles)

Figure 8.2: Tile partitioning of tested sequences

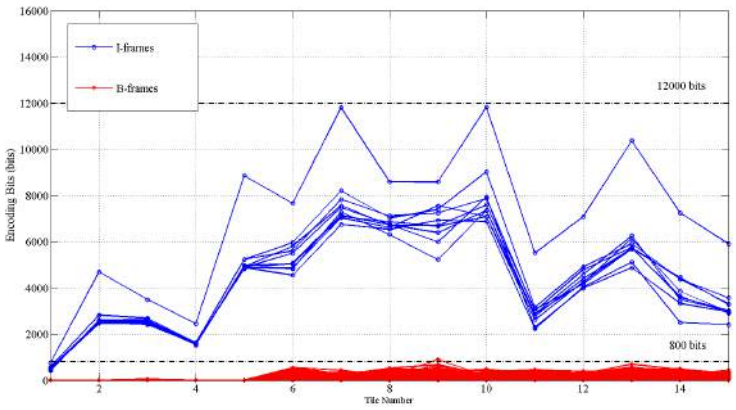
The structure of the proposed HEVC controller is represented in Fig.7.1. As explained in the previous chapter, we introduce a new level for region bit allocation and QP computing. At this level called "region level", the number of bits allocated per frame is allocated between

the two regions, considering a factor K as defined in equation (7.6). These bit budgets of the ROI and non-ROI are used independently to compute the number of encoding bits of CTUs of each region. Two independent R-λ models are then used for ROI and non-ROI. Thus, the allocated bits per CTU is used as input to the RD model of the corresponding region, to assign a QP per unit.

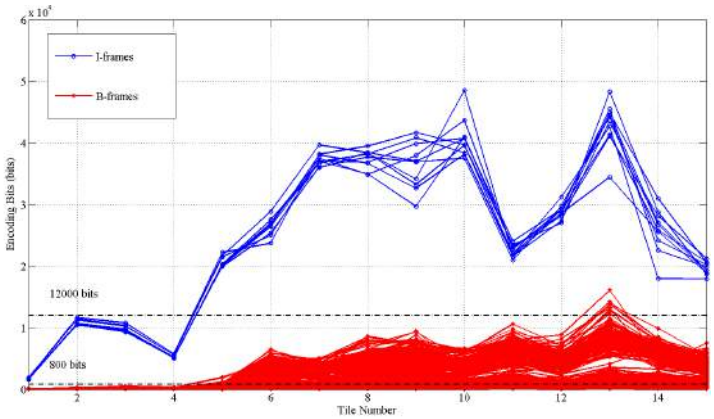
The novelty in the current work is not only independent rate allocation but also independent transmission and decoding of the ROI and the non-ROI which overcome the limitation of our ROI-based rate control algorithm. The ROI covers separate tiles from the non-ROI. Consequently, they are not affected by the quality decrease of the non-ROI when performing ROI-based bit allocation.

8.1.3 Adaptation at Network abstraction layer

Fig.8.3 represents the number of allocated bits per tile (15 tiles in the given example), while encoding "Kristen&Sara" sequence at low (a) and high (b) bit rates using the reference controller and tiling ("No ROI but tile" configuration represented in Fig.8.1(b)).



(a) 128 kbps



(b) 1.5 Mbps

Figure 8.3: Number of encoding bits per tile ("Kristen&Sara" sequence) at low and high bit rates

Each line corresponds to an encoded frame (600 frames per sequence). The figure shows that the number of bits to encode different tiles (for both I and B frames) is not homogeneous, as tiles have different sizes in the proposed partitioning. Moreover, the number of bits per tile may exceed the MTU size of the network.

By matching tiles and slice segments it is possible to divide each tile into data streams to fit the MTU size (12000 bits for IP network and 800 bits for wireless environment) [101]. Thus, tiles of the ROI and tiles of the rest of the frame (non-ROI) would be encapsulated in different NAL units. The video stream would contain two kinds of NALs to transmit in the channel. If we consider the partitioning presented in Fig.8.2(b) and if the number of bits allocated per tile does not exceed the MTU size, tile number 6 and tile number 7 will be encapsulated in separate NALs as illustrated in Fig.8.4.

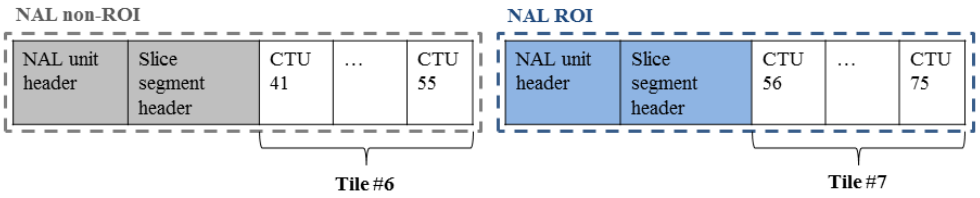


Figure 8.4: NAL unit formats

8.1.4 Packet loss and error concealment algorithm

Channel characteristics

For modeling the error characteristics of a wireless channel between two stations, a simple and widely used model is adopted, the Gilbert-Elliott model [1]. It considers two Markov chain with a good (G) and a bad (B) state, see Fig.8.5. Every state has a specific constant bit error rate, e_G in the good state and e_B in the bad one. The bit error in general depends on environmental conditions. Furthermore, the state transitions are determined by the values $1 - p_{(G \to B)}$ (for the probability that the next state is the be good state given that the current state is also good) and $1 - p_{(B \to G)}$ (for the probability that the next state is the be bad state given that the current state is also bad).

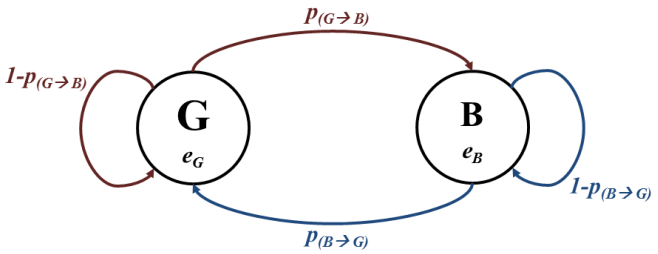


Figure 8.5: A two state Markov channel

Each NAL unit is then packetized and transmitted over Gilbert-Elliott channels. For our experiments we use parameters represented in Table 8.1.

$1-p_{(G\to B)}$	$1-p_{(B\to G)}$	e_G	e_B
0.995	0.96	10^{-4}	10^{-3}

Table 8.1: Gilbert-Elliot model parameters [1]

Error propagation and concealment

Two algorithms are proposed and compared to show the effectiveness of ROI- and tile-based controller:

- No ROI nor tile based coding (Configuration 8.1(a)): If one packet is lost, all the following packets of the same frame are lost and all dependent frames are not decoded. Considering temporal dependencies between successive images, error is propagated till the next intra frame. At the decoder side, each corrupted frame is replaced with the last correctly decoded frame.
- ROI- and tile-based coding (Configuration 8.1(d)): Since packets of the same tile are dependent, we decide that if one packet of a tile is lost, all packets of the same tile are considered as lost. Considering temporal dependencies between tiles in the same spatial position, error is propagated inside tiles of the same position till the next intra frame. Consequently, at the decoder side, each corrupted or lost tile is replaced with the tile of the last decoded frame and at the same spatial position. This is a simple way to conceal errors in corrupted streams.

8.2 Experimental results

The proposed algorithm has been implemented in HM.13. We tested class E sequences ("Johnny", "Kristen&Sara", "FourPeople") with the resolution of 1280 ×720 pixel, a frame rate equal to 60 fps and 600 frames per sequence [68]. We used a low delay configuration as the algorithm is designed for videoconferencing systems. We used open GOPs of size 4 and an intra period equal to 60 to limit temporal error propagation.

This section evaluates the impact of introduced features; ROI-based bit allocation and tiling. First, we study the impact of the K factor in the RD performance the encoder and visual quality of the ROI. Then, we analysis the effect of tiling in the visual quality of the sequence. We end up with a study of the performance of the proposed ROI- and tile-based rate control algorithm in a lossy network. To do so four streams are encoded using the four proposed configurations and their appropriate controllers:

- Stream 1 is the reference stream. It is obtained when using R-λ reference controller where no ROI neither tile are considered (Fig. 8.1(a)).

- Stream 2 is obtained when the R-λ reference controller is used without considering ROIs for bit allocation. However, tile partitioning is performed for independently decodable regions (Fig. 8.1(b)).
- Stream 3 is obtained with ROI-based R-λ controller proposed in the previous chapter. Tiling is not activated. All regions are dependently decodable (Fig. 8.1(c)).
- Stream 4 is the final stream. It represents the result of the proposed ROI- and tile-based R-λ controller (Fig. 8.1(d)).

8.2.1 Impact of the K factor in the RD performance

First test consists in evaluating the impact of the K factor in the RD performance when tiling option is activated by comparing visual quality and bit cost of stream 2 and stream 4.

Table 8.2 gives the RD performances for the three tested sequences and at low (128 kbps) and high (1.5 Mbps) bit rates, at various K factors, together with the improvement in the ROI quality. One can also remark that the budget limit is respected with good accuracy. Moreover, for repartitioning factors K bigger than the reference (gray lines represent stream 2), the quality of the ROI is improved. At low bit rates, we can have an increase in ROI peak signal-to-noise ratio (PSNR) up to 1.5 dB.

To conclude, introducing a bit allocation factor K helps improve budget partitioning between tiled regions. Depending on the encoded sequence and the tiling, ROI- and tile-based RC algorithm (Stream 4) gives equivalent or improved quality of ROI comparing to tile-based reference controller (Stream 2).

"Johnny" sequence at 128 kbps				"Johnny" sequence at 1.5 Mbps			
K	Bitrate (kbps)	PSNR (dB)	ΔPSNR ROI (dB)	K	Bitrate (kbps)	PSNR (dB)	ΔPSNR ROI (dB)
0.85	128.01	35.35		0.81	1500.03	42.78	
1.11	128.02	35.59	0.87	1.67	1498.22	42.72	0.07
1.33	127.98	35.66	1.37	2.02	1493.17	42.68	0.04
1.41	127.95	35.68	1.52	2.25	1487.54	42.64	0.01
"Kristen&Sara" sequence at 128 kbps				"Kristen&Sara" sequence at 1.5 Mbps			
K	Bitrate (kbps)	PSNR (dB)	ΔPSNR ROI (dB)	K	Bitrate (kbps)	PSNR (dB)	ΔPSNR ROI (dB)
0.95	130.59	32.38		0.92	1500.27	42.58	
1.16	129.56	32.58	0.60	1.66	1499.67	42.50	0.17
1.37	129.00	32.65	0.97	1.89	1498.42	42.46	0.19
1.47	128.77	32.68	1.00	1.99	1496.83	42.43	0.21
"FourPeople" sequence at 128 kbps				"FourPeople" sequence at 1.5 Mbps			
K	Bitrate (kbps)	PSNR (dB)	ΔPSNR ROI (dB)	K	Bitrate (kbps)	PSNR (dB)	ΔPSNR ROI (dB)
2.05	136.09	28.85		2.07	1499.96	41.31	
2.40	138.38	29.02	0.43	2.78	1499.68	41.22	0.15
2.83	130.95	29.11	0.93	3.24	1497.48	41.19	0.16
3.06	129.47	29.27	1.25	3.46	1493.06	41.15	0.12

Table 8.2: Global performance at low and high bit rates

8.2.2 Impact of the K factor in visual quality of ROI

In addition to the improvement in RD performance of the encoder, subjective quality increases. The proposed ROI- and tile-based controller provides an improvement in ROI quality, both in objective metrics and based on subjective quality evaluation as illustrated in Fig. 8.6. The example shows less block artifacts in the faces of stream 4 (a) than the reference stream 2 (b). The facial expression is clearer and we can see better details .

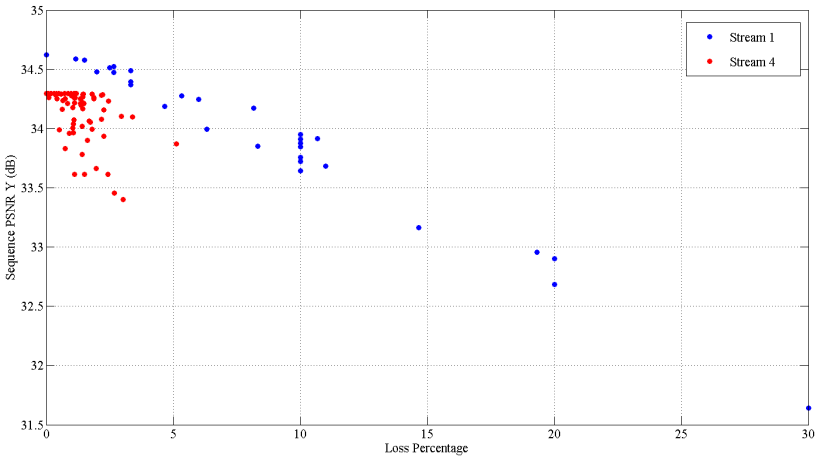
(a) Reference RC (Stream 2)

(b) ROI- and tile-based RC (Stream 4)

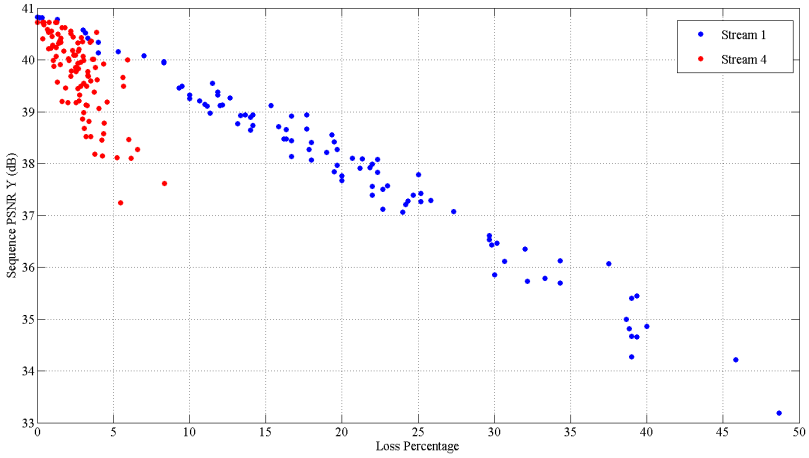
Figure 8.6: Comparison of subjective quality of "Kristen&Sara" sequence encoded at 256 kbps (Frame 593)

8.2.3 Analysis of tiling effect in visual quality

To evaluate the importance of tiles in reducing error propagation and limiting sequence quality decrease, we test 100 channel patterns and compare sequence quality for different loss percentage. Fig. 8.9 shows obtained sequence quality after decoding corrupted stream 1 (encoded with reference) and stream 4 (encoded with proposed). We notice that using the reference approach error is propagated and we can loose up to 30% of the data at 128 kbps and 50 % at 1.5 Mbps. However, using proposed approach tiling reduces error propagation. Thus, loss percentage does not exceed 5% at both low and high bit rates. Consequently, decoded sequence quality will be much more lower using the reference controller.

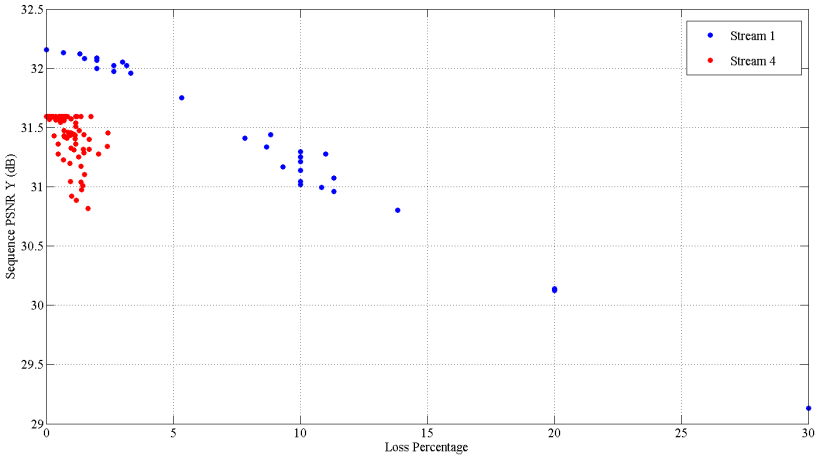


(a) 128 kbps

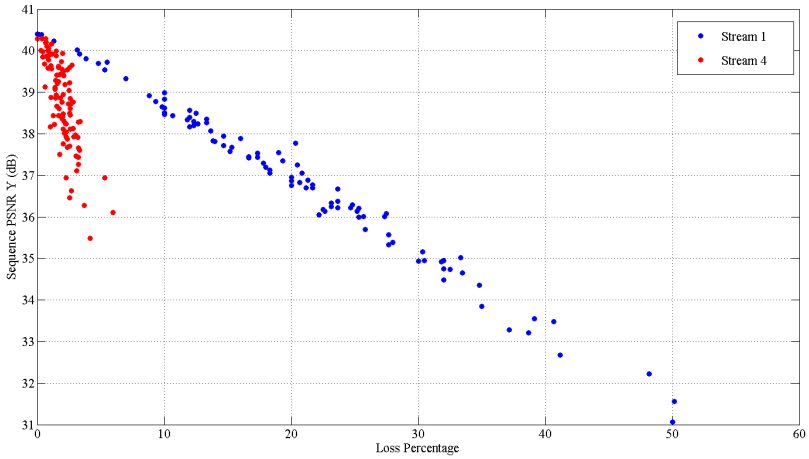


(b) 1.5 Mbps

Figure 8.7: PSNR of decoded corrupted stream for 100 tested loss patterns at low and high bit rates of “Johnny” sequence



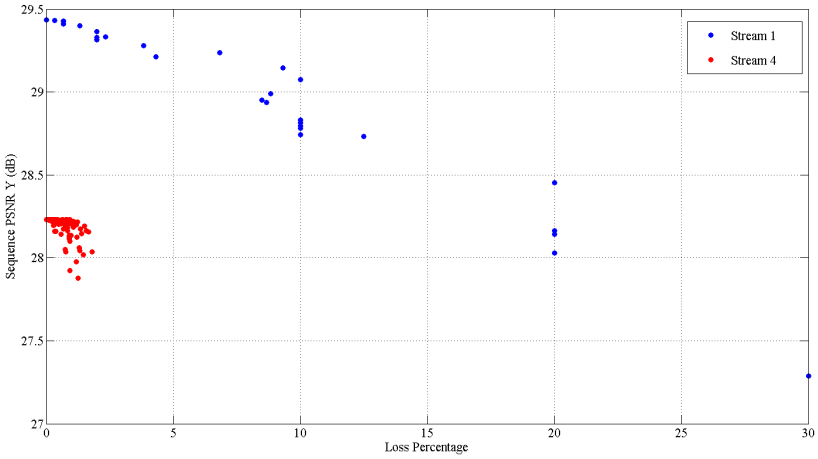
(a) 128 kbps



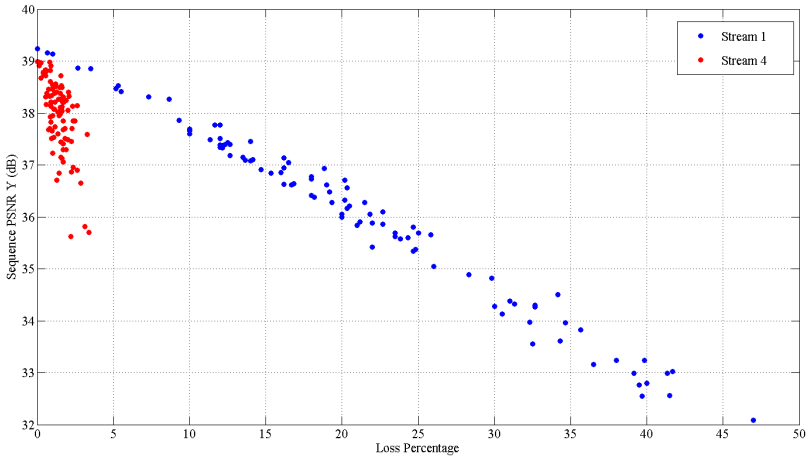
(b) 1.5 Mbps

Figure 8.8: PSNR of decoded corrupted stream for 100 tested loss patterns at low and high bit rates of "Kristen&Sara" sequence

Form these results we can also estimate the cost of tiles. In fact, if no loss is noticed (loss percentage is equal to 0%), decoded stream 4 may have lower PSNR than decoded stream 1. Fig. 8.7 ,Fig. 8.8 and Fig. 8.9 show that at low bit rate the impact of tiling is more important that at high bit rate. For example for "Johnny" we have a decrease in quality of 0.3 dB at 128 kbps, while no quality decrease at high bit rate. Moreover, the PSNR of "FourPeople" decreases of 1dB because of the chosen tiling map. A different tile partitioning could be relevant at low bit rate. In fact, when we have many small tiles, the partitioning become costly mainly at low bit rate. It affects our bit partitioning over regions. Consequently, the proposed tiling is interesting mainly at high bit rate. Thus, we will be presenting results at 1.5 Mbps in next subsections.



(a) 128 kbps



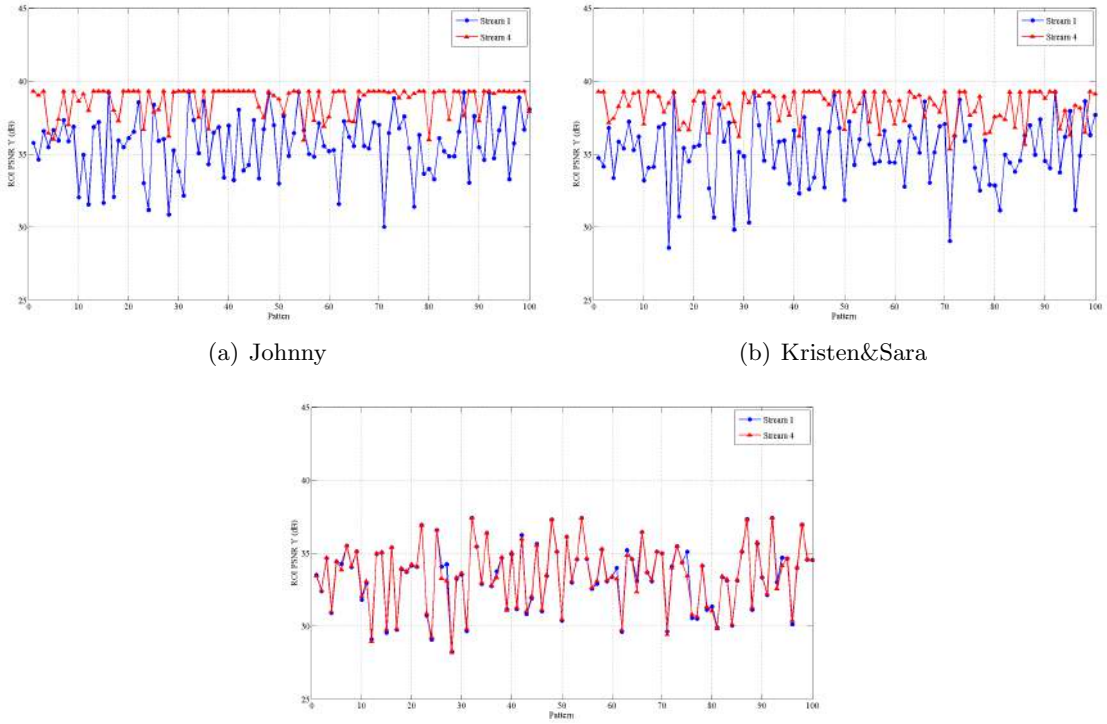
(b) 1.5 Mbps

Figure 8.9: PSNR of decoded corrupted stream for 100 tested loss patterns at low and high bit rates of "FourPeople" sequence

8.2.4 ROI quality after decoding corrupted streams

Fig.8.10 shows PSNR ROI of decoded stream 1 and stream 4 for 100 tested channel patterns. We notice that introducing a bit allocation factor K between regions over independent tiles helps improve budget partitioning between tiled regions and protect ROIs from error propagation. Depending on the encoded sequence and the tiling, ROI- and tile-based RC algorithm (Stream 4) gives equivalent (i.g. "FourPeople") or improved quality of ROI (i.g. "Johnny" and "Kristen&Sara") comparing to the reference scheme (Stream 1).

Moreover, when evaluating ROI quality frame by frame, we notice that for the reference scheme if a packet is lost from the non-ROI, all ROIs of depending frames are affected. On the contrary, our proposed scheme protects the ROI from non-ROI packet loss and



(a) Johnny (b) Kristen&Sara

(c) FourPeople

Figure 8.10: Comparison of ROI quality of decoded Stream 1 and Stream 4 at 1.5 Mbps for 100 tested patterns

corresponding tiles are not corrupted. This is visible in Fig.8.11, in particular we notice the error propagation affecting ROIs in the reference scheme.

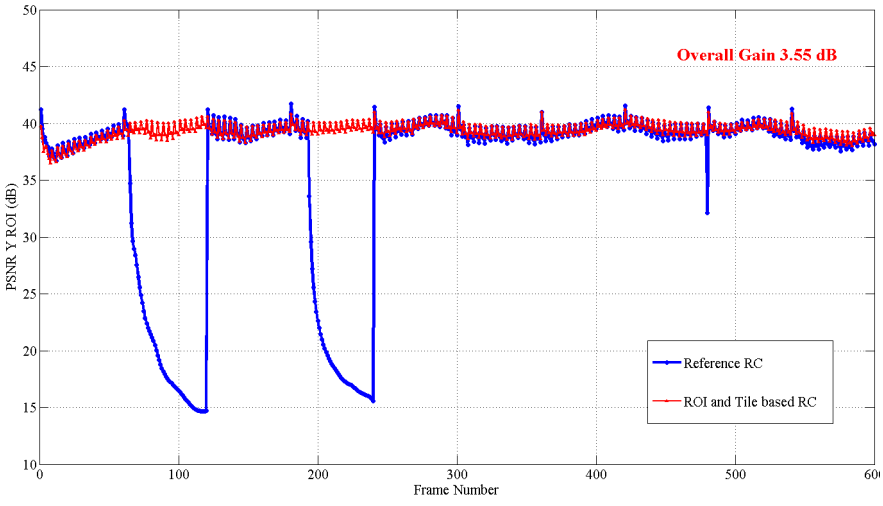


Figure 8.11: PSNR ROI of "Kristen&Sara" coded at 1.5 Mbps

Examples of the decoded class-E sequences available via the following link: `http:`

//cagnazzo.wp.mines-telecom.fr/en/?p=1092/ show the superiority of the proposed scheme with respect to the reference.

8.2.5 Impact of pattern loss in quality of decoded sequence

Finally, for different loss patterns the quality of the full sequence and the ROIs is much better using our method, even if some packets of the ROI are lost. "Kristen&Sara" example at 1.5 Mbps is given in Table 8.3. It shows that ROI and full frame quality is better using our algorithm than the reference encoder. Considering, three different patterns, we notice up to 4 dB in quality increase of the ROI. 100 patterns have been tested and results are given in Fig.8.10. Furthermore, it is possible to improve our architecture by proposing an improved tile partitioning, mainly for "FourPeople" sequence and by optimizing tile encapsulation in packets.

Number of lost packets	Reference		Proposed	
	PSNR (dB)	PSNR ROI (dB)	PSNR (dB)	PSNR ROI (dB)
4	37.34	35.75	39.21	39.30
5	36.89	35.44	39.50	39.03
7	35.91	33.81	37.86	37.13

Table 8.3: Comparison of "Kristen&Sara" decoding quality at 1.5 Mbps using different loss patterns

8.3 Conclusion

In this chapter, a new architecture for an ROI- and tile-based rate control scheme has been proposed to enhance the quality of independently decodable regions lying within an ROI, under poor channel conditions. The proposed scheme has been described in Section 8.1. The implementation has been done in HM.13 and the controller performance has been evaluated at video coding layer and network layer (Section 8.2).

The proposed architecture achieves better visual quality in ROIs thanks to independent rate allocation between regions, encoding and transmission of regions. At VCL layer the QP of LCUs of the same region are independently computed from the rest of the frame, the regions are coded in separate tiles and then transmitted in different NAL units. Consequently, at NAL layer, transmission errors do not affect both regions, they are limited to the affected tile and depending tiles at the same spatial position. However, it is important to optimize tile partitioning to reduce their cost. As a conclusion, this scheme allows a better representation of the ROI while respecting the global rate constraint.

Conclusions & future work

Thesis objectives

The purpose of this thesis was to introduce region-of-interest-based concept in High Efficiency Video Coding and develop accurate rate control methods aimed to increase the coding efficiency of regions with different importance levels. Two research phases have structured this thesis.

In the first phase, content-based rate distortion models have been proposed to perform a better distribution of quantization parameters over units of different characteristics. Theses models are thus more aimed to select optimal quantization parameter per CTU to improve rate distortion performance of the encoder.

The second phase was more dedicated to region-based coding. In fact, ROI-based rate control schemes have been developed. The implemented and tested methods during this second phase ensure independent processing of regions. Modifications are introduced in HEVC systems at two layers: video coding layer and network abstraction layer. On the first hand, rate control is performed over independently decodable regions. On the second had, different regions are transmitted in separate streams. The developed methods represent a complete ROI-based video coding approach for videoconferencing systems.

Summary

Rate Distrotion modeling for HEVC

Our first contributions in this thesis were based on rate distortion modeling. The first step was to study existing RD models initially used for QP computing at frame level by adapting them to CTU level. The proposed exponential model takes into account spatial dependencies inside a CTU. Thus, it give a good representation of the relationship between the rate and the distortion for independently decodable CTUs.

In a second work on rate distortion modeling, we derived content based RD models available for both intra- and inter-coded frames and at low and high bit rates. The idea is to fit the transform coefficient distribution using and uses its parameter to RD modeling. A study of transform distribution at CTU level helps us choose the probabilistic model to

use. Coefficients' distribution of intra-coded units was fitting using a GG while coefficients' distribution of inter-coded units was fitted using a BGG. Once the fitting is performed and the PDF parameter are derived, we used them to model the rate and the distortion and find the optimal QP distribution at CTU level that minimizes the RD cost of the frame. The obtained map of QPs is then used to encode the sequence. Experiments have shown a better RD performance comparing to the R-λ model (up to -88% in bit rate gain). Moreover, the proposed model is able to reach low bit rates that the existing R-λ model is not able to respect. As a conclusion, the

ROI-based rate control for HEVC

To perform ROI-based rate control, we first the performs of existing controller in HEVC test model. We find that the R-λ model give better RD performance than the quadratic one. Thus, the our work focused on adapting the R-λ algorithm to perform ROI-based bit allocation and QP computing. The proposed method aims at making differentiated bit allocation over different regions and compute QP parameters at CTU level after a classification of units in their corresponding region. Our ROI-based rate control for HEVC is proposed in HM.10 and improved in HM.13. The scheme takes into account Inter and Intra pictures and has been introduced at frame and CTU levels to ensure independent budget repartitioning in different regions.

To evaluate this model, we first compared it with the reference control without considering ROIs. We notice that the budget constraint is respect and an improvement in ROI quality (up to 2dB in terms of PSNR for hybrid configuration). Second, to evaluate performance of the proposed algorithm with state-of-the-art methods, we developed the ROI-based quadratic rate control algorithm. Comparative tests have shown that our algorithm give smoother bit partitioning and better RD performance.

The proposed algorithm achieves better visual quality in ROIs (Gain up to 2dB), while respecting the global bit rate constraint. This scheme is useful for videoconferencing systems to allow a better representation of the face expressions. However, as all the units (ROI and non-ROI) are in a single Tile (All units are dependent). The intra prediction may use a spacial components from the non-ROI to code a unit of the ROI. Thus in intra mode, the new bit repartition affects both non-ROI and ROI.

Consequently, we develop a new approach that introduces tiling in the ROI-based rate control scheme. This contribution show an important improvement in ROI quality as it opers in both VCL and NAL layes. Tiles help performing bit allocation over independently decodable regions so ROI and non-ROI can not affect each other and are transmitted in separate streams. This proposition help us limit error propagation across regions and have an improvement in the quality of the ROI.

Perspectives for future work

At the time of finalizing this manuscript, several interesting perspectives can be proposed to further continue the work done in this thesis. The primary points concern rate distortion modeling approaches, but some short-term improvements to the ROI-based rate control method can be proposed as well.

Rate distortion modeling improvement

- Rate distortion models based on spatial and temporal dependencies: The rate control problem is to decide how to distribute a bit budget to the units and so the frames of the GOP. The difficulty lies in the fact that CTUs are jointly coded in an hybrid video encoder based on motion estimation and compensation. In this context, exhaustive search of the coding parameters is very inefficient, because the distortion of a coded unit does not depend only of the affected rate, but also the distortion of all previously coded units in of the same frame or in previous frames used as reference. The idea is to model spatial and temporal dependencies between units by a recursive parametric RD model. This allows us to formulate allocation rate per CTU as a convex optimization problem that can effectively be solved with very recent algorithms.

- Rate distortion models for transform optimization: Transform is an important feature in HEVC standard as it reduces signal correlations in the spatial domain. As decried in the second chapter, only DCT and DST are introduced in HEVC. However, using rate distortion models it is possible to find the transform that optimized the encoder RD performance.

ROI-based rate control algorithm improvement

- Appropriate ROI detection methods and tile partitioning: Face detection is not part of our researches but it is an important step of our work and the first functionality of our scheme. Viola and Jones approach limits itself to a limited set of features and classifiers to reduce computation. Consequently, in some frames we may have false detection. Moreover, when tiling the frames, it is important to choose the appropriate repartition that does not cost a lot and can properly protect the ROI. To conclude, improvements to the detection and tiling steps would lead to a better processing of the ROI.

- ROI-based coding with error protection in tiles: The remainder of the thesis work will be essentially concentrated on improving the current ROI-based video coding algorithm by protecting the tiles of the ROI from errors by introducing a priority index in the ROI stream. Consequently, we reduce ROI packet loss and protect the transmission of the region-of-interest.

- ROI-based rate control for HEVC extensions: The scope of the JCT-VC group was extended to continue working on extensions to the HEVC standard. While the first version of HEVC is sufficient to cover a wide range of applications, needs for enhancing the standard in several ways have been identified. With the evolution of 3D technologies and devices, the standardization of extensions in 3d area is continuing. Consequently, future work may focus on rate control for multi-view content. Our algorithm can be adapted to 3D-HEVC, for more attractive applications. Furthermore, working on range extensions for embedded-bitstream scalability could be interesting. SHEVC can perfom at that moment ROI-based scalable video coding that takes into account new features of the extended version of HEVC.

Publications

Journal articles

1. 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* , 2014.

1. M. Meddeb, M. Cagnazzo and B. Pesquet-Popescu, "CTU-level operational rate and distortion modeling for HEVC", *CSVT* , 2016. In preparation.

Conference papers

1. M. Meddeb, M. Cagnazzo and B. Pesquet-Popescu, "Region-of-interest based rate control scheme for high efficiency video coding", Proceeding of *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* , Florence, Italy, May 2014.

2. M. Meddeb, M. Cagnazzo and B. Pesquet-Popescu, "ROI-based Rate control using tiles for an HEVC encoded video stream over a lossy network", Proceeding of *IEEE International Conference on Image Processing (ICIP)* , Quebec City, Canada, September 2015.

Bibliography

[1] J. Ebert and A. Willig, "A Gilbert-Elliot Bit Error Model and the Efficient Use in Packet Level Simulation," *tkn.tu-berlin.de*, 1999. *Cited in Sec.* (document), 8.1.4, 8.1

[2] Y. Q. S. H. Sun, *Image and Video Compression for Multimedia Engineering*. CRC Press, 2008. *Cited in Sec.* 1.1, 1.2.1, 1.2.2

[3] M.-T. Sun and A. R. Reibman, *Compressed Video over Networks*, 1st ed. New York, NY, USA: Marcel Dekker, Inc., 2000. *Cited in Sec.* 1.2.1

[4] R. J. Clarke, *Transform Coding of Images*. Orlando, FL, USA: Academic Press, Inc., 1985. *Cited in Sec.* 1.2.1

[5] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989. *Cited in Sec.* 1.2.1

[6] A. Gersho and R. Gray, "Scalar quantization i: Structure and performance," in *Vector Quantization and Signal Compression*, ser. The Springer International Series in Engineering and Computer Science. Springer US, 1992, vol. 159, pp. 133–172. [Online]. Available: http://dx.doi.org/10.1007/978-1-4615-3626-0_5 *Cited in Sec.* 1.2.2

[7] N. Jayant, "Adaptive quantization with a one-word memory," *Bell System Technical Journal, The*, vol. 52, no. 7, pp. 1119–1144, Sept 1973. *Cited in Sec.* 1.2.2

[8] Y. Yoo and A. Ortega, "Adaptive quantization without side information using scalar-vector quantization and trellis coded quantization," in *Signals, Systems and Computers, 1995. 1995 Conference Record of the Twenty-Ninth Asilomar Conference on* , vol. 2, Oct 1995, pp. 1398–1402 vol.2. *Cited in Sec.* 1.2.2

[9] A. Bovik, *Handbook of Image & Video Processing*. Elsevier Academic Press, 2000. *Cited in Sec.* 1.2.3

[10] K. Sayood, *Introduction to Data Compression*, fourth edition ed. Boston: Morgan Kaufmann Publishers Inc., 2012. *Cited in Sec.* 1.2.3

[11] I. E. G. Richardson, *Video Formats and Quality*. John Wiley & Sons, Ltd, 2004, pp. 9–25. [Online]. Available: http://dx.doi.org/10.1002/0470869615.ch2 *Cited in Sec.* 1.3.1

[12] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 9–10, 2008. [Online]. Available: http://digital-library.theiet.org/content/journals/10.1049/el_20080522 *Cited in Sec.* 1.3.1

[13] D. M. Rouse and S. S. , "Understanding and simplifying the structural similarity metric," in *Proceed. of IEEE Intern. Conf. Image Proc.*, 2008. *Cited in Sec.* 1.3.1

[14] T. Zhao, K. Zeng, A. Rehman, and Z. Wang, "On the use of SSIM in HEVC," in *Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, California, USA: Ieee, Nov. 2013, pp. 1107–1111. *Cited in Sec.* 1.3.1, 7.3.1

[15] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004. *Cited in Sec.* 1.3.1

[16] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *VCEG Meeting*, Austin, USA, Apr. 2001. *Cited in Sec.* 1.3.2

[17] R. Schafer and T. Sikora, "Digital video coding standards and their role in video communications," *Proceedings of the IEEE*, vol. 83, no. 6, pp. 907–924, 1995. *Cited in Sec.* 2.1.1

[18] T. Sikora, "MPEG digital video-coding standard," *IEEE Signal Processing Mag.*, Sep. 1997. *Cited in Sec.* 2.1.2

[19] C. Poynton, *Digital Video and HDTV Algorithms and Interfaces*, 1st ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003. *Cited in Sec.* 2.1.2

[20] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia*. British Library Cataloguing in Publication Data, 2003. *Cited in Sec.* 2.1.3

[21] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003. *Cited in Sec.* 2.1.4, 3.1.2

[22] N. Kamaci and Y. Altunbasak, "Performance comparison of the emerging h.264 video coding standard with the existing standards," in *Proceedings of the 2003 International Conference on Multimedia and Expo - Volume 2*, ser. ICME '03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 345–348. [Online]. Available: http://dl.acm.org/citation.cfm?id=1170745.1171559 *Cited in Sec.* 2.1.4

[23] *H.264/MPEG-4 AVC Reference Software Manual*, Joint Video Team of ISO/IEC MPEG & ITU-T VCEG, Jul. 2009. *Cited in Sec.* 2.1.4

[24] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003. *Cited in Sec.* 2.1.4

[25] G. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," vol. 5558, no. 1, Aug. 2004, pp. 454–474. *Cited in Sec.* 2.1.4

[26] 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 Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, 2012. *Cited in Sec.* 2.2, 2.2.4

[27] M. B. Vivienne Sze, Gary J. Sullivan, *High Efficiency Video Coding (HEVC) - Algorithms and architectures*. Springer, 2014. *Cited in Sec.* 2.2

[28] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012. *Cited in Sec.* 2.2.1, 2.2.2, 2.2.3

[29] W.-J. Han, J. Min, I.-K. Kim, E. Alshina, A. Alshin, T. Lee, J. Chen, V. Seregin, S. Lee, Y. M. Hong, M.-S. Cheon, N. Shlyakhov, K. McCann, T. Davies, and J.-H. Park, "Improved video compression efficiency through flexible unit representation and corresponding extension of coding tools," *IEEE Transactions on Circuits and Systems for Video Technology* , vol. 20, no. 12, pp. 1709 –1720, December 2010. *Cited in Sec.* 2.2.2

[30] K. Misra, A. Segall, M. Horowitz, S. Xu, A. Fuldseth, and M. Zhou, "An overview of tiles in hevc," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 6, pp. 969–977, Dec 2013. *Cited in Sec.* 2.2.2

[31] Y. Ye, Y. He, and Y. He, "SEI message: independently decodable regions based on tiles," in *JCT-VC L0049, 12th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2013. *Cited in Sec.* 2.2.2

[32] ——, "ROI tile sections," in *JCT-VC K0103, 11th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Shanghai, CN, 2012. *Cited in Sec.* 2.2.2

[33] R. Gregory-Clarke, L. D'Acunto, R. van Brandenburg, E. Thomas, G. Thomas, and O. Niamut, "Immersive Live event Experiences - Interactive UHDTV on Mobile Devices," *The best of IET and IBC*, vol. 6, pp. 38–43, 2014. *Cited in Sec.* 2.2.2

[34] C. Feldmann, C. Bulla, and B. Cellarius, "Efficient Stream-Reassembling for Video Conferencing Applications using Tiles in HEVC," in *5th International Conferences on Advances in Multimedia (MMEDIA)*, 2013, pp. 130–135. [Online]. Available: http://www.thinkmind.org/index.php?view=article&articleid=mmedia_2013_6_30_40108 *Cited in Sec.* 2.2.2

[35] *High Efficiency Video Coding (HEVC)*, Fraunhofer Heinrich Hertz Institute, 2013-2015. [Online]. Available: https://hevc.hhi.fraunhofer.de/ *Cited in Sec.* 2.2.3

[36] K. McCann, C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G. Sullivan, "High efficiency video coding (hevc) test model 16 (hm 16) improved encoder description," in *N14970, 19th meeting of Joint Collaborative Team on Video Coding*, Strasbourg, FR, October 2014. *Cited in Sec.* 2.2.3

[37] *H.265 : High efficiency video coding*, ITU Std., April 2015. [Online]. Available: http://www.itu.int/rec/T-REC-H.265 *Cited in Sec.* 2.2.3

[38] F. Bossen, D. Flynn, and S. Karsten, "HM Software Manual," in *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11.* Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 2014, pp. 1–18. *Cited in Sec.* 2.2.3

[39] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of h.265/mpeg-hevc, vp9, and h.264/mpeg-avc encoders," in *Picture Coding Symposium (PCS), 2013*, Dec 2013, pp. 394–397. *Cited in Sec.* 2.2.4

[40] G. Sullivan, J. Boyce, Y. Chen, J.-R. Ohm, C. Segall, and A. Vetro, "Standardized extensions of high efficiency video coding (hevc)," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 7, no. 6, pp. 1001–1016, Dec 2013. *Cited in Sec.* 2.2.4

[41] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423,623–656, Oct. 1948. *Cited in Sec.* 3.1, 3.1.1

[42] ——, "Coding theorems for a discrete source with a fidelity criterion," in *IRE National Convention Record*, vol. 4, 1959, pp. 142–163. *Cited in Sec.* 3.1.1

[43] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Mag.*, vol. 15, pp. 74–90, Nov. 1998. *Cited in Sec.* 3.1.2

[44] S. Ma, J. Si, and S. Wang, "A study on the rate distortion modeling for High Efficiency Video Coding," in *Proceed. of IEEE Intern. Conf. Image Proc.*, Sep. 2012, pp. 181–184. *Cited in Sec.* 3.2.1

[45] S. Yu and I. Ahmad, "A new rate control algorithm for MPEG-4 Video Coding," in *Proc. SPIE*, 2002. *Cited in Sec.* 3.2.1, 4.1.1

[46] Z. Li, W. Gao, F. Pan, S. Ma, K. Lim, G. Feng, X. Lin, S. Rahardja, H. Lu, and Y. Lu, "Adaptive rate control for h.264," *Elsevier J. Vis. Comm. and Image Repres.*, vol. 17, no. 2, pp. 376–406, Apr. 2006. *Cited in Sec.* 3.2.1, 4.1

[47] M. Naccari and F. Pereira, "Quadratic modeling rate control in the emerging HEVC standard," in *Proceed. of Pict. Cod. Symp.*, 2012, pp. 401–404. *Cited in Sec.* 3.2.1

[48] L. Sun, O. Au, and W. Dai, "An adaptive frame complexity based rate quantization model for intra-frame rate control of high efficiency video coding (hevc)," *APSIPA ASC*, pp. 1–6, 2012. *Cited in Sec.* 3.2.1

[49] J. Si, S. Ma, and W. Gao, "Adaptive rate control for HEVC," in *JCT-VC I0433, 9th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2012. *Cited in Sec.* 3.2.1

[50] Y. Yoon, H. Kim, S.-h. Jung, and D. Jun, "A new rate control method for hierarchical video coding in HEVC," in *IEEE International symposium on Broadband Multimedia Systems and Broadcasting*, 2012. *Cited in Sec.* 3.2.1

[51] Z. He and S. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840–849, 2002. *Cited in Sec.* 3.2.1

[52] L. Tian, Y. Zhou, and Y. Sun, "Novel rate control scheme for intra frame video coding with exponential rate-distortion model on H.264/AVC," *Elsevier J. Vis. Comm. and Image Repres.*, vol. 23, no. 6, pp. 873–882, Aug. 2012. *Cited in Sec.* 3.2.1

[53] Z. Ma, M. Xu, Y. Ou, and Y. Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 5, pp. 671–682, 2012. *Cited in Sec.* 3.2.1

[54] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by R-lambda model for HEVC," in *JCT-VC K0103, 11th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Shanghai, CN, 10-19 Oct. 2012. *Cited in Sec.* 3.2.1, 4.1, 4.2, 4.2.1, 4.2.3, 4.2.4, 7.2.1, 7.2.2, 7.3.1, 7.3.2

[55] B. Li, D. Zhang, H. Li, and J. Xu, "QP determination by lambda value," in *JCT-VC I0426, 9th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2012. *Cited in Sec.* 3.2.1

[56] F. Zhang and D. R. Bull, "An adaptive lagrange multiplier determination method for rate-distortion optimisation in hybrid video codecs," in *Proceed. of IEEE Intern. Conf. Image Proc.*, Quebec City, Canada, Oct. 2015. *Cited in Sec.* 3.2.1

[57] C.-Y. Wu and P.-C. Su, "A content-adaptive distortion-quantization model for intra coding in H.264/AVC," in *International Conference on Computer Communications and Networks ICCCN*, 2011, pp. 1–6. *Cited in Sec.* 3.2.2

[58] J. Xie, L.-t. Chia, and B.-s. Lee, "An Improved Distortion Model for rate control of DCT-based Video Coding," in *International Multi-Media Modelling Conference.* Ieee, 2006, pp. 88–95. *Cited in Sec.* 3.2.2

[59] L. Xu, X. Ji, W. Gao, and D. Zhao, "Laplacian distortion model (LDM) for rate control in video coding," in *Advances in Multimedia Information Processing*, 2007, pp. 638–646. *Cited in Sec.* 3.2.2

[60] Z. Wu, S. Xie, K. Zhang, and R. Wu, "Rate Control in Video Coding," in *Recent Advances on Video Coding*, J. Del Ser Lorente, Ed. InTech, 2011, pp. 79–117. *Cited in Sec.* 3.3

[61] J.-W. Lee and Y.-S. Ho, "Target bit matching for mpeg-2 video rate control," in *IEEE Region 10 International Conference on Global Connectivity in Energy, Computer, Communication and Control*, vol. 1, 1998, pp. 66–69 vol.1. *Cited in Sec.* 4.1.1

[62] J. Ribas-corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, 1999. *Cited in Sec.* 4.1.1

[63] H. Choi, J. Nam, J. Yoo, and D. Sim, "Rate control based on unified RQ model for HEVC," in *JCT-VC H0213, 8th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, San José, CA, USA, 2012. *Cited in Sec.* 4.2, 4.2.2, 7.1

[64] ——, "Improvement of the rate control based on pixel-based URQ model for HEVC," in *JCT-VC I0094, 9th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11* , Geneva, CH, 2012. *Cited in Sec.* 4.2, 4.2.2, 7.1, 7.1.1, 7.1.2, 7.3.1

[65] B. Li, H. Li, and L. Li, "Adaptive bit allocation for R-lambda model rate control in HM," in *JCT-VC M0036, 13th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Incheon, KR, 18-26 Apr. 2013. *Cited in Sec.* 4.2, 4.2.3, 7.3.1

[66] M. Karczewicz and X. Wang, "Intra Frame Rate Control Based SATD," in *JCT-VC M0257, 13th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Incheon, KR, 2013. *Cited in Sec.* 4.2, 4.2.1, 4.2.3, 7.2.1

[67] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, 2005. *Cited in Sec.* 4.2.1

[68] F. Bossen, "Common test conditions and software reference configurations," in *JCT-VC L1100, 12th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2013. *Cited in Sec.* 4.2.3, 4.2.4, 5.1.3, 7.3.1, 8.2

[69] *HEVC test model 13 (HM.13)*. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-13.0/ *Cited in Sec.* 4.2.3, 7.3.1

[70] A. Fiengo, G. Chierchia, M. Cagnazzo, and B. Pesquet-Popescu, "A convex-optimization framework for frame-level optimal rate allocation in predictive video coding," in *Proceed. of IEEE Intern. Conf. Acoust., Speech and Sign. Proc.*, Florence, Italy, 2014. *Cited in Sec.* 5, 5.1.1

[71] C. Pang, O. C. Au, F. Zou, and J. Dai, "An Analytic Framework for Frame-Level Dependent Bit Allocation in Hybrid Video Coding," *2011 IEEE 13th International Workshop on Multimedia Signal Processing*, 2011. *Cited in Sec.* 5.1.1

[72] C. Pang, O. Au, F. Zou, J. Dai, X. Zhang, and W. Dai, "An analytic framework for frame-level dependent bit allocation in hybrid video coding," vol. 23, no. 6, pp. 990—-1002, Jun. 2013. *Cited in Sec.* 5.1.1, 5.1.2, 5.1.3

[73] T. M. Cover and J. A. Thomas, *Elements of information theory.* New York, USA: Wiley-Interscience, 1991. *Cited in Sec.* 5.1.1

[74] T. André, M. Cagnazzo, M. Antonini, and M. Barlaud, "A JPEG2000-compatible full scalable video coder," *EURASIP Journal of Image and Video Processing*, vol. 2007, p. 11, 2007. *Cited in Sec.* 5.1.1

[75] M. Cagnazzo, T. André, M. Antonini, and M. Barlaud, "A model-based motion compensated video coder with JPEG2000 compatibility," in *Proceed. of IEEE Intern. Conf. Image Proc.* , vol. 4, Singapore, Oct. 2004, pp. 2255–2258. *Cited in Sec.* 5.1.1

[76] A. Fraysse, B. Pesquet-Popescu, and J.-C. Pesquet, "On the uniform quantization of a class of sparse sources," *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 3243–3263, 2009. *Cited in Sec.* 5.1.1, 5.3, 5.3.1, 5.5

[77] *HEVC test model 16 (HM.16)*. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.3/ *Cited in Sec.* 5.1.3

[78] J. L. Devore and N. R. Farnum, *Applied Statistic for Engineers and Scientists.* Duxbury, 1999. *Cited in Sec.* 5.1.3

[79] M. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance," *Image Processing, IEEE Transactions on*, vol. 11, no. 2, pp. 146–158, Feb 2002. *Cited in Sec.* 5.2.1

[80] B. Lee, M. Kim, and T. Nguyen, "A frame-level rate control scheme based on texture and non-texture rate models for High Efficiency Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, pp. 1–14, 2013. *Cited in Sec.* 5.3

[81] S. Boyd and L. Vandenberghe, *Convex Optimization.* New York, NY, USA: Cambridge University Press, 2004. *Cited in Sec.* 5.3.4

[82] W. Sun and Y.-X. Yua, *Optimization Theory and Methods.* Springer US, 2006. *Cited in Sec.* 5.3.4

[83] T.-H. Huang, K.-Y. Cheng, and Y.-Y. Chuang, "A collaborative benchmark for region of interest detection algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 296–303. *Cited in Sec.* 6.1

[84] D. N. Kanellopoulos, *Inteligent Multimedia Technologies for Networking applications: Techniques and Tools.* Information Science Reference, 2013. *Cited in Sec.* 6.1.1

[85] J. Hernandez, H. Morita, M. Nakano-Miytake, and H. Perez-Meana, "Movement detection and tracking using video frames," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, ser. Lecture Notes in Computer Science, E. Bayro-Corrochano and J.-O. Eklundh, Eds. Springer Berlin Heidelberg, 2009, vol. 5856, pp. 1054–1061. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-10268-4_123 *Cited in Sec.* 6.1.2

[86] M. Abdelkader, R. Chellappa, Q. Zheng, and A. Chan, "Integrated motion detection and tracking for visual surveillance," in *Computer Vision Systems, 2006 ICVS '06. IEEE International Conference on*, Jan 2006, pp. 28–28. *Cited in Sec.* 6.1.2

[87] N. B. Zahir, R. Samad, and M. Mustafa, "Initial experimental results of real-time variant pose face detection and tracking system," in *IEEE International Conference on Signal and Image Processing Applications*, Oct. 2013, pp. 264–268. *Cited in Sec.* 6.1.3

[88] I. Himawan, W. Song, and D. Tjondronegoro, "Automatic region-of-interest detection and prioritisation for visually optimised coding of low bit rate videos," in *IEEE Workshop on Applications of Computer Vision (WACV)*, Jan. 2013, pp. 76–82. *Cited in Sec.* 6.1.3

[89] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR*, vol. 1, 2001, pp. I–511–I–518. *Cited in Sec.* 6.1.3, 7.3.1

[90] Q. Li, U. Niaz, and B. Merialdo, "An improved algorithm on Viola-Jones object detector," in *Content-Based Multimedia Indexing (CBMI)*. Ieee, Jun. 2012, pp. 1–6. *Cited in Sec.* 6.1.3

[91] L. Yang, L. Zhang, S. Ma, and D. Zhao, "A ROI quality adjustable rate control scheme for low bitrate video coding," in *Proceed. of Pict. Cod. Symp.* IEEE, May 2009, pp. 1–4. *Cited in Sec.* 6.2.1, 6.2.4

[92] C.-Y. Wu and P.-C. Su, "A region-of-interest rate-control scheme for encoding traffic surveillance videos," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2009, pp. 194–197. *Cited in Sec.* 6.2.2, 6.2.4

[93] H. Hu, B. Li, W. Lin, W. Li, and M.-T. Sun, "Region-based rate control for H.264/AVC for low bit-rate applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 11, pp. 1564–1576, Nov. 2012. *Cited in Sec.* 6.2.3

[94] J. Chiang, C. Hsieh, G. Chang, F.-D. Jou, and W.-N. Lie, "Region-of-interest based rate control scheme with flexible quality on demand," in *Proceed. of IEEE Intern. Conf. on Multim. and Expo*, 2010, pp. 238–242. *Cited in Sec.* 6.2.4, 7, 7.1, 7.1.2

[95] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "Region-of-Interest Based Rate Control Scheme for High Efficiency Video Coding," in *Proceed. of IEEE Intern. Conf. Acoust., Speech and Sign. Proc.*, Florence, Italy, 2014. *Cited in Sec.* 7.2, 7.4

[96] ——, "Region-of-Interest Based Rate Control Scheme for High Efficiency Video Coding," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014. *Cited in Sec.* 7.2, 7.4

[97] *HEVC test model 9 (HM.9)*. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-9.0/ *Cited in Sec.* 7.3.1

[98] *HEVC test model 10 (HM.10)*. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-10.1/ *Cited in Sec.* 7.3.1

[99] I.-K. Kim, K. McCann, K. Sugimoto, B. Bross, and W.-j. Han, "High Efficiency Video Coding (HEVC) Test Model 10 (HM10) Encoder Description," in *JCT-VC L1002, 12th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2013. *Cited in Sec.* 7.3.1

[100] I.-K. Kim, K. McCann, K. Sugimoto, B. Bross, W.-j. Han, and G. Sullivan, "High Efficiency Video Coding (HEVC) Test Model 13 (HM13) Encoder Description," in *JCT-VC O1002, 15th meeting of Joint Collaborative Team on Video Coding of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, Geneva, CH, 2013. *Cited in Sec.* 7.3.1

[101] S. Brangoulo, N. Tizon, B. Pesquet-Popescu, and B. Lehembre, "Video Transmission over UMTS Networks Using UDP/IP," in *Proceed. of Europ. Sign. Proc. Conf.*, Florence, Italy, 2006, pp. 3–7. *Cited in Sec.* 8.1.3