Rate-distortion analysis and control in DCT-based scalable video coding

Xie, Jun

2007

Xie, J. (2007). Rate-distortion analysis and control in DCT-based scalable video coding. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/2628>

<https://doi.org/10.32657/10356/2628>

Nanyang Technological University

Downloaded on 26 Apr 2023 22:32:06 SGT

Rate-distortion Analysis and Control in DCT-based Scalable Video Coding

Xie Jun

School of Computer Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2007

Abstract

With the explosive development of computing and communication technologies, video compression technologies have advanced tremendously and have been widely used in the multimedia storage/communication in our daily life. There is a need to control the output of a video coder due to limited storage capacity (or transmission bandwidth) and the demand for high compression efficiency. Output bitrate of a video encoder and the reproduction quality at the decoder side, which are known as rate-distortion (R-D) theoretically, have been two major concerns in measuring the performance of a video compression system. To optimize the rate-distortion (R-D) performance of a video coder subject to a target R or D, video coding behaviors under different coding strategies have been extensively studied. Accurate R-D models should be based on correct understanding of video data processing in the video coder. Coding errors are introduced during the compression process and they can explicitly dictate the compressed video quality. Besides, the coding errors of the base-layer are the source of the enhancement-layer in signal-to-noise-ratio (SNR) scalable coding. Knowledge on the statistics of coding errors is important in R-D analysis of video coding. However, few studies have been conducted to illustrate what gives rise to the actual distribution of the coding errors. Consequently, variation of source statistics has not been taken into account in traditional R-D modeling for both generic discrete cosine transform (DCT) -based video coding and enhancement-layer coding in the SNR scalable coding. This dissertation first investigates the distribution of DCT residues. A distribution model for DCT residues is proposed by studying individual frequency components of DCT residues. It can quantitatively illustrate the distribution of DCT residues with respect to raw video and the quantization parameter. Subsequently, a distortion model and a rate model are derived based on the proposed probability distribution model. In practice, they are applied to control an MPEG-4 video coder. Second,

R-D behaviors of SNR scalable coding is also investigated. Since the coding errors of the base-layer depend on both the base-layer quantization strategy and raw video statistics, the enhancement-layer source statistics usually vary a lot from frame to frame. To address this problem, more source-independent R-D models and robust control algorithms are proposed for SNR scalable coders including the conventional SNR scalable coder and the recent fine-granularity-scalability coder. Compared with traditional R-D optimization schemes for SNR scalable coding, the proposed work is more robust and can achieve higher compression efficiency.

Acknowledgments

Many people have played an important part of my PhD education, as friends, teachers, and colleagues.

First and foremost, I would like to express my sincere appreciation and respect to Professor Chia Liang-Tien, my supervisor, who has led me into the research area of multimedia soon after I just finished my undergraduate courses. This work would not be made possible without his continuous support, constant encouragement, constructive comments and consistent enthusiasm throughout my PhD study and this research. Thank Professor Lee Bu-sung also for his help and advice, especially when I began my study here and was considering the research project.

In the lab, Center for Multimedia and Network Technology, I was surrounded by knowledgeable and friendly people who have always offered help. I am grateful to all the members. In particular, thanks to the former staff, Dr. Zhao Chengji, for his helpful advices during my first year research; other lab-mates who are Zou Zixuan, Yang Wenxian, Wang Susong, Tang Guosheng, Liu Song, Zhou Chen, Chu Yang, Yi Haoran, Chen Ling, Hu Yiqun, Wu Jianhua and Zhang Yu were each a great help in their own way.

In addition, thanks to the friends I have met in Singapore. I have been lucky enough to make so many good friends and I appreciate the friendship among us. Life would not have been so colorful without them. Thanks for the soccer games, the delicious cooking and so on.

Last but not least, I would like to express my deepest gratitude and love to my family and my girlfriend for their support. Without their love, understanding, support and encouragement, the completion of this study would have never been possible. I thank them for all their love and help they have given me.

Contents

Abstract	i
Acknowledgments	iii
List of Figures	viii
List of Tables	xii
Acronym Index	xiii
1 Introduction	1
1.1 Motivation	2
1.2 Objective and Approach	7
1.3 Major Contributions of the Dissertation	9
1.4 Dissertation Organization	11
2 Background and Related Work	13
2.1 Video Coding Basics	14
2.1.1 Non-scalable Video Coder	14
2.1.2 SNR Scalable Coder	20
2.2 Distributions of Source DCT Coefficients	26
2.3 Classical Rate Distortion Theory	27
2.4 Related Work on R-D Modeling and Control for Video Coding	31
2.4.1 Approaches of R-D Modeling for Video Coding	32
2.4.2 Typical Rate Control Algorithms	33
2.4.3 Quality Control	37
2.5 Summary	38

3	Mathematical Analysis on the Distribution of DCT Residues	39
3.1	Problem Description of Previous Study on the Distribution of DCT Residues	40
3.2	Proposed Distribution Model for DCT Residues	44
3.2.1	Frequency Processing of a Video Encoder	44
3.2.2	The Distribution of individual frequency components	46
3.2.3	The PDF for All the DCT Residues	51
3.3	Experiments: Goodness-of-fit Test	53
3.4	Summary	58
4	R-D Modeling and Control for Non-scalable DCT Video Coding	60
4.1	R-D Modeling for Non-scalable Video Coding	61
4.1.1	Distortion Model	61
4.1.2	Rate Model	63
4.1.3	R-D Model with respect to QP	64
4.2	Quality Control	65
4.2.1	Performance of the Proposed Distortion Model Versus QP	65
4.2.2	Video Coding at a Desired Fidelity Level	66
4.3	Rate Control	76
4.3.1	Rate Control Algorithm	76
4.3.2	Experimental Results of Rate Control	79
4.4	Summary	81
5	Rate Control for Conventional SNR Scalable Coding	86
5.1	Problem Description	87
5.1.1	The Idea of SNR Scalable Coding	87
5.1.2	Problem Description of Rate Control for SNR Scalable Coding	87
5.2	Some Useful Characteristics of EL Compression	90
5.2.1	Relationship Between R_E and r_Q	91
5.2.2	Relationship Between R_E and Non-zero Percentage Among Quantized Coefficients	95
5.2.3	Relationship between r_Q and P_E^{nz}	97
5.3	Optimum Bit Allocation for Enhancement-layer Coding	97

5.3.1	Theoretical Model for Optimum Bit Allocation	101
5.3.2	Challenges for Practical OBA	103
5.3.3	Practical OBA Schemes in MB-level	105
5.4	Rate Control	108
5.4.1	Rate Model	108
5.4.2	Frame-level Rate Control	109
5.4.3	MB-level Rate Control	109
5.5	Experimental Results	111
5.6	Summary	113
6	R-D Analysis and Optimal Bit Allocation for FGS Coding	117
6.1	Problems Description of Bit Allocation for FGS Coding	118
6.1.1	Optimal Bit Allocation in R-D Theory	118
6.1.2	Drawbacks of Existing Bit Allocation Schemes for FGS Coding . .	119
6.2	Proposed R-D Analysis for FGS Coding	124
6.2.1	Linear Rate Model for FGS Coding	124
6.2.2	Distortion Analysis for FGS Coding	127
6.3	Bit Allocation Scheme	130
6.3.1	Estimation of θ in the Rate Model	130
6.3.2	Bit Allocation Algorithm	131
6.4	Experimental Results	133
6.5	Summary	134
7	Conclusions and Future Work	137
7.1	Conclusion	137
7.2	Future Work	139
	Appendix	143
A.	Review of Kolmogorov-Smirnov Test for the Goodness-of-fit	143
B.	Derivation: PDF of Quantization Error When a Laplacian Source is Quan- tized by a Uniform Quantizer	144
C.	Proof: the optimal strategy of truncating NZBC	147
D.	Sample of test video sequences	150

References	151
List of Publication	163

List of Figures

1.1	3-stage processing of video coding	2
1.2	R-D curve	3
1.3	R-Q and D-Q functions	5
2.1	A hybrid motion-compensation and DCT video codec	15
2.2	Examples of DCT coefficient variance at different frequencies	18
2.3	Two-layer SNR scalable coder structure	22
2.4	The scan order of enhancement-layer DCT coefficients in FGS coding	25
2.5	FGS encoder	25
2.6	Percentage versus value of source DC coefficients in intra- frames.	27
2.7	Percentage versus value of source AC coefficients in Inter- frames	28
2.8	The Gaussian PDF and the Laplacian PDF	29
2.9	R-D system from source to user	29
3.1	The actual distribution of DCT residues.	42
3.2	Video coding system for individual frequency components	44
3.3	Examples: distribution of DC residues coefficients, percentage versus value. 47	
3.4	Examples: distribution of DC residues coefficients, percentage versus value. 48	
3.5	Experimental distribution of inter- DC coefficients against the Laplacian and Gaussian distributions,	49
3.6	Examples: distribution of AC residues coefficients, percentage versus value. 50	

3.7	PDF of Quantization error in the case of uniformly quantizing a Laplacian source, $f_{ac}(e, q, \lambda_x)$, with $t_0 = 0$ and $\lambda_x = 0.2$	51
3.8	Comparison of average KS test statistic t_{ks} at different QP	55
3.9	Average optimal ν in GGD versus QP in Foreman	58
3.10	PMF of DCT residues. The experimental DCT residues are from one P Foreman frame quantized with QP = 16, and the PMFs modeled by Gaussian, Laplacian, GGD and the proposed distributions are plotted.	59
4.1	Normalized distortion model of the AC component. Vertical axis: $D \cdot \frac{\lambda_x^2}{2} = \frac{D}{\sigma_x^2}$, horizontal axis: $\lambda_x \cdot q$	63
4.2	Performance comparison of distortion models with respect to QP	67
4.3	Flowchart of the quality control algorithm	68
4.4	Iterative algorithm to get the QP solution at a target distortion level	69
4.5	PSNR comparison between the uniform quality control and the proposed quality control when raw sequences are encoded at the target PSNR of 35db.	72
4.6	Bitrate of individual frames when raw sequences are encoded at the target PSNR of 35db.	73
4.7	Value of adaptive parameter (K_D^t) in the quality control.	75
4.8	Flowchart of the rate control algorithm	78
4.9	Iterative algorithm to get the QP solution at a target bitrate	79
4.10	PSNR comparison between VM18 rate control, ρ -domain rate control and the proposed rate control.	82
4.11	Rate (bit per pixel) versus frame number by the VM18 rate control, ρ -domain rate control and the proposed rate control respectively.	83
4.12	Buffer fullness comparison between VM18 rate control, ρ -domain rate control and the proposed rate control.	84
5.1	Quantizer in the SNR scalable coder	88
5.2	Probability distribution of AC (0,1) coefficients. (a)(c): the BL and EL in Mobile & Calendar respectively; (b)(d): the BL and EL in Stefan respectively	89

5.3	Plot of the function R_E of $\frac{Q_B}{Q_E}$	93
5.4	Relationship between bitrate (R_E^0) in terms of bit per pixel (bpp) and the ratio of quantization parameters (r_Q) in different test sequences, where the horizontal axis is r_Q and the vertical axis is R_E^0	94
5.5	R_E^0 (bpp) versus P_E^{nz} in different test sequences	96
5.6	K_E^R versus P_E^{nz} in different sequences	98
5.7	K_E^R versus R_E^0 (bpp) in different sequences	99
5.8	P_E^{nz} versus r_Q in different sequences	100
5.9	Normalized distortion curves of different frames. Normalized distortion D_{E0} versus the percentage of non-zero coefficients P_E^{nz}	102
5.10	Example of the MB variance distribution and the corresponding allocated bit number with the MB-level OBA	104
5.11	Plot of α_E versus coding bitrate using Mobile & Calendar	106
5.12	Flowchart of rate control at MB-level	110
5.13	PSNR performance comparison between TM5 and the proposed algorithm for EL coding, when the BL is encoded in the VBR mode and the EL is encoded at 3Mbit/s.	114
5.14	PSNR performance comparison between TM5 and the proposed algorithm for EL coding when the BL and the EL are encoded at 3Mbit/s respectively.	115
5.15	Comparison of the distortion MSE of the reconstructed frame in MB-level, where distortion is measured by MSE	116
6.1	ϵ_i^2 of different frames in the FGS-layer	121
6.2	Allocated number of bits for different frames (R_i) by existing bit allocation schemes	123
6.3	Percentage of non-zero binary coefficients ($P1$) in different BP	126
6.4	Rate (bpp) of different BP	126
6.5	The rate R (bit) versus the percentage of NZBC ($P1$) in different BP from one frame in Foreman	128
6.6	θ versus $P1$ in different BP from 300 frames of Foreman	131

6.7	PSNR Comparison between the proposed work (red solid line) and uniform bit allocation (black dotted line)	135
7.1	Uniform quantizer	145

List of Tables

3.1	Number of essential mathematical functions in the PDF of the generalized Gaussian distribution (GGD) and the proposed PDF.	57
3.2	Example of total numbers of mathematical functions executed in the KS test with the GGD model and the proposed distribution model.	57
4.1	Control error comparison, in terms of PSNR, between the proposed quality control (PQC) and the uniform quality control (UQC), when video sequences are encoded at a target PSNR level.	71
4.2	Experimental result comparison between VM18 rate control (VM18), ρ -domain rate control (ρ -D) and the proposed rate control (new).	80
5.1	Performance comparison when the BL is encoded in VBR model	112
5.2	Performance comparison when the BL is encoded in CBR model	112
6.1	Definitions of notations for the FGS analysis	124
6.2	PSNR comparison between the uniform bit allocation (uniform) and the proposed bit allocation (proposed)	134

Symbols and Abbreviations:

< notations >: < description >, < page >

B_c : current buffer level, 77	f_{ac} : probability density function of the AC component, 50
B_s : buffer size, 77	f_{eG} : quantization error PDF when a Gaussian source is uniformly quantized, 46
C_E : header and syntax bits in the EL, 108	q : quantization stepsize, 35, 45
D : distortion, 3	t_0 : dead-zone of uniform quantizer, 45
D_T : target distortion, 66	t_{ks} : test statistics, 54
K_D^t : ratio between actual distortion and the result of the distortion model, 66	B-frame: bidirectional interpolated frame, 14
P_E^{nz} : percentage of non-zeros among quantized coefficients, 95	BL: base layer, 87
Q : quantization parameter, 19	BP: bitplane, 23
Q_B : quantization parameters in the BL, 87	BP: BP, 117
Q_E : quantization parameters in the EL, 87	bpp: bit per pixel, 94
R : encoding bitrate, 3	CDF: cumulative distribution function, 53, 144
R_E^0 : bitrate for texture information in the EL, 108	DCT: discrete cosine transform, 2
R_E^0 : bitrate of all those coefficients in the EL excluding the header and syntax, 92	DVD: digital video disk, 37
T_{GoF} : target number of bits for one group of frames, 131	EL: enhancement-layer, 87
ρ : the percentage of zeros among quantized transform coefficients, 33	FGS: fine-granularity-scalable, 8, 117
σ_E^2 : the variance of the EL difference frame pixels, 91	GOP: group of pictures, 33
f_{dc}^0 : PDF of intra-DC residues, 47	I-frame: intra- frame, 14

KS: Kolmogorov-Smirnov, 48

LSB plane: least significant bitplane, 125

MAD: mean-absolute-difference, 35

MB: macroblock, 97

MSB plane: most significant bit plane, 24

MSB plane: most significant bitplane, 125

MSE: mean-squared-error, 61, 70, 118

NZBC: non-zero binary coefficients, 125, 139

NZBC: non-zero binary coefficients in the
bitplane coding, 11

OBA: optimal-bit-allocation , 97

P-frame: prediction coded frame, 14

PDF: probability density function, 9, 26, 61

QP: quantization parameter, 4, 64

QPs: quantization parameter, 60

SNR: signal-to-noise ratio, 6

UTQ: uniform threshold quantizer, 45

Chapter 1

Introduction

Vision is the most advanced of all human senses, so it is not surprising that it plays the single most important role in human perception. Without doubt, video is thus one of the most vital mediums of information communication in modern life. However, raw visual data captured by photographic devices are tremendously large. They are not suitable for storage or delivery over networks due to limited resources. Modern video compression techniques offer the possibility to store or transmit the vast amount of data necessary to represent digital video in an efficient and robust way [1]. With the explosive development of computing and communication technologies in recent years, video compression technologies have received a lot of attention from both academia and industry.

Over the last decade, video compression techniques have advanced a lot targeting at achieving higher compression gains, error robustness and network/device adaptation. They have been widely-used in multimedia storage [2] and communication [3, 4]. For the sake of hardware implementations and commercial interoperability, digital video compression standards have been proposed since the early 1990's. In the meantime, technological progress of video coding has promoted the finalization of a series of video compression standards, which are H.261 [5], MPEG-1 [6], MPEG-2 [7, 8], H.263 [9–11], MPEG-4 [12, 13] and H.264/MPEG-4 part 10 [14] covering various applications such as video

coding, digital storage media, television broadcasting, Internet streaming and so on.

1.1 Motivation

In various video compression standards [5–14], a generalized design of video coders employs a 3-stage structure (Fig 1.1). The 3 stages are transform, lossy-coding and entropy-coding. In the first stage, raw video data in the spatial domain is transformed to coefficients in the frequency domain in order to remove redundancy between data samples and to concentrate the energy of video data into a small number of transform coefficients [15]. The most widely-used method in video compression is discrete cosine transform (DCT). In the second stage, quantization [16, 17] is employed to considerably reduce the magnitude of coefficients for a high compression ratio. In the last stage, entropy coding techniques are applied to further increase the compression ratio. Some of coded video frames are reconstructed with inverse-quantization and inverse-transform in the encoder. They are then used as a reference to encode coming frames for the removal of temporal redundancies, which is a major difference between a video coder and an image coder.

Figure 1.1: 3-stage processing of video coding

In terms of the coder structure, the one-layer structure, as shown in Fig 1.1, is called a non-scalable coder. It is the main part of video compression standards. With recent development in networking technologies, scalable coding (also named layered coding)

CHAPTER 1. INTRODUCTION

[18–23] has been incorporated into standards to provide more flexible video compression. A scalable coder has a structure of at least two layers. One is the base-layer and the others are enhancement-layers. A base-layer can be independently decoded and it is usually encoded in a low bitrate to provide basic visual quality. Enhancement-layers can only be decoded together with the base-layer. They are responsible for encoding errors of base-layer pictures so as to further enhance picture quality of the base-layer.

Various video compression standards define the coder structure, syntax and semantics of video bitstreams only. Beyond their scopes, there is an open issue left that is how to control the video encoder. Encoding bitrate, denoted by R , and quality of decoded video, indicated by distortion (D), have been the key pair of parameters in measuring the performance of video coders. Ideally, both R and D are expected to be as small as possible. However, for a given coder and a given identical source, distortion (D) varies inversely with encoding bitrate (R) [24, 25], as qualitatively depicted in Fig 1.2. Therefore, a tradeoff has to be made between R and D , when video coding is conducted.

Figure 1.2: R-D curve

CHAPTER 1. INTRODUCTION

In order to obtain an optimal R-D tradeoff, R-D characteristics of video coding need to be studied. In the process of video coding (Fig 1.1), quantization is the only stage which is lossy. It plays a crucial role on the final R-D output, because quantized transform coefficients, instead of original transform coefficients, are coded in compressed video bitstreams. Video coder [5–13] employ a set of quantization parameters (QP) to adjust the quantization stepsize for transform coefficients. The QP is hence responsible for the adjustment of the coder output. Accordingly, finding an appropriate QP solution is the key to achieve the optimal tradeoff between R and D. One of keen topics for video coding is to study the relationship between R-D and QP [26–29] which is usually characterized by the rate-quantization function ($R(Q)$) and the distortion-quantization function ($D(Q)$). Their characteristics curves are shown in Fig 1.3. With the joint consideration of $R(Q)$ and $D(Q)$ functions, the QP solution can be achieved subject to certain target bitrate (R_T) or target distortion (D_T). In such an approach, plenty of work [27–37] has been done to model R-D behaviors of video coding. Some classical algorithms have been adopted as standardized rate control algorithms, such as TM5 [38], TMN10 [39] and VM18 [40] rate control algorithms. They represented the state-of-the-art of rate control for its associated coder at that time. In this dissertation, these standardized algorithms are the reference to compare with the proposed rate control algorithm.

Accurate control of video coding cannot be performed with inaccurate R-D models. It relies on good understanding of the statistical distribution of source video data, video data processing and statistics of ultimate coding errors. The statistics of source DCT coefficients have been studied in [41–43] for natural images and in [44] for motion-compensated pictures. And there is a lot of work on the analysis for the video coding system [28, 34], including R-D models such as [27–37]. Surprisingly, we found that few researchers had modeled the distribution of coding errors and analyzed its relationship to video source and quantization resolution. Knowledge on the distribution of coding errors

Figure 1.3: R-Q and D-Q functions

is instrumental in R-D modeling, as coding errors cause the distortion of encoded video and also have a strong correlation to the encoding bitrate. The lack of this knowledge will induce the inaccuracy of the distortion model and it will hereby lower the overall performance of the R-D analysis framework. The standardized rate control algorithms [38–40] employed an oversimplified distortion model defined by QP only. Underlying ideas of their R-D models assume that the magnitude of coding errors has a uniform distribution over a range defined by the adopted QP. In spite of simplicity, their solutions are not accurate enough, especially in low bitrate coding [45–47]. Some recent studies intended to model the distribution of coding errors by more sophisticated models such as the Gaussian distribution in [48], the generalized Gaussian distribution in [45] and the addition of two Laplacian distributions in [49]. They employed a statistical method to fit parameters of their conjectured models to collected coding errors, so their models cannot illustrate what gives rise to the actual distribution of coding errors. Additionally, these models presumed that coding errors were distributed over a boundless range which is inconsistent with the actual distribution limited by the quantization stepsize. Therefore,

CHAPTER 1. INTRODUCTION

it is needed to investigate the distribution of coding errors, from which it is promising to develop a more accurate R-D model and then to perform better control of video coding.

Another challenge relates to control of signal-to-noise ratio (SNR) scalable coding. The base-layer coding of an SNR scalable coder is quite the same as non-scalable coding, and the source of the enhancement-layer is the coding errors of the base-layer. As mentioned earlier regarding coding errors in non-scalable coding, very few studies have been done on the distribution of the enhancement-layer source. Due to limited knowledge on distribution of the enhancement-layer source, most of existing R-D analysis frameworks for enhancement-layer coding [50–52] employed oversimplified solutions which does not take the variation of the statistics of enhancement-layer sources into account. The simplification deviates the actual observation and it will also cause a large control error [53]. Moreover, those standardized rate control algorithms [38–40], which were originally proposed for non-scalable coding, are not applicable to enhancement-layer compression. It is because the source and the quantization stepsize of the enhancement-layer in SNR scalable coding are different from those in non-scalable (and base-layer) coding. Specifically, the statistical distribution of the enhancement-layer source are highly dependent on both raw video and the base-layer coding strategy, and the quantization stepsize in the enhancement-layer must be smaller than that in the base-layer. Therefore, more accurate R-D analysis for SNR scalable coding is needed to take into consideration the characteristics of SNR scalable coding.

The dominant design of video coders is DCT-based. Motivated by the above challenges, we will analyze the distribution of coding errors in DCT domain and then investigate R-D behaviors for the DCT-based coders, including the non-scalable coder and the SNR scalable coder. Therefore, this dissertation is titled **Rate-distortion Analysis and Control in DCT-based Scalable Video Coding**.

1.2 Objective and Approach

This research work looks into the distribution of coding errors and R-D analysis and control for DCT-based coders. The following targets have been set towards fulfilling the theme of this research.

- **Mathematical analysis on the distribution of coding errors**

Following denotations are used in the manuscript for clarity. *source DCT coefficients* stand for DCT coefficients which are to be quantized in the non-scalable coding (or base-layer coding). *DCT residues* are defined as coding errors in the DCT domain in non-scalable coding (or base-layer coding). The DCT residues are not only the distortion in non-scalable coding (or base-layer coding) but also the enhancement-layer source in SNR scalable coding.

The first objective of this dissertation is to quantitatively analyze what gives rise to the actual distribution of the DCT residues. According to the video compression process, it can be found that the distortion is caused solely by the quantization, so the DCT residues are quantization errors in essence. In common uniform quantization, the distribution of quantization error depends on two factors which are the source statistics and the quantization stepsize. However, the encoder processes all the video data in a blocked-based mode, for example, the typical size of the block is 8×8 . The quantization scheme in video coding is complicated, as the quantization stepsize for block-wise source DCT coefficients is the product of the QP and a quantization weight matrix to emphasize some low frequency information [40, 54]. It causes difficulty in deriving the distribution model of DCT residues. To the best of the author's knowledge, the distribution of DCT residues have not been studied quantitatively.

The quantizer in each frequency is uniform and the statistics of individual frequency components have been formulated in [41–43, 55]. This makes it feasible to model the distribution of DCT residues by the means of studying individual frequency components.

- **Accurate R-D analysis and control for non-scalable coding**

The R-D analysis framework consists of R-Q modeling and D-Q modeling. Its effectiveness depends on the accuracies of both the R-Q model and the D-Q model. The second objective of this dissertation is to propose a more accurate R-D model than conventional R-D models. It is supposed to be scalable to both the variation of video contents and the quantization resolution. To continue our analysis on the distribution of DCT residues, we look forward to developing a more accurate distortion model. Then, a better rate model can be achieved by combining the proposed distortion model with the classical rate-distortion function. When the proposed R-D model is applied to practical video coding, it is expected to provide better performance than the standard rate control algorithm.

- **R-D analysis and control for SNR scalable coding**

In SNR scalable coding, the enhancement-layer source is DCT residues of the base-layer. Its enhancement nature makes the coder design of the enhancement-layer different from that of the non-scalable coder, which results in dissimilar R-D characteristics of the enhancement-layer coding. The third objective is to analyze, model and control the R-D behaviors of the enhancement-layer in the SNR scalable coding. In terms of the technique of dealing with DCT transform coefficients, we categorize standardized SNR scalable coders into the conventional QP-based coder [7, 18, 20] and the bitplane-coding based fine-granularity-scalable (FGS) coder [22, 23, 56]. The QP-based enhancement-layer coder adopts the same quantization scheme in the common non-scalable coder which quantizes each DCT coefficient as a decimal integer. The FGS coder employs the bitplane coding technique considering each quantized DCT coefficient as a binary representation, instead of a decimal integer of a certain value [22, 57]. Hence, FGS bitstream can be truncated at an arbitrary bitrate, so FGS coding is more flexible than traditional SNR scalable coding. New R-D analysis for the enhancement-layer coding

should take those characteristics into account. The proposed R-D optimization method is expected to achieve more robust and efficient compression than traditional ones.

1.3 Major Contributions of the Dissertation

The work reported in this dissertation aims to provide better solutions for DCT video compression. The attempt has led to contributions and original results as follows.

- **Proposal of a quantitative model for analyzing the distribution of DCT residues**

A solution to model the distribution of DCT residues is provided by studying their individual frequency components. Based on uniform quantization theory and widely-adopted statistical models for raw video data, the probability density function (PDF) of DCT residues is proposed. Experimental results suggest it is closer to the actual distribution than the PDF of commonly-used distributions such as uniform, Gaussian and Laplacian distributions. Furthermore, the proposed PDF can quantitatively explain the relationship between the distribution of DCT residues and both video source and the quantization stepsize. This gives the proposed approach an advantage over other statistical approaches [48, 49, 58] which conjectured the PDF according to known residue data. As the proposed PDF is a function of source video variance and the quantization stepsize, it can predict the possible distribution of DCT residues prior to actual video coding. This advantage will be useful in the study of realtime video compression.

- **Development of R-D modeling for non-scalable coding**

Based on the proposed PDF of DCT residues, an accurate distortion model is developed with good adaptation to the variation of both quantization stepsize and video source contents. The proposed distortion model quantifies the distortion both with respect to the video source and with respect to the quantization parameter. It can estimate

the distortion more accurately than the existing distortion models in the standardized rate control. Combining the proposed distortion model with the classic R-D function in information theory [16, 24, 59], the R-Q model is obtained.

• **Development of quality control and rate control for non-scalable coding**

With the proposed R-D model for non-scalable coding, a quality control algorithm and a rate control algorithm are developed to regulate the R-D behaviors. In the quality control, the quality of encoded video can be controlled in one-pass mode at a desired fidelity level. The proposed quality control outperforms existing quality control based on the uniform distortion model and is different from the traditional quality control algorithms [60, 61] that are in essence an empirical approach. Another application of the proposed R-D model is to develop a rate control algorithm. In comparison with the VM18 rate control [40] for MPEG-4, the proposed rate control algorithm is more accurate and can achieve higher compression efficiency.

• **Development of a rate control algorithm for QP-based SNR scalable coding**

To cope with the possible large variation of the statistical distribution of the enhancement-layer source, a source-independent rate control algorithm is developed. According to R-D theory and quantization theory, an approximate linear relationship between the bitrate and the ratio of quantization stepsize of different layers is formulated. Furthermore, the linear rate model, namely ρ -domain rate model proposed for non-scalable coding, is investigated for the enhancement-layer and it is observed that the linear rate model has large varying slope. We provide a solution to overcome this limitation. For robust video compression, a practical optimal bit allocation strategy in the macroblock-level is proposed.

- **Development of R-D analysis and bit allocation for FGS coding**

To deal with the large variation of the FGS-layer source statistics and take the bitplane-coding characteristics into account, we study the relationship between R-D and the number of truncated non-zero binary coefficients (NZBC) in the FGS coding. We propose a linear rate model based on the linear relationship between the rate of the truncated bitplane and the number of truncated NZBC. The advantage of the proposed rate model is that it is less source-dependent and more accurate. The relationship between the distortion and NZBC is also analyzed. From the theoretical derivation, the optimal strategy of truncating the NZBC is given to minimize the overall distortion at a targeted number of NZBC. An optimal bit allocation algorithm is then developed to minimize the overall distortion at a target bitrate. The proposed bit allocation algorithm can provide higher video quality and is more robust compared to traditional bit allocation algorithms.

1.4 Dissertation Organization

This dissertation is organized into seven chapters.

Chapter 2 is literature review. Video coding basics are introduced first, and subsequently, related work on R-D analysis and control for video coding is reviewed.

Chapter 3 presents the proposed mathematical analysis on the distribution of DCT residues. By the method of frequency decomposition, individual frequency components of DCT residues are modeled first. Combining them together, the PDF for all DCT residues is proposed.

Chapter 4 presents the proposed R-D analysis framework for non-scalable coding on the basis of the proposed PDF of DCT residues. Subsequently, the proposed R-D model is applied to control an MPEG-4 video coder.

Chapter 5 investigates the rate control for the convenient QP-based SNR scalable coder. Subsequent to studying the quantization resolution of different layers and dis-

CHAPTER 1. INTRODUCTION

cussing the validity of the rate model for non-scalable coding in the enhancement-layer, a rate control algorithm for the enhancement-layer is proposed.

Chapter 6 presents a bit allocation scheme for FGS coding. The relationship between R-D and the non-zero binary coefficient is investigated. Then, an optimum bit allocation scheme is proposed to truncate the FGS bitstream.

Chapter 7 concludes the dissertation and discusses some possible future work.

Chapter 2

Background and Related Work

Video compression techniques represent raw video data by removing spatial and spectral redundancies in response to the human visual system and exploiting the temporal correlation among consecutive frames. A video compression system, commonly known as codec, comprises an encoder and a decoder. Its design has been defined in a series of video coding standards [5–13], and it is dominantly a three-stage process consisting of DCT, quantization and entropy-coding. Beyond the scope of the video coder design, controlling the video coder has been a hot topic all the time. The practical control of video coding is complex, because it relates to the variation of video contents, the adopted coding strategy and their effects on the final R-D output of a video coder. In order to achieve efficient video compression, video coding behaviors have been extensively studied and many R-D models have been proposed to analyze the R-D performance of the video coder. Based on the models, practical R-D control algorithms have been developed to control the video coders. This chapter first introduces basic video coding technologies including the generic DCT-based non-scalable coder and SNR scalable coders. Subsequently, well-known statistical models for video sources in the DCT domain are briefly introduced. Then, R-D modeling for video coding, as well as the classical R-D theory, is reviewed.

2.1 Video Coding Basics

In this section, basic video coding techniques are introduced with the illustration of a typical hybrid motion-compensation and DCT coder. Subsequently, the SNR scalable coder is presented.

2.1.1 Non-scalable Video Coder

Non-scalable video coding is the primary concern of various video compression standards and also the basis of scalable video coding. Most of the non-scalable video coders adopt the motion-compensation and DCT approach to compress raw video sequences. The concept of such an approach comes mainly from three parts. First, motion compensation is performed to remove temporal redundancies among image frames. Second, DCT is employed as a means to remove spatial redundancies within an image frame [62]. Third, entropy coding techniques are used to express the information in a minimal set. Such a video coder structure is shown in Fig 2.1, and every stage is described in detail as follows.

2.1.1.1 Motion Compensation

Frames of a raw video sequence can be encoded either in the intra mode or in the inter mode. Intra frames (I-frame) are encoded without reference to any past or coming frames. Inter frames undergo motion compensation with reference to previously encoded frames, before they are transformed to frequency coefficients. There are two types of inter frames, namely prediction coded frames (P-frame) and bidirectional interpolated frames (B-frame). The purpose of motion compensation is to remove the temporal redundancies by the exploitation of the similarities among successive frames [63, 64]. Accordingly, motion vectors are required to indicate the distance between the current area and the reference area. They are coded in the bitstreams necessarily for the proper decoding.

CHAPTER 2. BACKGROUND AND RELATED WORK

2.1.a: Encoder

2.1.b: Decoder

Figure 2.1: A hybrid motion-compensation and DCT video codec

CHAPTER 2. BACKGROUND AND RELATED WORK

2.1.1.2 Discrete Cosine Transform

Frame pixels in the intra frame or in the motion-compensated frame are divided into non-overlapped blocks. These block-wise pixels are transformed from the spatial domain to the frequency domain where video data is better prepared for next-stage operation, namely quantization.

Currently, DCT [65] is the most widely-used transform in video compression. The basis vectors of one-dimensional (1-D) N -point DCT, $\mathbf{H}[u, n]$, are defined by (2.1), and the matrix containing all basis vectors is denoted by $\mathbf{H}$. Using $\mathbf{F} = [f_0, f_1, \dots, f_{N-1}]^T$ to represent a vector of original samples, 1-D DCT result of $\mathbf{F}$ is $\mathbf{X} = \mathbf{H}\mathbf{F}$ in (2.2) where $\mathcal{N}$ denotes the set $\{0, 1, \dots, N-1\}$. In video compression, DCT is applied to blocks consisting of $N \times N$ pixels, which is a two-dimensional (2-D) case. The 2-D DCT is similar to 1-D DCT, as the basis image of 2-D DCT is the outer-product of two 1-D DCT basis vectors. The above 1-D DCT can be extended to express 2-D DCT as (2.3), where $\mathbf{F} = [\mathbf{F}(i, j), (i, j) \in \mathcal{N}^2]$ represents a matrix of pixel intensities in a block, $\mathbf{X} = [\mathbf{X}(u, v), (u, v) \in \mathcal{N}^2]$ represents the matrix of DCT coefficients corresponding to $\mathbf{F}$ and $\mathcal{N}^2$ denotes the 2-D space. The DCT coefficient $\mathbf{X}(0, 0)$ is the DC coefficient, representing the average intensity of pixels in the block, and the others are AC coefficients.

$$\mathbf{H}[u, n] = a(u) \cos \frac{(2n+1)u\pi}{2N} \quad 0 \leq n < N \quad (2.1)$$

$$\text{with } a(u) = \begin{cases} \sqrt{\frac{1}{N}} & u = 0 \\ \sqrt{\frac{2}{N}} & 1 \leq u < N \end{cases}$$

$$\mathbf{X}[u] = \sum_{n \in \mathcal{N}} \mathbf{H}[u; n] \mathbf{F}[n] \quad (2.2)$$

$$\mathbf{X} = \mathbf{H}^T \mathbf{F} \mathbf{H} \quad (2.3)$$

DCT can concentrate the energy of images into a small number of transform coefficients. This advantage is shown in Fig 2.2 by two examples of variance (indicating

CHAPTER 2. BACKGROUND AND RELATED WORK

energy) of 8×8 DCT coefficients at different frequencies. The left figures are original pictures from the Foreman sequence, of which the upper one is intra-coded and the lower one is inter-coded. The right figures show the variance of DCT coefficients at different frequencies. It can be seen that only a small number of DCT coefficients have comparatively large variance. These DCT coefficients with small variance have small magnitudes correspondingly. They are considered insignificant in terms of statistical and subjective measures, so they are usually quantized in a low quantization resolution and need not be coded for transmission [13]. Therefore, this property of DCT allows compressed video information to be represented by a few DCT coefficients with larger magnitudes.

Furthermore, DCT is an orthogonal transform in which DCT coefficient is the difference of the value of original pixels. DCT coefficients have less correlation compared with original image pixels, so quantization of DCT coefficients is more efficient than quantization of the original pixels themselves [66].

The inverse DCT is then $\mathbf{F} = \mathbf{H}^{-1}\mathbf{X}$. Because DCT is a real unitary transform, $\mathbf{H}^{-1} = \mathbf{H}^T$ and IDCT can be expressed by $\mathbf{F} = \mathbf{H}^T\mathbf{X}$. There is no information loss in the operation of DCT, because an inverse DCT can be performed to reproduce original pixels in the spatial domain.

2.1.1.3 Quantization

After the DCT is conducted, quantization is carried out to considerably reduce the magnitude of the DCT coefficients. In this stage, a DCT coefficient is mapped to an integer by rounding off the ratio between the coefficient and the quantization stepsize. There is no way to reproduce the original value from the quantized value, so quantization is an irreversible process and information is lost to some extent. Consequently, quantization has a significant impact on the distortion of compressed video and the encoding bitrate. Coarse quantization results in poorer reproduction quality and a lower bit rate compared with fine quantization.

CHAPTER 2. BACKGROUND AND RELATED WORK

2.2.a: Original Foreman picture

2.2.b: Variance of DCT coefficients in an intra-coded frame

2.2.c: Original Foreman picture

2.2.d: Variance of DCT coefficients in an inter-coding frame

Figure 2.2: Examples of DCT coefficient variance at different frequencies

The quantization in video coding is designed in a block-based manner. Every block consisting of DCT coefficients is quantized by a QP multiplied by the quantization matrix. The quantization matrices are defined prior to video coding. Usually, the default quantization matrices provided in the standards are employed. They are set to represent lower frequency coefficients in more detail, because the human viewer is more sensitive to the reconstruction errors related to low spatial frequencies than those related to high frequencies [67]. The quantization matrix for intra blocks is different from the quantization matrix for inter-blocks. For instance, the typical MPEG quantization method

CHAPTER 2. BACKGROUND AND RELATED WORK

[40] can be described by (2.4), where $\mathbf{X}(u, v)$ represents the DCT coefficient, Q denotes the quantization parameter and $\mathbf{Y}(u, v)$ represents the corresponding quantization result of $\mathbf{X}(u, v)$. Once QP is obtained, quantization can be carried out by the means of the pre-defined quantization scheme. QP is the control parameter responsible for adjusting the R-D output of a video coder, so the selection of QP is the main concern in the decision-making of the coding strategy.

$$\mathbf{Y}(u, v) = \begin{cases} \lfloor \frac{\mathbf{X}(0, 0)}{Q} + 0.5 \rfloor, & \text{if } \mathbf{X}(u, v) \text{ is DC in an intra block} \\ \lfloor \frac{\frac{16\mathbf{X}(u, v)}{\mathbf{M}^0(u, v)} + \frac{3}{4}Q}{2Q} \rfloor, & \text{if } \mathbf{X}(u, v) \text{ is AC in an intra block} \\ \lfloor \frac{16\mathbf{X}(u, v)}{\mathbf{M}^1(u, v)} \rfloor, & \text{if } \mathbf{X}(u, v) \text{ is AC in a non-intra block} \end{cases} \quad (2.4)$$

where,

$\mathbf{M}^0$ is the intra quantization matrix,

$$\mathbf{M}^0(u, v) = \begin{bmatrix} 8 & 17 & 18 & 19 & 21 & 23 & 25 & 27 \\ 17 & 18 & 19 & 21 & 23 & 25 & 27 & 28 \\ 20 & 21 & 22 & 23 & 24 & 26 & 28 & 30 \\ 21 & 22 & 23 & 24 & 26 & 28 & 30 & 32 \\ 22 & 23 & 24 & 26 & 28 & 30 & 32 & 35 \\ 23 & 24 & 26 & 28 & 30 & 32 & 35 & 38 \\ 25 & 26 & 28 & 30 & 32 & 35 & 38 & 41 \\ 27 & 28 & 30 & 32 & 35 & 38 & 41 & 45 \end{bmatrix} \quad (2.5)$$

$\mathbf{M}^1$ is the inter quantization matrix,

$$\mathbf{M}^1(u, v) = \begin{bmatrix} 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 \\ 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \\ 18 & 19 & 20 & 21 & 22 & 23 & 24 & 25 \\ 19 & 20 & 21 & 22 & 23 & 24 & 26 & 27 \\ 20 & 21 & 22 & 23 & 25 & 26 & 27 & 28 \\ 21 & 22 & 23 & 24 & 26 & 27 & 28 & 30 \\ 22 & 23 & 24 & 26 & 27 & 28 & 30 & 31 \\ 23 & 24 & 25 & 27 & 28 & 30 & 31 & 33 \end{bmatrix} \quad (2.6)$$

CHAPTER 2. BACKGROUND AND RELATED WORK

2.1.1.4 Entropy-coding

Quantized DCT coefficients, which are the set of smaller integers compared with original DCT coefficients, still have some redundancies. In order to further increase the compression ratio, entropy-coding techniques are applied to represent those quantized DCT coefficients by the utilization of their statistical properties. The commonly-used methods of entropy coding are run-length-coding (RLC) and variable-length-coding (VLC) [68]. Entropy coding techniques are lossless, so they will not introduce any distortion in the video encoding.

Eventually, the bitstream goes into a buffer before transmission, where the buffer can smooth the generated variable bitrate of different frames and also provides a feedback to control coding behaviors of coming frames. After one frame is encoded, if it is a reference frame (either an I-frame or a P-frame), it is stored for the coming frames to perform motion estimation and motion compensation. The decoder, as shown in Fig 2.1.b, executes the reverse process of the encoder to reproduce the video sequence.

2.1.2 SNR Scalable Coder

Scalable coding can be viewed as an extension of non-scalable coding. In scalable coding, the base-layer is encoded at a low bitrate to provide a coarse quality, so the received video quality will not degrade drastically if only the enhancement-layer data are lost; and the overall video quality can be enhanced when extra bandwidth is available for the enhancement-layer bitstream. Compared with non-scalable coding, scalable coding can provide more scalable and more flexible service to satisfy different video quality requirements from different clients [69]. Moreover, its layered nature increases the capability of dealing with the problems during multimedia communication such as packet loss, congestion and so on [70–72]. Generally, there are three basic scalable mechanisms, namely SNR, spatial and temporal scalability, referring to the scalability of quality, image size

CHAPTER 2. BACKGROUND AND RELATED WORK

and frame rate respectively. In a wider sense, data partition [73] is also considered as another scalable mechanism, although it is not truly a scalable encoding process and only partitions the non-layered video data into different parts according to their importance. SNR scalable coding is investigated in this thesis.

SNR Scalable video coding was originally proposed to increase the robustness of video coders against packet/cell loss in ATM network [18]. It has been extensively studied since it was incorporated into the MPEG-2 video coding standard. The approach of SNR scalability generates at least two layers of video bitstreams that provide the same spatial-and-temporal resolutions but different levels of video quality. The base-layer encodes the raw video sequence to provide the basic video quality, and the enhancement-layer encodes the residues of the base-layer to enhance the SNR of decoded pictures of the base-layer, so it is called SNR scalability [74]. According to the technique employed for dealing with DCT coefficients, we categorize standard SNR scalable coders into the QP-based SNR coder and the bitplane-based FGS coder. The QP-based SNR coder follows the conventional coding of DCT coefficients that quantizes DCT coefficients with a given QP. The BP-based FGS coder adopts the BP coding technique [57] by the means of encoding binary representations of DCT coefficients.

2.1.2.1 QP-based SNR Scalable Coder

The conventional QP-based SNR scalable coder refers to the SNR scalable coder whose enhancement layer adopts the same quantization scheme as the common non-scalable coder that rounds off DCT coefficients to a set of smaller integers. Its structure [7] is shown in Fig 2.3.a. The structure of both the base-layer coder and the enhancement-layer coder is very similar to that of a non-scalable coder composed of DCT, quantizer and the entropy coder. The difference is the source to be encoded. The enhancement-layer is to encode the coding residues of the base-layer differing from raw video pictures for the

CHAPTER 2. BACKGROUND AND RELATED WORK

base-layer. Accordingly, to make the magnitude of the base-layer coding errors smaller, the enhancement-layer is supposed to use a finer quantizer which is implemented by the adoption of a smaller QP than the base-layer. In the events of I-frames or P-frames, compressed data of both the base-layer and the enhancement-layer are reconstructed in the encoder. The reconstruction frame is used as the reference frame for motion prediction and motion compensation.

2.3.a: A two-layer SNR scalable encoder with drift

2.3.b: A two-layer SNR scalable encoder without drift

Figure 2.3: Two-layer SNR scalable coder structure

CHAPTER 2. BACKGROUND AND RELATED WORK

Such a standard compliant coder in Fig 2.3.a has a drawback due to the tight coupling between two layer bitstreams [68]. The enhancement-layer also contributes to the base-layer motion compensation during the encoding. If there is data loss in the enhancement-layer bitstream, the enhancement-layer cannot be decoded entirely and cannot work properly on the motion compensation at the decoder. Therefore, decoded video of the base-layer will suffer from picture drift [68] that will cause picture quality degradation. The drawback of picture drift becomes the most severe when only the base-layer bitstream can be decoded. Thus, drift-free pictures are required and the solution can be achieved by loosening the tight coupling between the two layers.

Many sophisticated coder structures have been proposed in [68, 74–76] to eliminate such impairments. They decrease the coupling of the base-layer and the enhancement-layer while trying to maintain the coding efficiency as high as possible. A simple structure of those drift-free coders is shown in Fig 2.3.b, which cuts off the contribution of the enhancement-layer on motion prediction and motion compensation in the base-layer. In this work, the drift-free SNR scalable coder in Fig 2.3.b will be used as the test platform for the study on the rate control for SNR scalable coding in Chapter 5.

2.1.2.2 Bitplane-based FGS Coder

Another technique of coding the DCT coefficient in the enhancement-layer is bitplane (BP) coding [57]. The BP-coding based FGS coding [22, 23, 56, 77] has been adopted in the amendment of MPEG-4. The major difference of the BP coding technique from the conventional QP-based method is that the BP coding method considers each quantized DCT coefficient as a binary number of several bits instead of a decimal integer of a certain value. In detail, the magnitude of the DCT coefficient can be represented in a binary expression (2.7), where $|x|$ is the magnitude of the DCT coefficient x , and $\mathbf{b}_n$ is the binary coefficient corresponding to the stepsize of 2^n . And 2^n indicates the significance of $\mathbf{b}_n$,

CHAPTER 2. BACKGROUND AND RELATED WORK

for example, $\mathbf{b}_0$ corresponding to 2^0 is considered to have the least significance. Every BP_n is composed of all the $\mathbf{b}_n$ in the same significant position that are decomposed from all the DCT coefficients in the frame. For example, all the $\mathbf{b}_0$ from DCT coefficients in a frame construct BP_0 , and all the $\mathbf{b}_1$ construct BP_1 and so on. A frame has BP_{max} (2.8) number of BP, and BP_{max} depends on the maximum of all DCT coefficients (magnitude) in the frame denoted by x_{max} . The BP plane, BP_{max} , is taken as the most significant bit plane (MSB plane), followed by MSB-1 plane, MSB-2 plane, $\dots$. The scanning order of BP-coding is shown in Fig 2.4. The scanning of one frame starts from MSB to least significant bitplane (LSB). At the BP level, all the bitplane-blocks are scanned one by one in the traditional way, and those binary coefficients in one bitplane-block are zigzag ordered in an array. The BP coding technique encodes the binary representations of DCT coefficients; if the information of any binary coefficient is lost, it does not affect the decoding of the other received binary coefficients. Therefore, it has an advantage that once an additional bit is received, it can contribute to the overall picture quality enhancement. It should be noted that contribution of each bit to the overall picture quality is not the same, which will be investigated in Chapter 6.

$$|x| = \mathbf{b}_0 \cdot 2^0 + \mathbf{b}_1 \cdot 2^1 + \dots + \mathbf{b}_n \cdot 2^n + \dots \quad (2.7)$$

$$\text{BP}_{max} = \lfloor \log_2 |x_{max}| \rfloor + 1 \quad (2.8)$$

FGS adopts the BP-coding technique to encode DCT coefficients. A basic coding system with FGS has both the base-layer and the enhancement-layer, as shown in the Fig 2.5. The base-layer uses a non-scalable coder to encode the raw video sequence at a low quality. The enhancement-layer, namely the FGS-layer, encodes the DCT residues of the base-layer in a fine-granular mode by the BP coding technique. In the FGS layer, if there are some regions of interest or frequencies to be weighed, adaptive quantization

CHAPTER 2. BACKGROUND AND RELATED WORK

Figure 2.4: The scan order of enhancement-layer DCT coefficients in FGS coding

can be applied to highlight them by BP-shifting [78, 79] or frequency weighing techniques [80, 81].

Figure 2.5: FGS encoder

In a practical video streaming application, FGS coding can provide continuous variation in quality in accordance to the channel condition. On the server side, the base-layer is encoded at a lower bound of the bitrate range, R_B , and the FGS-layer is encoded at the bitrate up to the maximum available bandwidth R_{max} . The server has a high flexibility

to send the FGS-layer bitstream at any arbitrary bitrate in the range $[R_B, R_{max}]$. On the receiver side, the decoder can enhance the picture quality proportional to the received number of bits in the FGS-layer bitstream on the condition that the base-layer can be decoded properly.

2.2 Distributions of Source DCT Coefficients

Raw video sequences are the input of a video coding system. They are compressed in the DCT domain, so the knowledge on the statistics of source DCT coefficients is essential in analyzing the performance of a video coding system. Source DCT coefficients are original frames or motion-compensated frames that are transformed from the spatial domain to the DCT domain. Experimental DCT distributions are shown in Fig 2.6 for intra-frames and in Fig 2.7 for inter-frames, where the Foreman sequence is encoded. DCT coefficients in Fig 2.6 are taken from the average of 6 intra-frames. If the block size is 8×8 and pixels are 8-bit, the range of DC coefficients in intra-frames is $[0, 2048]$ and the range of AC(1,0) coefficients is $[-927.87, 927.87]$ according to the definition of DCT. In Fig 2.7, experimental data is taken from the average of 12 P frames and 24 B frames respectively when motion compensation in half-pixel accuracy was applied. The distribution of DCT coefficients in inter-frames, as shown in Fig 2.7, depends on both raw pixel values and the result of motion compensation.

In order to reveal the best statistical model for those source DCT coefficients, extensive experiments for natural images [41, 42, 55] and motion-compensated differential frames [44] have been carried on. And a mathematical justification has been presented in [43]. A widely-adopted conclusion is that the statistics of DCT coefficients for image and video are best approximated by a Gaussian distribution for DC coefficients and a Laplacian distribution for AC coefficients. The approximation is a tradeoff between the simplicity and accuracy to experimental data. The probability density function (PDF)

CHAPTER 2. BACKGROUND AND RELATED WORK

Figure 2.6: Percentage versus value of source DC coefficients in intra- frames.

of the Gaussian distribution and the PDF of the zero-mean Laplacian distribution are given by (2.9) and (2.10). The two distributions are also plotted in Fig 2.8 so as to act as a reference for actual distribution of source DCT coefficients. The proposed analysis on the distribution of DCT residues in Chapter 3 will be presented based on the two statistical models for source DCT coefficients.

$$f^G(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left[-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right] \quad (2.9)$$

where, μ_x : mean; σ_x^2 : variance

$$f^L(x) = \frac{\lambda}{2} \exp(-\lambda_x \cdot |x|) \quad (2.10)$$

where, $\lambda_x = \frac{\sqrt{2}}{\sigma_x}$

2.3 Classical Rate Distortion Theory

The classical rate-distortion (R-D) theory [16, 24, 25, 59, 82, 83] has provided the theoretical foundation for the analysis of a video coding system. R-D theory is a field of

CHAPTER 2. BACKGROUND AND RELATED WORK

Figure 2.7: Percentage versus value of source AC coefficients in Inter- frames

information theory that specifies the rate required to represent a source with some less-than-perfect fidelity. An essential topic of R-D theory is to study the minimum expected distortion achievable at a particular rate for a given source distribution and a given distortion measure; or equivalently to say, it studies the minimum rate description required to achieve a particular distortion [82].

A generic communication system can be illustrated by Fig 2.9 where the vector $\mathbf{X}_N$ denotes the source sequence and $\hat{\mathbf{X}}_N$ is the corresponding reconstruction sequence. Since practical coders are infinite, it is impossible to evaluate the specific performance of indi-

CHAPTER 2. BACKGROUND AND RELATED WORK

Figure 2.8: The Gaussian PDF and the Laplacian PDF

vidual coders. Significantly, R-D theory points out if the source is stationary and ergodic, there is a monotonically non-increasing distortion-rate function $D(R)$, or equivalently to say that, there exists a monotonically non-decreasing rate-distortion function $R(D)$.

Figure 2.9: R-D system from source to user

The performance of a coding system is analyzed by the means of the entropy and the distortion. The entropy of the source itself is called the average self-information. It is defined by (2.11) for discrete-amplitude memoryless sources, where $P(x_k) = P(X = x_k)$ is the probability mass function of the discrete random variable X . And the differential entropy for continuous-amplitude sources is given in (2.12), where $f(x)$ is the probability density function of a continuous random variable X . Another important concept is the average mutual information, which is also known as channel capacity of a communication

CHAPTER 2. BACKGROUND AND RELATED WORK

system. Its definition is given by (2.13) for discrete-amplitude memoryless sources and (2.14) for continuous-amplitude memoryless sources.

$$H(X) = E[I(X)] = - \sum_{k=1}^K P(x_k) \log_2 P(x_k) \quad (2.11)$$

$$h(X) = E[-\log_2 f(x)] = - \int_{-\infty}^{\infty} f(x) \log_2 f(x) dx \quad (2.12)$$

$$I(X; \hat{X}) = H(X) - H(X|\hat{X}) \quad (2.13)$$

$$I(X; \hat{X}) = h(X) - h(X|\hat{X}) \quad (2.14)$$

The R-D function in R-D theory is based on the concept of mutual information as a measure of the transmission of information from source to user. Besides the mutual information, the other necessary condition for the study of the rate-distortion performance is the distortion measure to define the reproduction fidelity. For example, one commonly-used distortion measure is mean-square-error (MSE) which is given by (2.15). With a given distortion measure, the R-D function can be derived to describe the achievable lower and upper bound of the R-D performance (2.16) in theory [59, 82]. The R-D function for continuous-amplitude sources is usually utilized to analyze the R-D performance of video coders. Its lower bound is called the Shannon lower bound and it is given by (2.17) where distortion is measured by MSE. The upper bound corresponds to the R-D function of an ideal coder with the zero-mean memoryless Gaussian source, as given by (2.18) where σ_x^2 is variance of Gaussian source. Those theoretical R-D bounds are instructive in analyzing the performance of video coders in spite of being too optimistic for practical applications. For example, rate models in the well-known TMN10 and VM18 rate control algorithms [27, 29, 33] were derived based on the classical R-D function.

$$D_{\text{mse}} = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \hat{x}_i)^2 \quad (2.15)$$

$$R_L(D) \leq R(D) \leq R_G(D) \quad (2.16)$$

$$R_L(D) = h(X) - \frac{1}{2} \log(2\pi e D) \quad (2.17)$$

$$R_G(D) = \frac{1}{2} \log \frac{\sigma_x^2}{D} \quad 0 < D < \sigma_x^2 \quad (2.18)$$

In the analysis of practical video coding, the output bitrate of a video coder is estimated by the entropy of coded information at some reproduction quality evaluated by a given distortion measure. The peak signal to noise ratio (PSNR) is mostly used to assess the decoded picture quality. Its definition is given by (2.19) where X_i is the 8-bit value of the source picture pixel and $\hat{X}_i$ is the corresponding reconstruction of X_i . Since the R-D function for an idealized communication system does not consider the characteristics of video coding. Hence, there will be some mismatch between the result of the rate-distortion function and the actual coding rate of practical video coders, if it is applied to estimate the performance of a practical video coder. Therefore, the performance of video coders is usually modeled both on the basis of the classic R-D function and with the consideration of the characteristics of video coding.

$$\text{PSNR (db)} = 10 \log_{10} \frac{255 \times 255}{\frac{1}{N} \sum_{i=0}^{N-1} (X_i - \hat{X}_i)^2} \quad (2.19)$$

2.4 Related Work on R-D Modeling and Control for Video Coding

Various video coding standards specify only the syntax of the decoder. The flexibility is retained for the encoder to compress the raw video sequence at different bitrate corresponding to different visual quality. The encoding behaviors can be adjusted by the

means of changing the frame rate [84, 85], picture size [86, 87] and quantization resolution (picture fidelity) [27]. The adjustment of quantization resolution is the most popular way to control the output of the encoder, so the relationship between the QP and R-D output of the video coder has been extensively studied in order to achieve efficient video compression.

2.4.1 Approaches of R-D Modeling for Video Coding

The problem of controlling video coding is complex because of the variation of video contents, the variation of motion and the difficulty in coding strategy selection prior to obtaining exact R-D data of the current frame [88]. In previous R-D modeling for video coding, there are two basic approaches: the analytical approach and the empirical approach [34].

The analytical approach [27–29, 35, 89] formulates R-D behaviors by a set of mathematical functions that match the properties refined from the video coding system. The video coder is first viewed as a system composed of different functional modules to process the video data. Then, the statistical characteristics of the video source and the property of every module are characterized by existing theoretical models. With the joint consideration of them, a set of R-D functions are selected to describe the overall R-D performance. The empirical approach [30, 31, 90–94] extracts R-D models according to experimental R-D data collected from coded information. In the empirical approach, R-D models are obtained by mathematical processing of those collected data. The analytical approach can provide some insights into video coding behaviors that the empirical approach cannot [28, 34]. Moreover, the empirical approach has a drawback of high computational complexity and poor adaptability to the variation of video contents. It should be noted that the analytical approach usually employs an empirical method to compensate the mismatch between the theoretical model and the actual result.

2.4.2 Typical Rate Control Algorithms

In the study of efficient video compression, rate control is an important application of R-D modeling, because the output bitrate of source video encoder is usually constrained by the limited bandwidth or storage capacity, such as Internet video streaming. The objective of rate control is to maximize the video quality at a given bitrate. Many rate control algorithms have been proposed to control the bitrate of video coders. Some of them have become recommendation models for video coding standards, such as TM5 [38] for MPEG-2, TMN10 [39] for H.263, VM18 [40] for MPEG-4. Besides, the ρ domain rate control algorithm [95] has been proposed recently based on the linear relationship between the ultimate rate and " $1-\rho$ ", where ρ is the percentage of zeros among quantized transform coefficients. In the following, ideas of those well-known rate control algorithms are reviewed.

2.4.2.1 TM5 Rate Control Algorithm

The TM5 rate control algorithm [38] was proposed for MPEG-2 high bitrate video coding. In the scenarios of MPEG-2 video coding, a video sequence can be divided into group of pictures (GOP). The first frame of the GOP is the I-frame; and the rest are either P-frames or B-frames. The TM5 rate control algorithm works on one-by-one GOP. In more detail, first, a frame level bit allocation is performed. Target bits for each picture type t ($t \in \{I, P, B\}$), T_t , are computed by (2.20), where T_{min} ensures a minimum number of bits for each frame, R_r represents the remaining number of bits in the current GOP, X_t is the global complexity, K_t is a constant and N_t is the number of frames not yet encoded in the current GOP. The K_t is set to maintain consistent quality among frames, and the relationship between K_t and QP for different frame types is defined by (2.21), where default setting of K_I , K_P and K_B is 1.0, 1.0 and 1.4 respectively. The global complexity X_t is updated as (2.22) after a frame of the same type is encoded, where

CHAPTER 2. BACKGROUND AND RELATED WORK

MQ_t is the average QP of different MBs and S_t is the generated bits for the frame. Second, a reference QP is determined for MB in the current frame according to the occupancy of the virtual buffer and a uniform distribution of bits over all MB. Such an assumption that bits are uniformly distributed over all MB is an over-simplified solution rather than the fact. Third, an adaptive QP for a certain MB is computed according to spatial activities of MB. TM5 rate control tries to encode the zones with higher activity more accurately by allocating more bits.

$$T_t = \max \left\{ R_r \cdot \frac{X_t/K_t}{\sum_{t \in I, P, B} (N_t \cdot X_t/K_t)}, T_{min} \right\} \quad (2.20)$$

$$\frac{Q_I}{K_I} = \frac{Q_P}{K_P} = \frac{Q_B}{K_B} \quad (2.21)$$

$$X_t = MQ_t \cdot S_t \quad (2.22)$$

The R-D solution of the TM5 rate control algorithm can be concluded that coding bitrate R is inversely proportional to QP; and the distortion D increases linearly with QP. The R-D model with respect to QP is quite simple. Moreover, the TM5 rate control algorithm was proposed for a high bitrate coder. Thus, it has a large control error especially for low bitrate coders.

2.4.2.2 VM18 Rate Control Algorithm

The VM18 rate control algorithm [40] was developed for the MPEG-4 video coder. It employed a quadratic rate model of QP proposed in [27, 33]. In the derivation of its rate model, video source was assumed to have a Laplacian distribution and the distortion was measured by the absolute-error given by (2.24). The Shannon R-D lower bound with distortion measured by absolute-error was used, which is given by (2.23) [96], where λ is the model parameter of the Laplacian distribution. The Taylor series expansion of the R-D function is then simplified to be a quadratic R function of D . Eventually, the

CHAPTER 2. BACKGROUND AND RELATED WORK

quadratic rate model in (2.25) is obtained by substituting the quantization stepsize (q) for D , where a_1 and a_2 are model parameters. Such a rate model does not have any parameter relating to video source, so it is not scalable with video contents. To overcome this drawback, a factor of mean-absolute-difference (MAD) is added in the later version of the rate control. It is given by (2.26), where H is set for bits used for header, motion vectors and shape information [33].

$$R(D) = \ln\left(\frac{1}{\lambda \cdot D}\right) \quad 0 < D < \frac{1}{\lambda} \quad (2.23)$$

$$D(x, \hat{x}) = |x - \hat{x}| \quad (2.24)$$

$$R = a_1 \times q^{-1} + a_2 \times q^{-2} \quad (2.25)$$

$$\frac{R - H}{MAD} = a_1 \times q^{-1} + a_2 \times q^{-2} \quad (2.26)$$

The ideas of the VM18 rate control algorithm can be concluded as follows. Encoding bitrate (R) has a quadratic relationship with QP, and distortion (D) is deemed to be proportional to QP. The accuracy of VM18 rate control mainly depends on the model parameters of a_1 and a_2 in (2.25) that are adjusted according to information of several previous frames, so VM18 rate control does not have a strong adaptability to the variation of video contents, for example, in scene change. Additionally, VM18 rate control algorithm does not control the QP in the I-frames and the buffer control does not work for I- frames either, which is not realistic.

2.4.2.3 TMN10 Rate Control Algorithm

The TMN10 rate control algorithm [39] was developed for the H.263 video coder. The algorithm is based on the D-Q model in (2.27) and the R-Q model in (2.28) for high bitrate coding and (2.29) for low bitrate coding [29]. Because H.263 is a standard for low bitrate coding, the rate model (2.29) for low bitrate coding is combined with the

CHAPTER 2. BACKGROUND AND RELATED WORK

distortion model to estimate the possible R-D result prior to actual coding. Given a target bitrate, the optimal quantization solution for different MB is obtained by the Lagrangian multiplier optimization technique. The TMN10 rate control algorithm controls video coding at the MB level. The coding statistics of previous MB are utilized to update the model parameters for the current MB. The TMN10 rate control algorithm only controls P-frames; however, intra blocks may occur frequently when scene changes, which will introduce a large control error.

$$D(q) = \frac{q^2}{12} \quad (2.27)$$

$$R(q) = \frac{1}{2} \log_2 2e^2 \frac{\sigma^2}{q^2} \quad \frac{\sigma^2}{q^2} > \frac{1}{2e} \text{ high bitrate coding} \quad (2.28)$$

$$R(q) = \frac{2}{\ln 2} \frac{\sigma^2}{q^2} \quad \frac{\sigma^2}{q^2} \leq \frac{1}{2e} \text{ low bitrate coding} \quad (2.29)$$

The ideas of R-D modeling in the TMN-8 rate control algorithm can be concluded as follows. The distortion, measured by the MSE criterion, is modeled by a function of QP only and is the same as the distortion model in the case of high bitrate uniform quantization. The rate model is essentially from the classical rate-distortion function for a Laplacian source [16, 59].

2.4.2.4 ρ -domain Rate Control Algorithm

Different from conventional R-D models formulating the relationship between R-D and QP, a novel ρ -domain rate control algorithm [34, 35, 95, 97] was proposed to study the R-D with respect to ρ , the percentage of zeros among quantized transform coefficients. It was observed that zeros among quantized transform coefficients have a substantial effect on the final R-D output. Given that there is a one-to-one mapping between q and ρ according to the actual R- ρ curves, a linear rate control model was established. The basic idea of ρ -domain rate control is expressed in (2.30), where θ is a statistical constant

CHAPTER 2. BACKGROUND AND RELATED WORK

in the rate model. In further study of ρ -domain R-D analysis, a distortion model of ρ in (2.31) was proposed by fitting the parameters of its exponential model to the observed distortion curves, where σ^2 is the picture variance and α is a statistical parameter.

$$R = \theta \cdot (1 - \rho) \quad (2.30)$$

$$D(\rho) = \sigma^2 e^{-\alpha(1-\rho)} \quad (2.31)$$

The ρ -domain rate model is quite simple in terms of model complexity and it is also accurate. To the best of the author's knowledge, it can be considered as the most efficient R-D analytical framework, because video coders are designed to encode non-zeros among quantized coefficients physically and therefore the ρ -domain rate model works close to actual video coding behaviors. However, its distortion model is concluded based on the empirical observations. Additionally, the two model parameters, θ in the rate model and α in the distortion model, are not constant or stable especially in enhancement layer coding.

2.4.3 Quality Control

The main task of rate control is to maintain the encoder output bitrate at a value not higher than a predefined constraint. Thus, it usually encodes the raw sequence with variable frame complexity in the constant-bit-rate mode. Some other applications may prefer constant visual quality, and one example of those applications is the digital video disk (DVD). It has been found in perceptual experiments that human observers give worse rating for video with a larger quality variation [98], so video coded with small quality variation or even constant quality has been the focus in the research on quality control.

Quality control tries to control the distortion in a predefined range. The D-Q model (2.27) for uniform distributed quantization errors was the most popular for its simplicity.

CHAPTER 2. BACKGROUND AND RELATED WORK

However, this assumption is true only for high-bitrate video coding, and it has a large mismatch for low bitrate video coding. Besides the uniform distribution model for coding error, a Gaussian distribution was proposed in [48] to model the coding error in the low bitrate coding by fitting its model parameters to collected experimental data. It is essentially an empirical method and has not provided us with the proper understanding on the distribution of coding errors. Some algorithms employed pre-processing [99] or two-pass [98] methods to gather information for successfully encoding video with constant quality. They are not feasible for real-time video coding which cannot tolerate long delay. In addition, a simple frame bit allocation scheme was used to allocate more bits to frames with larger MAD in [100, 101], and QP was selected empirically to provide one-pass quality control in [61].

Through the review on quality control, it can be seen that previous D-Q models could not provide accurate quality estimation; and accordingly, previous quality control algorithms were mainly based on empirical approaches which cannot enable us to develop accurate and simple control algorithms.

2.5 Summary

This chapter reviews basic technologies of video coding and related work on R-D analysis of video coding. First, an overview of video coding is presented by introducing the design of DCT-based video coders including non-scalable and SNR scalable coders. Then, well-known statistical models for source DCT coefficients are introduced to illustrate the input characteristics of a video coding system. Subsequently, the rate-distortion function in information theory is briefly reviewed to present the R-D relationship in theory. Lastly, approaches and typical work on R-D modeling and control for video coding are reviewed.

Chapter 3

Mathematical Analysis on the Distribution of DCT Residues

DCT is widely employed in many video compression schemes to represent the visual information in the frequency domain. The DCT-based video coder can compress raw video data efficiently while suffering from distortion to some extent due to quantization. Quantization errors generated in the quantization of DCT coefficients are known as *DCT residues*. These DCT residues can implicitly dictate visual quality of compressed video. Besides, they are the source of the enhancement-layer in SNR scalable coding [22, 102]. Therefore, knowledge on the distribution of DCT residues is of interest to analyze the performance of DCT-based video compression systems. It also makes sense in those applications regarding postprocessing of coded video [46] and so on. However, few analysis have been provided to illustrate what gives rise to the actual distribution of DCT residues. This chapter investigates the distribution of DCT residues and quantifies its relationship regarding video source and the quantization strategy. First, some limitations of previous study on the distribution of DCT residues is presented. Subsequently, individual frequency components of DCT residues are studied on the basis of quantization theory and well-accepted statistical models for source DCT coefficients. The probability density functions (PDF) of individual frequency components of DCT residues, as well as the overall PDF of DCT residues, are presented. Lastly, the goodness-of-fit test is done

to justify that the proposed distribution model fits well to the actual distribution of DCT residues.

3.1 Problem Description of Previous Study on the Distribution of DCT Residues

Since knowledge on the distribution of DCT residues is significant in understanding and modeling the behaviors of a video coding system, various studies have been conducted. Basically, they can be categorized into three approaches: a uniform distribution model, a conjectured model with parameters empirically obtained from collected residues data and qualitative analysis.

The first approach employs a uniform distribution model [27, 29, 33, 38]. This approach is widely adopted for its simplicity, and its application to R-D modeling is highlighted in our review. Those R-D modeling work usually presented their distortion models directly and did not assume any distribution for DCT residues explicitly. However, ideas behind distortion models in typical rate control algorithms can be briefly analyzed as follows. TM5 rate control [38] for high bitrate video coding uses a simple rate-quantization model and one essential idea is that distortion increases linearly with the quantization parameter. In TMN8 rate control [29] for H.263 video coding, distortion measured by mean-square-error is modeled as (3.1), where q is quantization stepsize. The distortion model (3.1) is the same as the formula for quantization-error variance when a uniform-distributed source is uniformly quantized with q . VM18 rate control for MPEG-4 video coding employs a quadratic rate-quantization model [27, 33]. In the derivation of its rate model, distortion measured by absolute-error was assumed to be proportional to q . By reviewing these distortion models, it can be concluded that the implicit assumption that DCT residues are uniformly distributed in a range solely defined by q has been employed. The reason for employing those simple distortion models is the simplicity

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

in terms of model complexity and implementation cost. However, its underlying idea, namely the uniform distribution, is an oversimplified model for DCT residues and has a large mismatch to the actual distribution.

$$D = \frac{q^2}{12} \quad (3.1)$$

According to uniform quantization theory, if inputs do not have a uniform distribution, quantization errors can be considered to have a uniform distribution only when quantization stepsize is infinitesimal [16]. The statistical distribution of source DCT coefficients in each frequency is rather a Gaussian or Laplacian distribution than a uniform distribution [41, 43]. Thus, the uniform distribution model is reasonable in the case of very small quantization stepsize corresponding to very high bitrate coding. Furthermore, the effect of statistics of source DCT coefficients on the distribution of DCT residues cannot be ignored in the same manner as the uniform distribution model does. For example, an experimental distribution of DCT residues in all frequencies is plotted in Fig 3.1, where the DCT residues are from one I-frame in Foreman. It shows that the experimental distribution is far from a uniform distribution. Therefore, the uniform distribution model is not a good description for the distribution of DCT residues especially in low bitrate video coding.

Another approach conjectures the distribution of DCT residues by more complex models empirically, since a simple distribution model is inadequate to express the distribution well. In [45], the generalized Gaussian distribution (GGD) was used. The PDF of GGD is given by (3.2), where ν is the shape parameter describing the exponential rate of decay, μ is the mean and σ is the standard deviation. G. Yovanof etc. [45] investigated the most approximate ν for a best fit to actual DCT residues by the well known non-parametric Kolmogorov-Smirnov and chi-squared (χ^2) tests [103]. In another paper [49], the mixture of two Laplacian distributions (one having smaller-variance and the other

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

Figure 3.1: The actual distribution of DCT residues.

having larger variance) was proposed. M. Dai etc. estimated the shape parameter of the two Laplacian distributions and the weight parameter defining the contribution of either Laplacian distribution to the overall distribution by the Expectation-Maximization algorithm [104]. These complicated models are more scalable than commonly-used models due to having the flexible shape parameter such as ν in [45] and the combination of λ_1 and λ_2 in [49]. Hence, they are able to express the distribution of collected DCT residues more accurately. However, this empirical approach has two major drawbacks. One is lots of computational complexity required to find 'good' model parameters. The other is the lack in adaptability to the variation of either video contents (e.g. scene change) or encoding strategies, because this approach constructs their model by using data from

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

previous frames and not the current frame. In addition, their models are still mathematically flawed. They assumed that the DCT residues were distributed over the range $(-\infty, \infty)$, but as a matter of fact, the actual DCT residues definitely have a bounded distribution range defined by the quantization stepsize. Hence, this approach cannot provide a correct representation of the distribution of DCT residues. Moreover, they are not very applicable to real-time video coding applications due to the high computational complexity.

$$p(x) = \left[\frac{\nu \eta(\nu, \sigma_x)}{2\Gamma(1/\nu)} \right] \exp(-|\eta(\nu, \sigma_x)(x - \mu)|^\nu) \quad (3.2)$$

$$\text{where } \eta(\nu, \sigma_x) = \sigma_x^{-1} \left[\frac{\Gamma(3/\nu)}{\Gamma(1/\nu)} \right]^{\frac{1}{2}},$$

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt, \quad x > 0$$

Besides the above two dominant approaches, paper [22] qualitatively mentioned the distribution of DCT residues cannot be modeled by a simple statistical model, but no in-depth insights into this topic has been presented. In [46, 105], the probability of DCT residues at certain quantities was studied based on the quantization of a Laplacian-distributed DCT coefficient. However, they were originally employed for removing blocking artifacts in the postprocessing of compressed video. Hence, their focus was on maximum a posteriori estimation when the reconstructed DCT coefficients were given. Moreover, no mathematically-tractable model has been provided to formulate the distribution of DCT residues with regard to source DCT coefficients and the quantization stepsize. Therefore, their study cannot be directly extended to estimate the R-D performance of video coders prior to actual video coding.

3.2 Proposed Distribution Model for DCT Residues

This section provides details of the modeling of the distribution of DCT residues by studying the compression of individual DCT components.

3.2.1 Frequency Processing of a Video Encoder

DCT is used to de-correlate original video data. In the light of this, original video data, viewed as a composite video signal, are decomposed into individual frequency components. The video coding system can then be understood as encoding those components separately, as illustrated in Fig 3.2 [28]. The coding process of each frequency coefficient can be taken as passing it through a quantizer. By considering all frequency components together, all video signals can be brought back together and therefore their coding behaviors can be analyzed and modeled.

Figure 3.2: Video coding system for individual frequency components

DCT residues are exactly quantization errors at different frequencies. The distribution of quantization errors depends entirely on the statistics of the source signal and the characteristics function of the quantizer. However, the complicated quantization scheme in the video coder results in a complex quantizer characteristics function, making it difficult to derive the probability density function (PDF) of DCT residues directly. To overcome this challenge, individual frequency components of DCT residues are studied instead due to the following two conditions.

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

The quantizer for individual frequency coefficients is a uniform scalar quantizer with a dead-zone. Its characteristic function is much simpler than that of the synthesized quantizer for all DCT frequency components, so it can facilitate deriving the PDF of quantization errors at individual frequencies. Such a uniform threshold quantizer (UTQ) can be defined by (3.3), where x is input, q is the quantization stepsize and t_0 ($t_0 \geq 0$) is the dead-zone to remove the noise around zero. For example, the MPEG-style quantization scheme can be written by (3.4) in terms of UTQ, where $\mathbf{q} = [\mathbf{q}^t(u, v)]$ represents a matrix of the quantization stepsize for a t -type (intra- or inter-) coded block of source DCT coefficient in the frequency (u, v) . $\mathbf{q}^t(u, v)$ is given by (3.5), where $\mathbf{M}^t$ denotes the quantization matrix as given in (2.5) and (2.6). In addition, (3.5) is rewritten by (3.6) merely in order for the simplicity of expression, where the $\mathbf{w}^t(u, v)$ is the corresponding frequency weight.

$$\text{UTQ}(q, t_0, x) = \begin{cases} 0, & |x| \leq t_0 \\ q \cdot \text{Round}[\frac{x - t_0}{q}], & x > t_0 \\ q \cdot \text{Round}[\frac{x + t_0}{q}], & x < -t_0 \end{cases} \quad (3.3)$$

$$Y(u, v) = \begin{cases} \text{UTQ}[\mathbf{q}^t(u, v), 0, \mathbf{X}(u, v)], & X(u, v) \text{ is DC in an intra block} \\ \text{UTQ}[\mathbf{q}^t(u, v), \frac{\mathbf{q}^t(u, v)}{8}, \mathbf{X}(u, v)], & X(u, v) \text{ is AC in an intra block} \\ \text{UTQ}[\mathbf{q}^t(u, v), \frac{\mathbf{q}^t(u, v)}{2}, \mathbf{X}(u, v)], & X(u, v) \text{ is in a non-intra block} \end{cases} \quad (3.4)$$

$$\text{where, } \mathbf{q}^t(u, v) = \frac{Q \cdot \mathbf{M}^t(u, v)}{8} \quad (3.5)$$

$$\mathbf{q}^t(u, v) = Q \cdot \mathbf{w}^t(u, v) \quad (3.6)$$

The statistics models for source DCT coefficients are another necessary condition for studying the distribution of DCT residues. As reviewed in Section 2.2, the statistics of DCT coefficients for image and video are best approximated by a Gaussian distribution

for DC coefficients and a Laplacian distribution for AC coefficients¹. The above analysis shows that it is feasible to derive the PDF of individual frequency components of DCT residues.

3.2.2 The Distribution of individual frequency components

All source DCT coefficients in the same frequency (u, v) are assumed to be quantized with the same quantization stepsize $q^t(u, v)$. In other words, source DCT coefficients in the same frequency (u, v) have the same quantizer. The assumption will facilitate the derivation of the distribution of quantization errors in individual frequencies. In the following, the distribution of DCT residues in individual frequencies will be investigated by studying quantization errors in the case that a Gaussian or Laplacian source is uniformly quantized.

DC component: source DC coefficients are modeled by a Gaussian distribution. However, it can be seen from Fig 2.6 and Fig 2.7 that the domain and variance of source intra- DC coefficients are much different from that of source inter- DC coefficients. Moreover, the quantizer for intra- DC component is a uniform quantizer without dead-zone (t_0 in (3.3) = 0), and the quantizer for inter- DC component has a dead-zone ($t_0 \neq 0$). They will result in different statistics of their own DCT residues. It can be seen from experimental distributions of DC residues in Fig 3.4, where DC residues are from Foreman frames quantized by the constant QP equal to 20. Therefore, the PDF of intra-DC residues and that of inter-DC will be discussed respectively.

The case of quantizing intra- DC is modeled that a Gaussian source is uniformly quantized by a quantizer with the quantization stepsize q and without dead-zone. Quantization errors is within the range $[-\frac{q}{2}, \frac{q}{2})$. The generalized PDF of quantization errors, f_{eG} , has been given by (3.7) in [106]. However, such an expression of f_{eG} in the form of

¹Modern digital video compression techniques process video data in terms of integers. For convenience, source DCT coefficients and DCT residues are assumed to be continuous variables in the analysis.

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

3.3.a: DCT(0,0) from one I-frame in Foreman

3.3.b: DCT(0,0) from one P-frame in Foreman

Figure 3.3: Examples: distribution of DC residues coefficients, percentage versus value.

the Fourier series consists of infinite items and certainly requires excessive computational complexity in practical applications. Hence, we should further study it in order to find a simplified expression.

$$f_{eG}(e, q) = \frac{1}{q} \left[1 + 2 \sum_{n=1}^{\infty} \cos \frac{2\pi n(e - \mu_x)}{q} \exp\left(-\frac{2\pi^2 n^2 \sigma_x^2}{q^2}\right) \right] \quad e \in \left[-\frac{q}{2}, \frac{q}{2}\right) \quad (3.7)$$

where, μ_x : the mean of Gaussian source

σ_x^2 : the variance of Gaussian source

The quantization stepsize (q) of the intra-DC quantizer is relatively small and the variance of source DC coefficients (σ_x^2) is quit large, so $\frac{\sigma_x}{q}$ is much larger than 1. On this condition, all the terms in (3.7) except for the first (the constant one) can be eliminated. Therefore, the PDF of intra-DC residues (f_{dc}^0) is modeled by a uniform distribution model as given by (3.8).

$$f_{dc}^0(e, q) = \frac{1}{q} \quad e \in \left[-\frac{q}{2}, \frac{q}{2}\right) \quad (3.8)$$

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

3.4.a: DCT(0,0) from one I-frame in Foreman

3.4.b: DCT(0,0) from one P-frame in Foreman

Figure 3.4: Examples: distribution of DC residues coefficients, percentage versus value.

In the case of quantizing inter-DC coefficients, it is not reasonable to assume that quantization errors have a uniform distribution. As a dead-zone is usually employed in the quantization of inter- DC coefficients, the formula (3.7) for the quantizer without dead-zone is not applicable either. If the dead-zone is taken into account in modeling the quantization error PDF, a mathematically tractable model cannot be achieved with a Gaussian source model. This drives us to study the distribution of inter- DC source to find an appropriate solution. Inter- DC coefficients are considered as Laplacian distributed like AC components due to a tradeoff between accuracy and model simplicity. To verify the plausibility, Kolmogorov-Smirnov (KS) Test (Appendix A) is employed to test experimental distributions of inter- DC coefficients against Gaussian and Laplacian distributions. The result is given in Fig 3.5, where the Foreman sequence is encoded in P frames except for the first I- frame. The figure shows the Gaussian distribution works slightly better than the Laplacian distribution in terms of the overall statistics. It also suggests that it is feasible to model the inter- DC by the Laplacian distribution, which can be observed from the actual distribution in Fig 2.7. In this way, the PDF of

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

inter- DC can be derived as (3.9) by studying quantization errors of a Laplacian source (Appendix. B).

Figure 3.5: Experimental distribution of inter- DC coefficients against the Laplacian and Gaussian distributions,

$$f_{dc}^1(e, q, \lambda_x) = \begin{cases} \frac{\lambda_x}{2} \exp(-\lambda_x |e|) & e \in [-\frac{q}{2} - t_0, -\frac{q}{2}) \cup [\frac{q}{2}, \frac{q}{2} + t_0) \\ \frac{\lambda_x \{ \exp(\lambda_x |e| - \lambda_x t_0 - \lambda_x q) + \exp(-\lambda_x |e|) \cdot [1 - \exp(-\lambda_x q) + \exp(-\lambda_x q - t_0)] \}}{2[1 - \exp(-\lambda_x q)]} & e \in [-\frac{q}{2}, \frac{q}{2}) \\ 0 & \text{otherwise} \end{cases} \quad (3.9)$$

With the PDF of intra-DC residues (3.8) and the PDF of inter-DC residues (3.9), the PDF of DC residues is summarized as (3.10), where superscript t is set to indicate the coding type (intra- or inter-).

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

Figure 3.6: Examples: distribution of AC residues coefficients, percentage versus value.

$$f_{dc}^t(e, q, \sigma_x) = \begin{cases} \frac{1}{q}, & t = 0 \\ f_{dc}^1 \text{ in (3.9),} & t = 1 \end{cases} \quad e \in \left[-\frac{q}{2}, \frac{q}{2}\right) \quad (3.10)$$

Each AC component: All the DCT components except the DC component are AC components, so the distribution of AC residues plays an important impact on the overall distribution of all DCT residues. Examples of the distribution of AC residues are shown in Fig 3.6. The process of compressing each AC component can be modeled as a Laplacian source passing through a uniform threshold quantizer. The PDF of AC residues, denoted by f_{ac} is given by (3.11) and the derivation is in Appendix B. It is the same as the mentioned inter- DC PDF. A theoretical plot of Fig 3.6 is given in Fig 3.7 to show the relationship between f_{ac} and the value of quantization error e . If the uniform quantizer has no dead-zone, namely $t_0 = 0$, f_{ac} can be rewritten as (3.12) which can be considered as a special case of (3.11). In practice, the dead-zone t_0 for inter- AC and intra - AC may be different, we can substitute the corresponding t_0 according to the pre-defined quantization scheme such as the example given in (3.4).

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

Figure 3.7: PDF of Quantization error in the case of uniformly quantizing a Laplacian source, $f_{ac}(e, q, \lambda_x)$, with $t_0 = 0$ and $\lambda_x = 0.2$.

$$f_{ac}(e, q, \lambda_x) = \begin{cases} \frac{\lambda_x}{2} \exp(-\lambda_x |e|) & e \in [-\frac{q}{2} - t_0, -\frac{q}{2}) \cup [\frac{q}{2}, \frac{q}{2} + t_0) \\ \frac{\lambda_x \left\{ \exp(\lambda_x |e| - \lambda_x t_0 - \lambda_x q) + \exp(-\lambda_x |e|) \cdot [1 - \exp(-\lambda_x q) + \exp(-\lambda_x q - t_0)] \right\}}{2 [1 - \exp(-\lambda_x q)]} & e \in [-\frac{q}{2}, \frac{q}{2}) \\ 0 & \text{otherwise} \end{cases} \quad (3.11)$$

$$f_{ac}(e, q, \lambda_x)|_{t_0=0} = \frac{\lambda_x}{2 \cdot [1 - \exp(-\lambda_x q)]} [\exp(\lambda_x |e| - \lambda_x q) + \exp(-\lambda_x |e|)] \quad e \in [-\frac{q}{2}, \frac{q}{2}) \quad (3.12)$$

3.2.3 The PDF for All the DCT Residues

The PDF of individual frequency components have been presented on the assumption that each frequency component is quantized with the same quantization stepsize. Before

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

the PDF for all DCT residues, f_e , is given, the following two factors need to be elaborated. First, if all source DCT coefficients in the same frequency (u, v) but belonging to different blocks adopt the same $q^t(u, v)$, it corresponds to a constant QP for a frame. It is consistent with our assumption and hence no modification is needed. However, QP of different blocks may not be the same. It indicates that $q^t(u, v)$ for all the source DCT coefficients $X(u, v)$ belonging to different blocks may be different. To be rigorous, the average quantization stepsize $\bar{q}^t(u, v)$ is substituted for $q^t(u, v)$. Such a substitution is feasible, as typical control strategies such as [29, 40] limit QP to vary in a small range for maintaining comparatively constant visual quality within a frame. Second, intra-coded blocks may occur in the inter-frames sometimes due to no matching reference blocks [40] or for the purpose of error resilience [107]. To make our model more complete, such cases are taken into account, since residue statistics in intra blocks are largely different from residue statistics in inter blocks, as shown in Fig 3.4. The solution is to treat those inter-blocks and intra-blocks as two independent sources. The PDF of residues in the frequency (u, v) can be obtained by adding up the PDF of intra-blocks residues and the PDF of inter-blocks residues according to their percentages. Hence, $f_{(u,v)}$, representing the PDF of quantization error $\mathbf{E}(u, v)$, can be rewritten by (3.13) where superscript t denotes the block coding type and p^t is the percentage of t -type blocks.

$$\begin{aligned} f_{(u,v)}(e) &= \sum_t p^t \cdot f_{(u,v)}^t[e, \bar{q}^t(u, v), \sigma_x^t(u, v)], \\ \text{subject to: } &p^0 + p^1 = 1 \\ \text{where, } f_{(u,v)}^t &= \begin{cases} f_{dc}^t & (u, v) = (0, 0) \\ f_{ac}^t & (u, v) \neq (0, 0) \end{cases} \end{aligned} \quad (3.13)$$

The complete PDF of DCT residues in individual frequencies has been given above. The cumulative distribution function (CDF) of DCT residues in all frequencies can be synthesized to be (3.14). The PDF is the derivative of the CDF and is then given by (3.15). In practice, QP, instead of quantization stepsize in individual frequencies, is the commonly-used control parameter. Recall (3.6), there is one to one mapping between

QP and $q^t(u, v)$. Combing (3.6) and (3.15), the PDF of DCT residues with regard to QP can be obtained by substituting $\overline{\text{QP}} \cdot \mathbf{w}(u, v)$ for $\bar{q}(u, v)$ in (3.15). It can be rewritten by (3.16) with all the parameters.

$$F_e(e) = \frac{\sum_{(u,v) \in \mathcal{N}^2} \int_{-\infty}^e f_{(u,v)}(x) dx}{\sum_{(u,v) \in \mathcal{N}^2} \int_{-\infty}^{\infty} f_{(u,v)}(x) dx} \quad (3.14)$$

$$f_e(e) = F'_e(e) = \frac{1}{N^2} \sum_{(u,v) \in \mathcal{N}^2} f_{(u,v)}(e) \quad (3.15)$$

$$f_e(e, \bar{Q}, \sigma_{\mathbf{x}}) = \frac{1}{N^2} \sum_{(u,v) \in \mathcal{N}^2} f_{(u,v)}[e, \bar{Q} \cdot \mathbf{w}(u, v), \sigma_{\mathbf{x}}(u, v)] \quad (3.16)$$

3.3 Experiments: Goodness-of-fit Test

To verify the accuracy of the proposed analysis on the distribution of DCT residues, goodness-of-fit of the proposed PDF is tested to compare with other typical probability models that were used for the study of DCT residues previously.

The KS test (Appendix A) is used to measure the goodness-of-fit. In the KS test, the empirical distribution function $G(z_i)$ is constructed by all the DCT residues in a frame. $G(z_i)$ is the actual distribution of DCT residues. The specified distributions are Gaussian, Laplacian, GGD and the proposed distribution model respectively. The cumulative distribution function (CDF) of the specified distribution, denoted by $F(z_i)$, is calculated by the integral of its PDF. To be rigorous in mathematics, DCT residues are discrete integers rather than continuous variables which are usually supposed in the analysis. $F(z_i)$ at the integer point z_i is calculated by (3.17), where $f(e)$ represents the specified PDF.

$$F(z_i + 0.5) = \int_{-\infty}^{z_i + 0.5} f(e) de \quad (3.17)$$

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

Model parameters of the specified PDF $f(e)$ are necessary to compute $F(z_i)$. To distinguish between the proposed distribution model and traditional statistical models, parameters of different PDFs and the approach to calculate these model parameters are explained in the following. The Gaussian model parameters are the mean (μ_e) and the variance (σ_e^2) of DCT residues. The Laplacian model parameters are μ_e and $\lambda_e = \sqrt{2}/\sigma_e$. Both Gaussian and Laplacian distributions depend on μ_e and σ_e^2 , while they have a different "skewness". The parameters, μ_e and σ_e^2 , are estimated by the mean and the variance of DCT residues in an usual manner. Besides μ_e and σ_e^2 , the PDF of GGD has another parameter ν responsible for controlling the "skewness" (3.2), for example, if $\nu = 1$, GGD reduces to a Laplacian distribution. In the following experiments, ν is the optimal solution of a KS test the same as presented in [45], and it is calculated by the iterative method. In other words, the optimal ν corresponding to the smallest KS test statistic is chosen to parameterize the GGD. The estimation of ν also requires that the whole set of DCT residues have been known already. In contrast to the statistical approach, model parameters of the proposed PDF are the matrix of source DCT variance in individual frequencies σ_x^2 and the matrix of quantization stepsize $\mathbf{q}$ in individual frequencies (or to say equivalently, QP). Our model can work without collecting any information from DCT residues.

Experiments are performed on an MPEG-4 coder adopting the MPEG- style quantization method. DCT residues are collected from coded Foreman and Carphone sequences (qcif, 4:2:0, 10 frame/s and 100 frames). The comparison on the average test statistics (t_{ks}) is plotted in Fig 3.8, where video sequences are encoded with different fixed QP. It can be seen that the t_{ks} of GGD is the smallest at all QP samples, which suggests that GGD with an optimal ν has the best match to the actual DCT residues. It is expected that GGD can achieve the best match, because GGD has a more flexible shape parameter ν making GGD optimized to approximate the actual distribution.

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

3.8.a: Carphone

3.8.b: Foreman

Figure 3.8: Comparison of average KS test statistic t_{ks} at different QP

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

However, optimal estimation of ν introduces lots of extra computational complexity. Different from the proposed PDF, GGD has the shape parameter (ν) whose optimal value varies with experimental data. In order to obtain the optimal value of ν , the KS test using GGD has to be repeated by enumerating ν , and then the value of ν with the minimal KS test statistics is considered as the optimal solution. The optimal ν is usually distributed in the range of $(0.2, 2.0]$. If the precision of ν solution is 0.01, 180 times of KS tests have to be carried out with 180 different value of ν . To give quantitative analysis, we tool the example of an inter-frame without intra-blocks. Table.3.1 lists the number of essential mathematical functions of the PDF of GGD (3.2) and the proposed PDF (3.11), where the number 64 in the column of "exp(x)" is the frequency dimension (the DCT block size) in the proposed PDF. Suppose the QP is equal to 16, the quantization stepsize for all frequency components is $2 \cdot \text{QP}$ provided that frequency weights are the same, and variance of the coded frame (σ_e^2) is 26.0 (corresponding to PSNR 34). In the KS test, the variable is assumed to be distributed in $[-3\sigma_e^2, 3\sigma_e^2]$ and the integration step to compute the CDF in (3.17) is 1. Total numbers of those mathematical functions which are executed in the KS test are given in Table.3.2. It shows that computational complexity by using GGD is much higher than the complexity by using the proposed PDF. Therefore, repetition of KS tests for the optimal solution of ν results in the high computation complexity with the GGD model. For a reference, value of the optimal ν solution at different QP is plotted in Fig 3.9 by averaging the optimal ν of all test frames. The variation of ν across QP indicates that DCT residues should be modeled by a scalable distribution model instead of by a simple statistical model.

Still seen from Fig 3.8, experimental results suggest the proposed model is more accurate than the Gaussian distribution model at all QP samples. Moreover, in the comparison to the Laplacian distribution, the proposed distribution model can also achieve a smaller test statistic at most of QP samples except the smallest QP sample. If we go

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

PDF model	number of functions		
	$\Gamma(x)$	$\exp(x)$	$x^{1/2}$
GGD	5	1 $(0 \leq x < \infty)$	1
proposed	0	0 $(x > \frac{q}{2} + t0)$	0
		64×1 $(\frac{q}{2} \leq x < \frac{q}{2} + t0)$	
		64×5 $(0 \leq x < \frac{q}{2})$	

Table 3.1: Number of essential mathematical functions in the PDF of the generalized Gaussian distribution (GGD) and the proposed PDF.

PDF model	number of executed functions		
	$\Gamma(x)$	$\exp(x)$	$x^{1/2}$
GGD	140,400	28,080	28,080
proposed	0	10572	0

Table 3.2: Example of total numbers of mathematical functions executed in the KS test with the GGD model and the proposed distribution model.

to check the optimal shape parameter ν plotted in Fig 3.9, it is found that ν is around 1, which approximates to a Laplacian distribution. It suggests that the Laplacian PDF is much like an optimal match to the actual DCT residues at the smallest QP sample. This is the reason why Laplacian works better than the proposed distribution model at the smallest QP sample in Fig 3.8. Anyway, it can be concluded with no doubt that the overall performance of the proposed distribution model is more scalable and more accurate than Laplacian and Gaussian distribution models. In addition, the actual distribution of DCT residues and the probability mass function (PMF) estimated by different distribution models are plotted in Fig 3.10. It provides us one example about the shape and skewness of different distributions.

Lastly in this section, it is worth highlighting that the proposed model is to "predict" the distribution of DCT residues prior to actual video coding rather than to "reproduce" the distribution of known DCT residues by statistical processing of DCT residues. It can explain what gives rise to the actual distribution of DCT residues in a quantitative method. It quantifies the distribution of DCT residues with respect to the quantization stepsize and with respect to source information. Practically, it can estimate the possible

Figure 3.9: Average optimal ν in GGD versus QP in Foreman

distribution of DCT residues with different quantization strategies for a given video source. This is an advantage over the traditional statistical models in R - D modeling and the control of one-pass video coding that requires effective adaptation to the variation of video contents and coding strategies, which will be presented in the following chapter.

3.4 Summary

The distribution of DCT residues is quantitatively analyzed in this chapter. After some limitations of existing work are addressed, the proposed work investigates on the distribution of DCT residues by studying individual frequency components of DCT residues. The PDF of DCT residues at different frequencies is derived based on understanding the distribution of source frequency coefficients and the uniform quantization theory.

CHAPTER 3. MATHEMATICAL ANALYSIS ON THE DISTRIBUTION OF DCT RESIDUES

Figure 3.10: PMF of DCT residues. The experimental DCT residues are from one P Foreman frame quantized with $QP = 16$, and the PMFs modeled by Gaussian, Laplacian, GGD and the proposed distributions are plotted.

Subsequently, the overall PDF for DCT residues is achieved by combining the PDF of individual frequency components together. Lastly, the KS test is done to verify the accuracy of the proposed work in comparison to Gaussian, Laplacian and GGD distributions. Experimental results suggest that the proposed distribution model is promising in the analysis for DCT residues. Besides the accuracy, the proposed model has another advantage that it can estimate the possible distribution prior to actual encoding. This will be useful to R-D analysis for one-pass video coding.

Chapter 4

R-D Modeling and Control for Non-scalable DCT Video Coding

R-D models are to analyze the R-D performance of a video coder and hence to facilitate the decision-making of coding strategies prior to actual video coding. R-D output of a video coder depends largely on the selection of quantization parameters (QPs). The relationship between R-D and QP, characterized by $R(Q)$ and $D(Q)$ functions, has been extensively studied [27, 29, 33, 38, 39]. As reviewed in Section 2.4.2, distortion models in standardized rate control algorithms [27, 29, 33, 38, 39] employed an over-simplified solution which is solely defined by QP. In spite of mathematical simplicity, they are not accurate and lack the adaptability to the variation of video contents. The proposed analysis on the distribution of DCT residues in Chapter 3 indicates that distortion of coded video depends on both statistics of video sources and the quantization strategy, since DCT residues, frequency-domain representations of spatial coding error, can dictate the quality of decoded video. This chapter investigates R-D modeling for DCT-based non-scalable coding. A novel distortion model is presented based on the proposed distribution model for DCT residues, and a rate model is obtained by combining the distortion model with the classic R-D function. Subsequently, a quality control algorithm is developed for video coding at target visual fidelity, and a rate control algorithm is developed to control video coding at the target bitrate.

4.1 R-D Modeling for Non-scalable Video Coding

In this section, a distortion model is proposed based on the probability density function (PDF) of DCT residues. Subsequently, a rate model is derived by combining the distortion model with the classic R-D function.

4.1.1 Distortion Model

Mean-squared-error (MSE) is used as the distortion criterion in this work. The MSE calculated in the spatial domain is equal to the MSE calculated in the DCT domain due to the unitary property of DCT [72]. Therefore, the distortion can be modeled by studying DCT residues instead of the decoded images in spatial domain. Only the stage of quantization causes information loss of visual information in video compression, so factors determining the magnitude and the distribution of quantization errors are essential in distortion modeling. Chapter 3 has investigated the distribution of DCT residues in terms of the PDF of DCT residues in individual frequencies. In the following, our distortion model is proposed based on the PDF of DCT residues concluded in (3.13).

If the PDF ($f(e)$) of a continuous random variable (e) is known, the expectation ($\text{EX}[e]$) and the variance ($\text{VAR}[e]$) can be calculated by (4.1) and (4.2) respectively. Let $\mathbf{E} = \{e(u, v) | (u, v) \in \mathcal{N}^2\}$ denote the matrix of DCT residues at different frequencies. It is easy to conclude that the expectation of DCT residues in individual frequencies ($\text{EX}[e(u, v)]$) is equal to zero, because the PDFs of individual frequency components, $f_{(u,v)}(e)$, are even functions with respect to e . Therefore, the distortion is equal to the variance of DCT residues, as the formula given in (4.3).

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

$$EX[e] = \int_{-\infty}^{\infty} ef(e) de \quad (4.1)$$

$$\text{VAR}[e] = \int_{-\infty}^{\infty} (e - EX[e])^2 f(e) de \quad (4.2)$$

$$D_{(u,v)} = \int_{-\infty}^{\infty} e^2 f_{(u,v)}(e) de \quad (4.3)$$

Combining the distortion formula (4.3) with the proposed PDF model given earlier in Chapter 3, the distortion model for individual frequency component can be obtained. Specifically, the distortion model for the DC component is derived as (4.4) by substituting the PDF of DC residues (3.10) into (4.3). The distortion model for AC components (4.5) can be derived by substituting the PDF of AC residues (3.11) into (4.3), where $(u, v) \neq (0, 0)$. In the case that the quantizer had no dead-zone ($t_0 = 0$), (4.5) can be rewritten by (4.6). It should be noted that t_0 is a constant totally dependent on the quantizer at individual frequencies. For instance, t_0 at different frequencies are usually different, as an intra- and an inter- weighting quantization matrices are set to highlight low frequency information. The normalized AC distortion with respect to the product of λ_x and q is plotted in Fig 4.1. From this figure as well as the model expression itself, it can be seen that the proposed distortion model estimates the distortion of AC components with taking both the source variance and the adopted quantization stepsize into account.

$$D_{(0,0)}^t(q, \sigma_x) = \begin{cases} \frac{q^2}{12} & (t=0) \\ \frac{2}{\lambda_x^2} \left[1 - \frac{\lambda_x q \exp(-\lambda_x t_0 - \frac{\lambda_x q}{2})}{1 - \exp(-\lambda_x q)} \right] - \exp(-\lambda_x t_0 - \frac{\lambda_x q}{2}) (t_0^2 + t_0 q + \frac{2t_0}{\lambda_x}) & (t=1) \end{cases} \quad (4.4)$$

$$D_{(u,v)}(q, \sigma_x) = \frac{2}{\lambda_x^2} \left[1 - \frac{\lambda_x q \exp(-\lambda_x t_0 - \frac{\lambda_x q}{2})}{1 - \exp(-\lambda_x q)} \right] - \exp(-\lambda_x t_0 - \frac{\lambda_x q}{2}) (t_0^2 + t_0 q + \frac{2t_0}{\lambda_x}) \quad (4.5)$$

$$D_{(u,v)}(q, \sigma_x) = \frac{2}{\lambda_x^2} \left[1 - \frac{\lambda_x q \exp(-\frac{\lambda_x q}{2})}{1 - \exp(-\lambda_x q)} \right] \quad (t_0 = 0) \quad (4.6)$$

Figure 4.1: Normalized distortion model of the AC component. Vertical axis: $D \cdot \frac{\lambda_x^2}{2} = \frac{D}{\sigma_x^2}$, horizontal axis: $\lambda_x \cdot q$

With those distortion models for individual frequency components, the overall distortion model can be obtained by the average distortion of all frequency components as (4.7).

$$D_q(\mathbf{q}, \sigma_{\mathbf{x}}) = \frac{1}{N^2} \sum_{(u,v) \in \mathcal{N}^2} \sum_{t \in \{0,1\}} p^t D_{(u,v)}^t[\mathbf{q}^t(u,v), \sigma_x^t(u,v)] \quad (4.7)$$

4.1.2 Rate Model

The output rate of an encoder is usually estimated by the entropy of coded video information. The rate-distortion function in information theory [16, 59, 82] is the mathematical foundation of analyzing the achievable R-D performance of a lossy communication system. It is often used to model the rate-distortion in practical video coding as well [27, 29].

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

If distortion is measured by squared-error, the rate-distortion function is given by (2.17) [16], where h_x is the differential entropy of the source and is given by (2.12). Individual frequency components of video signals are mostly uncorrected and they are looked on as independent sources in the modeling. The rate model for individual frequency components can be computed by substituting the proposed distortion model into (4.8) and computing h_x according to the statistical models of source DCT coefficients. Specifically, the rate of the intra-DC component is modeled in (4.9) by the rate-distortion function for a Gaussian source, and the rate of the inter-DC and AC components is modeled in (4.10) by the rate-distortion function for a Laplacian source. Based on the rate models for different frequency components, the overall rate model is given by (4.11).

$$R(D) = h_x - \frac{1}{2} \log_2(2\pi e D) \quad (4.8)$$

$$R_{(u,v)}^t = \frac{1}{2} \log_2 \frac{\sigma_x^{2^t}(u,v)}{D_{(u,v)}^t} \quad \text{if } (u, v, t) = (0, 0, 0) \quad (4.9)$$

where, $0 \leq D_{(u,v)}^t \leq \sigma_x^{2^t}(u, v)$

$$R_{(u,v)}^t = \frac{1}{2} \log_2 \frac{e \sigma_x^{2^t}(u, v)}{\pi D_{(u,v)}^t} \quad \text{if } (u, v, t) \neq (0, 0, 0) \quad (4.10)$$

where, $0 \leq D_{(u,v)}^t \leq \frac{e}{\pi} \sigma_x^{2^t}(u, v)$

$$R_q(\mathbf{q}, \sigma_x) = \frac{1}{N^2} \sum_{(u,v) \in \mathcal{N}^2} \sum_{t \in \{0,1\}} p^t R_{(u,v)}^t[\mathbf{q}^t(u, v), \sigma_x^t(u, v)] \quad (4.11)$$

4.1.3 R-D Model with respect to QP

The R-D models with respect to the quantization stepsize in individual frequencies $\mathbf{q}(u, v)$ have been presented. As the quantization parameter (QP) is the control parameter used to adjust the output of a video coder, the R-D model with respect to QP will be illustrated as follows.

The conversion between $\mathbf{q}(u, v)$ and QP can be formulated by (3.6) according to pre-defined quantization schemes in the standards. It should be noted that the distribution model for DCT residues is obtained based on the statistics of DCT coefficients from a whole frame. In other words, all those DCT coefficients in the same frequency are assumed to have the same quantizer. However, it is possible that all the blocks do not adopt a constant QP. It means DCT coefficients in the frequency (u, v) are probably quantized with different quantization stepsize $\mathbf{q}(u, v)$. In the same manner as used in Section 3.2.3, the average QP ($\bar{Q}$) is substituted for QP for the mapping between $\mathbf{q}(u, v)$ and the QP. In this way, the distortion model (4.7) and the rate model (4.11) can be rewritten by (4.12) and (4.13) respectively, where $\mathbf{w}^t(u, v)$ is the frequency weight.

$$D_Q(\bar{Q}, \sigma_x) = \frac{1}{N^2} \sum_{(u,v) \in \mathcal{N}^2} \sum_{t \in \{0,1\}} p^t D_{(u,v)}^t [\bar{Q}^t \cdot \mathbf{w}^t(u, v), \sigma_x^t(u, v)] \quad (4.12)$$

$$R_Q(\bar{Q}, \sigma_x) = \frac{1}{N^2} \sum_{(u,v) \in \mathcal{N}^2} \sum_{t \in \{0,1\}} p^t R_{(u,v)}^t [\bar{Q}^t \cdot \mathbf{w}^t(u, v), \sigma_x^t(u, v)] \quad (4.13)$$

4.2 Quality Control

In this section, the performance of the proposed distortion model is examined in the full range of QP. Subsequently, a quality control algorithm is developed based on the proposed distortion model to encode raw video sequences at a desired fidelity level.

4.2.1 Performance of the Proposed Distortion Model Versus QP

Performance of the proposed distortion model in (4.12) at different QP is tested to verify its accuracy. In the experiment, the result of the proposed distortion model is compared to the actual distortion and the result of the uniform distortion model (2.27). An MPEG-4 encoder is used to encode the Carphone and Foreman raw sequences. The first frame is

an I-frame and the rest are P frames encoded with different QP ranging from 1 to 31. The comparison among the actual distortion, distortion estimated by the uniform distortion model and the distortion estimated by the proposed distortion model is shown in Fig 4.2. Fig 4.2.(a) and Fig 4.2.(c) show the comparison for the I-frame, and Fig 4.2.(b) and Fig 4.2.(d) show the average performance of P-frames. It can be seen that the proposed distortion model is closer to the actual distortion at the most range of QP, especially for large QP. Thus, the advantage of the proposed distortion model over the traditional uniform distortion model is that it can estimate the distortion more accurately in the overall range of QP. This advantage indicates that the proposed distortion model can work well in a wide range of coding bitrate.

4.2.2 Video Coding at a Desired Fidelity Level

It is observed that the human visual system feels more comfortable with constant visual quality [98]. In some variable-bit-rate (VBR) video coding applications, visual quality is given priority over output bitrate. Existing quality control algorithms [60, 61] control video quality to be constant in an empirical approach. Different from them, a quality control algorithm is developed based on the proposed distortion model.

4.2.2.1 Quality Control Algorithm

When a theoretical distortion model is applied to practical video coding, it is usually found that there is mismatch to some extent between the actual result and the theoretical estimation. For a robust and accurate quality control, the distortion control model is written by (4.14) based on our distortion model (4.12), where K_D^t is set to make up for the model mismatch. Based on (4.14), a quality control algorithm is developed to encode every frame of a video sequence at a target fidelity level (D_T) in a one-pass manner. It can be illustrated by the flowchart in Fig 4.3.

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

4.2.a: intra-frame in Carphone

4.2.b: inter-frame in Carphone

4.2.c: intra-frame in Foreman

4.2.d: inter-frame in Foreman

Figure 4.2: Performance comparison of distortion models with respect to QP

$$D(\bar{Q}, \sigma_{\mathbf{x}}) = \frac{1}{N^2} \sum_{t \in \{0,1\}} \sum_{(u,v) \in \mathcal{N}^2} p^t D_{(u,v)}^t [\bar{Q} \cdot \mathbf{w}^t(u,v), \sigma_x^t(u,v)] \cdot K_D^t \quad (4.14)$$

Figure 4.3: Flowchart of the quality control algorithm

First, the model parameter K_D^t is initialized to be an empirical constant. 1.0 was used in the experiment. Variance of individual frequency components of a new frame is also computed for subsequent calculating QP.

Second, the QP solution for the new frame is calculated for the target distortion (D_T) according to (4.14). It is not easy to obtain the function, $\bar{Q}^t(D_T, \vec{\sigma}_x^t)$, by the

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

transformation of the distortion model in (4.14). Thus, the QP solution ($\overline{Q}^t$) cannot be computed directly by a formula of $\overline{Q}^t(D_T, \vec{\sigma}_x^t)$. From another point of view, the distortion increases monotonically with respect to QP, and QP is an integer distributed in the limited range [1,31]. Therefore, the iterative method is applicable to get an approximate QP solution for a given D_T . The pseudo code of the iterative algorithm is given in Fig 4.4, where $D(Q)$ is computed according to (4.14) and the initialized Q_x can be set as 16 or the same as the QP solution in the previous frame. The QP solution may not be an integer. In this case, a rounded QP solution will be adopted.

PRECISION: the threshold to end the loop

```

let  $Q_{max} = 31$ ,  $Q_{min} = 1$ ; initialize  $Q_x$ 

while  $|D(Q_x) - D_T| > \text{PRECISION}$ 
{
  if  $D(Q_x) < D_T$ 
  {
    if  $Q_x < Q_{max}$ 
    {  $Q_{min} = Q_x$ ;  $Q_x = \frac{Q_x + Q_{max}}{2}$ ; }
    else
    {  $Q_x = Q_{max}$ ; }
  }
  else
  {
    if  $Q_x > Q_{min}$ 
    {  $Q_{max} = Q_x$ ;  $Q_x = \frac{Q_x + Q_{min}}{2}$ ; }
    else
    {  $Q_x = Q_{min}$ ; }
  }
}

```

Figure 4.4: Iterative algorithm to get the QP solution at a target distortion level

Third, K_D^t is adjusted for the next frame, after one frame is encoded. It is updated according to the difference between the actual distortion and the result of the proposed distortion model. It is updated by (4.15) where i is the frame index, D_A denotes the actual distortion, D_E denotes the expected distortion calculated by the model, and α is

an empirical constant representing the correlation among consecutive encoded frames. α was set as 0.7 in the experiment. The updated K_D^t will be used for the next frame.

$$K_D^t(i) = \alpha \cdot \frac{D_A(i-1)}{D_E(i-1)} + (1 - \alpha) \cdot K_D^t(i-1) \quad (4.15)$$

4.2.2.2 Experimental Results of Quality Control

The proposed quality control algorithm is applied on an MPEG-4 video coder. Carphone, Foreman, News and Stefan raw sequences (qcif, 4:2:0, 300 frames) are used in the experiment. Every frame will be maintained at a desired fidelity level in terms of a specified PSNR value.

In comparison, another quality control algorithm, which is based on the uniform distortion model, is used to perform quality control as well. It is called uniform quality control and given by (4.16) where the model parameter (K_D^t) is defined the same as the proposed quality control algorithm in (4.14). It is also updated in the same manner as in the proposed quality control algorithm in order for fair comparison.

$$D(\bar{Q}) = \sum_t p^t \cdot \frac{\bar{Q}^2}{12} \cdot K_D^t \quad (4.16)$$

Test sequences are encoded at different target PSNR. The frame rate is 10 frames per second (f/s). The first frame is intra-coded, and the following frames are P frames. Average PSNR of coded frames by the proposed quality control, average PSNR by the uniform quality control and mean-squared-error (MSE) of PSNR are shown in Table.4.1. The MSE of PSNR is given by (4.17), where i is the frame index, PSNR_A is the actual PSNR of coded frames and PSNR_T is the target PSNR. In addition, PSNR of every frames is shown in Fig 4.5 and their corresponding frame bitrate is shown Fig 4.6. In the experiments with "Carphone" and "News", the proposed quality control algorithm and the uniform quality control have a similar performance; in some cases, the uniform quality

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

control can even achieve a little smaller quality control error. In the experiments with "Foreman" and "Stefan" sequences, the proposed quality control algorithm can always achieve a better performance in maintaining the quality to the target level.

$$\text{MSE}_{\text{psnr}} = \sum_i [\text{PSNR}_A(i) - \text{PSNR}_T(i)]^2 \quad (4.17)$$

video sequence	target PSNR	average PSNR		bitrate (kbit)		MSE of PSNR		
		UQC	PQC	UQC	PQC	UQC	PQC	gain
carphone	32	32.021	32.040	42.61	43.00	0.103	0.046	0.057
	35	35.001	35.048	71.49	73.23	0.100	0.166	-0.066
	38	37.883	38.080	123.56	126.71	0.097	0.208	-0.112
foreman	32	32.047	32.009	52.81	52.87	0.237	0.071	0.166
	35	35.124	35.000	88.50	87.05	0.271	0.137	0.133
	38	38.117	37.910	149.95	144.17	0.429	0.179	0.250
news	32	31.984	32.027	35.41	33.92	0.008	0.038	-0.030
	35	34.998	35.038	57.06	51.06	0.111	0.099	0.013
	38	38.096	37.970	94.83	78.68	0.090	0.178	-0.089
stefan	32	31.771	31.964	208.39	216.46	0.181	0.118	0.063
	35	34.830	34.914	326.71	330.21	0.201	0.105	0.096
	38	38.558	38.595	520.93	523.55	0.726	0.591	0.134

Table 4.1: Control error comparison, in terms of PSNR, between the proposed quality control (PQC) and the uniform quality control (UQC), when video sequences are encoded at a target PSNR level.

In the analysis of the experimental result, the effect of the adaptive parameter (K_D^t) should be included, because it contributes to the resultant video quality more or less. In some scenarios, it can maintain the quality constant even though the model has a large discrepancy. An ideal value for the adaptive value is 1.0, which means that the experimental result is absolutely the same as the result of the theoretical model. The experimental value at individual frames is plotted in Fig 4.7, when "Caphone" and "Foreman" sequences were encoded with a target PSNR of 35. It shows the value of K_D^t with the proposed quality is much closer to 1 than that with the uniform quality control. Some values with the uniform quality control are as large as up to 2, which indicates there was

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

4.5.a: Carphone

4.5.b: Foreman

Figure 4.5: PSNR comparison between the uniform quality control and the proposed quality control when raw sequences are encoded at the target PSNR of 35db.

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

4.6.a: Carphone

4.6.b: Foreman

Figure 4.6: Bitrate of individual frames when raw sequences are encoded at the target PSNR of 35db.

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

a very large discrepancy between the actual distortion and the theoretical model result. On the other hand, the variation of the adaptive parameter can indicate the adaptability of the distortion model. "Carphone" and "News" sequences has nearly no scene changes and no fast motion. The adaptive parameter was comparatively stable in rather long video segments. It indicates that the quality control was maintained in quite a stable state of an accurate estimation. However, in the experiments with sequences having more scene changes and fast motion, such as "Foreman" and "Stefan", the adaptive parameter with the uniform quality control varied a lot due to its poor adaptability to variation of video contents, while that with the proposed quality control varied in a comparatively small range. In other words, the proposed quality control can achieve a rather constant quality even though video contents vary much.

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

4.7.a: Carphone

4.7.b: Foreman

Figure 4.7: Value of adaptive parameter (K_D^t) in the quality control.

4.3 Rate Control

Rate control is a main concern in most video coding applications. For instance, when real-time video streams are delivered over a bandwidth-constrained network, if encoding bitrate exceeds the bandwidth, some portion of video packets will have to be lost. In consequence, visual quality at the receiver side will dramatically decrease. In the following, a rate control algorithm is developed for low-delay DCT-based video coding to control encoding bitrate at constant bitrate.

4.3.1 Rate Control Algorithm

The practical rate model for a frame is given by (4.18) [29, 40]. R_Q in (4.13) denotes those bits for encoding texture information. Since statistical models used to derive the rate model is approximate, there is some mismatch between the model and actual coding bitrate. K_R is a model parameter to make up for the mismatch. It is considered as a function of QP in (4.19), where K_{R1} and K_{R0} are coefficients which are obtained in a linear regression approach. C is a constant, standing for bits used for header information and syntax. Both K_R and C are adjusted adaptively according to coded information in previous frames.

$$R(\bar{Q}, \sigma_x) = R_Q(\bar{Q}, \sigma_x) + K_R(\bar{Q}) + C \quad (4.18)$$

$$K_R(Q) = K_{R1} \cdot Q + K_{R0} \quad (4.19)$$

Buffer Control:

The buffer control strategy in the VM18 rate control [40] is employed in the proposed rate control algorithm except that the buffer status is adjusted from the first frame. The adjustment is necessary, because the VM18 rate control does not calculate bits for the first

frame in the buffer, which is not realistic [37]. Thus, buffer fullness is updated starting from the first frame both in the VM18 rate control and the proposed rate control. The buffer feedback (B_f), which is used to adjust target bits for the current frame, is given by (4.20), where B_c is the current buffer level and B_s is the buffer size. The initial buffer level is set to be 50% of B_s , and B_s depends on the maximal accumulated delay. For example, if the maximal delay is 0.125s, B_s is equal to 0.25 multiplying the target bitrate (R_T). The next frame will be skipped if the current buffer level (B) is over $90\% \cdot B_s$.

$$B_f = \frac{B_c + 2 \cdot (B_s - B)}{2B_c + (B_s - B)} \quad (4.20)$$

Frame-level Rate Control flowchart:

The rate control flow can be described by Fig 4.8.

First, the rate model parameter C and current buffer status B_c are initialized. C is set empirically. B_c is set to 50% of B_s .

Second, a target number of bits for current frame (T_f) is calculated. T_f depends on the target bitrate (R_T) and the buffer feedback (B_f). It is computed by (4.21) where $\frac{R_T}{30}$ ensures the minimal number of bits to be assigned, R_r is the remaining number of bits in the sequence (or segment) and N_r is the number of P frames remaining in the sequence (or segment).

$$T_f = \max\left(\frac{R_T}{30}, \frac{R_r}{N_r}\right) \cdot B_f \quad (4.21)$$

Third, QP for macroblocks is calculated. In this step, QP, or equivalently to say $\bar{Q}$, is calculated at the frame level. The proposed rate model is complicated, so it is difficult to obtain the function, $\bar{Q}^i(R_T, \sigma_x^t)$ and then to compute the QP solution. The rate decreases monotonically with QP, and QP is an integer in the limited range $[1, 31]$.

Figure 4.8: Flowchart of the rate control algorithm

Thus, an iterative method can be applied to get an QP solution ($\overline{Q^t}$) for a given bitrate (R_t). Pseudo code is given in Fig 4.9. If the QP solution is not an integer, it will be rounded to the nearest integer and be the practical QP for macroblocks in the current frame.

```

PRECISION: threshold to end the loop;
 $R_t$ : allocated bitrate for the current frame
let  $Q_{max} = 31$ ,  $Q_{min} = 1$ ; initialize  $Q_x$ 

while  $|R(Q_x) - R_t| > \text{PRECISION}$ 
{
    if  $R(Q_x) < R_t$ 
    {
        if  $Q_x > Q_{min}$ 
        {  $Q_{max} = Q_x$ ;  $Q_x = \frac{Q_x + Q_{min}}{2}$ ; }
        else
        {  $Q_x = Q_{min}$ ; }
    }
    else
    {
        if  $Q_x < Q_{max}$ 
        {  $Q_{min} = Q_x$ ;  $Q_x = \frac{Q_x + Q_{max}}{2}$ ; }
        else
        {  $Q_x = Q_{max}$ ; }
    }
}

```

Figure 4.9: Iterative algorithm to get the QP solution at a target bitrate

Fourth, rate control parameters are updated for the following frames, after one frame is encoded.

4.3.2 Experimental Results of Rate Control

The proposed rate control is applied on an MPEG-4 encoder to control low-delay video coding. The encoder employs the MPEG style quantization strategy. Video sequences (4:2:0, qcif, 300 frames) are controlled in the frame-level in one-pass mode. The maximum accumulated delay is set to be 125ms and the frame rate is 10 frame/s. The first frame

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

is intra-coded with a pre-determined QP equal to 16, and the rest are P-frames. The first frame usually uses up a substantial amount of bits, and this probably results in the skipping of subsequent P-frames. In the following result comparison, the number of skipped frames is counted from the first encoded P-frame.

The VM18 rate control and the ρ -domain rate control¹ are compared with the proposed rate control. In order for fair comparison, experiments settings, such as the buffer control strategy and so on, are the same. To eliminate the impact of different optimization techniques applied at the MB-level², all rate control algorithms are conducted in the frame-level. In Table.4.2, the number of skipped frames and average PSNR are shown. Experimental results show that the proposed rate control achieves up to 0.15dB higher PSNR than VM18 rate control and up to 0.10dB higher PSNR than the ρ -domain rate control. The proposed rate control can achieve less number of skipped frames than VM18 rate control. In comparison to the ρ -domain rate control, numbers of skipped frames are the same with Carphone, Foreman and Stefan sequences. However, with the News sequence, the proposed rate control has one more skipped frame. In terms of the overall performance, the proposed rate control can surpass the VM18 rate control and has a similar performance with the ρ -domain rate control.

Table 4.2: Experimental result comparison between VM18 rate control (VM18), ρ -domain rate control (ρ -D) and the proposed rate control (new).

sequence name	target rate (kbit/s)	actual bitrate (kbit/s)			skipped frames			average PSNR (db)		
		VM18	ρ -D	new	VM18	ρ -D	new	VM18	ρ -D	new
Carphone	48.00	48.53	48.20	48.21	1	0	0	33.27	33.34	33.35
Foreman	64.00	63.93	63.77	64.05	2	0	0	33.09	33.20	33.24
News	32.00	32.00	32.05	32.04	2	0	1	31.62	31.64	31.74
Stefan	100.00	100.17	100.20	100.07	1	1	1	27.95	27.96	27.99

¹The slope θ in the ρ -domain rate model (2.30) is estimated by first-order linear regression. Namely, $\theta(\rho) = a \cdot \rho + b$, where a and b are parameters fitting the rate model to data from previous frames.

²At the MB-level, the VM18 rate control adjusts model parameters by using linear-regression [40]; the ρ -domain rate control increases/decreases the QP for remaining MBs adaptively according to information of coded MB within the current frame [95].

PSNR of different frames is shown in Fig 4.10. The frame rate is shown in Fig 4.11. And the buffer occupancy is shown in Fig 4.12, where the buffer fullness is a normalized number. In the experiments, after the first intra-frame was encoded, buffer occupancy increased to a level much larger than the half. Subsequently, the buffer control strategy tried to maintain the buffer occupancy at the half of buffer size. More constant buffer occupancy indicates more accuracy of rate control. It can be seen that the proposed rate control can achieve more constant buffer than the VM18 rate control, while the proposed rate control and the ρ -domain rate control have a very similar performance. Besides the picture quality comparison, computational complexity of ρ -domain rate control needs to be addressed. Conventional QP-based rate control algorithms usually need to calculate variance of the frame being encoded and then to compute the desired QP subject to a given bitrate. The ρ -domain rate control calculates a desired percentage of zero among quantized DCT coefficients (ρ) and then the corresponding QP solution. The ρ solution can be calculated with ease according to its simple rate model. However, there is no simple formula to map the ρ solution to a QP solution, so computation like quantization of current frame has to be repeated with different QP till the QP corresponding to the ρ solution is found. The mapping between ρ and QP will introduce extra computation complexity.

4.4 Summary

In this chapter, an R-D analysis framework for non-scalable DCT video coding is presented, and it is also applied to develop a quality control algorithm and a rate control algorithm. The proposed distortion model is derived based on the probability distribution model of DCT residues at individual frequencies. It is shown that the proposed distortion model is closer to the actual distortion than the existing distortion model in a wide range of QP. A quality control algorithm is developed to encode real-time video

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

Figure 4.10: PSNR comparison between VM18 rate control, ρ -domain rate control and the proposed rate control.

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

4.11.a: Carphone

4.11.b: Foreman

4.11.c: News

4.11.d: Stefan

Figure 4.11: Rate (bit per pixel) versus frame number by the VM18 rate control, ρ -domain rate control and the proposed rate control respectively.

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

4.12.a: Carphone

4.12.b: Foreman

4.12.c: News

4.12.d: Stefan

Figure 4.12: Buffer fullness comparison between VM18 rate control, ρ -domain rate control and the proposed rate control.

CHAPTER 4. R-D MODELING AND CONTROL FOR NON-SCALABLE DCT VIDEO CODING

at a desired fidelity level. Experimental results show that the proposed quality control algorithm works well to control the coded video quality to a target PSNR. Subsequently, the distortion model is combined with the classic R-D function to derive the rate model. Based on the proposed rate model, a rate control algorithm is developed to control video coding at a given bitrate. Experimental results justify the effectiveness of the proposed rate control. The proposed R-D framework has the advantage over previous work that it can adapt well to the variation of both the video source and the quantization resolution.

Chapter 5

Rate Control for Conventional SNR Scalable Coding

To cope with the challenges of fluctuating transmission bandwidth [71, 72] and different visual quality requirements from multiple clients [3], scalable coding has been incorporated into video compression standards to provide more scalable and more flexible service than non-scalable coding. Among different kinds of video coding scalability, SNR scalability allows for the delivery of two services with the same spatial and temporal resolution but different levels of quality. In order to maximize the visual quality that is provided by different layers of an SNR scalable coder, an efficient coding strategy is needed to control coding behaviors of the scalable encoder. However, most rate control algorithms are proposed for non-scalable coding, which are developed based on the R-D characteristics of non-scalable coding. Thus, those algorithms originally for non-scalable coding are not appropriate, or need to be improved, when they are applied to the enhancement-layer (EL) compression. In this chapter, rate control for the EL of the conventional SNR scalable coder is investigated. First of all, problems of existing rate control for the EL is addressed. And subsequently, a rate model is derived through analyzing some R-D characteristics of the EL compression. Lastly, the proposed rate control algorithm is applied to a drift-free MPEG-2 SNR scalable encoder.

5.1 Problem Description

In this section, the idea of the SNR scalable coder is explained. Subsequently, the problem of rate control for the enhancement-layer is addressed.

5.1.1 The Idea of SNR Scalable Coding

The detailed structure of the convenient SNR scalable coder has been introduced in Section 2.1.2.1. The conventional SNR scalability was incorporated into the MPEG-2 video compression standard. It is called QP-based SNR scalability to differentiate from recent bitplane-based FGS. The idea of the QP-based SNR scalable coder can be described by the block diagram in Fig 5.1. The base-layer (BL) encodes raw video data with a coarse quantizer to provide the basic and most vital video information. And the enhancement-layer (EL) encodes the BL residues with a finer quantizer to provide additional picture details. Both the BL and the EL adopt the same quantization scheme and the same entropy coder which are conformed to the quantization scheme and the entropy coder for the non-scalable coder [7]. The main differences between the BL and the EL are two points. The first difference is the source to be encoded. If it is viewed in the DCT domain, the BL source (X_B) is transform coefficients of raw video data, and the EL source (X_E) is DCT residues of the BL. The second difference is the value of the adopted QP. In order to enhance the picture quality of the BL, the QP in the EL (Q_E) should be smaller than the QP in the BL (Q_B); otherwise the EL will not reduce the magnitude of BL coding errors at all.

5.1.2 Problem Description of Rate Control for SNR Scalable Coding

When multiple layers of video bitstreams are delivered over channels with fluctuating quality-of-service or even without quality-of-service guarantee, there is a need to regulate

Figure 5.1: Quantizer in the SNR scalable coder

the output bitrate of different layers. At the same time, it is also desired to maximize the picture quality provided by multiple layers at the different levels of bitrate constraints. According to the design of the QP-based SNR scalable coder, the BL is similar to the general non-scalable coding, so it can usually be controlled by those well-known rate control algorithms [38, 40] proposed for non-scalable coding. However, there are some challenges when those traditional rate control algorithms are applied to the EL coding.

First of all, the EL has different sources and different quantization stepsize from those of the BL. It will result in different R-D characteristics. Examples of the probability distribution of source data both in the BL and in the EL are given in Fig 5.2. The base-layer source depends solely on raw video pictures. The enhancement-layer source is DCT residues in the BL. According to the analysis on the distribution of DCT residues in Section 3.2, the EL source depends on both raw video pictures and the base-layer quantization stepsize (Q_B). Moreover, the EL coder has the effect of quality enhancement. In order to encode the noise of BL pictures, the EL quantization stepsize (Q_E) should be smaller than the base-layer quantization stepsize (Q_B). Another difference is that some blocks of the EL can be skipped without dramatically degrading the visual quality, if no enough bits are available to encode those blocks.

Take another case of recent ρ -domain rate control [35, 95], which is much source-independent. The slope in the linear rate model, which is the ratio between the output

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

Figure 5.2: Probability distribution of AC (0,1) coefficients. (a)(c): the BL and EL in Mobile & Calendar respectively; (b)(d): the BL and EL in Stefan respectively

bitrate and the non-zero percentage among quantized coefficients, is nearly a constant and stable in non-scalable coding. However, it is observed to vary quite significantly in EL coding. Therefore, the ρ -domain rate model still needs to be improved when it is applied to the EL compression.

What's more, standardized rate control algorithms control the video coding at the macroblock level. The macroblock-level rate control is optimal bit allocation among all the macroblocks essentially. It is observed that if the traditional optimal-bit-allocation solution [16] is employed in EL coding, many macroblocks will probably be allocated a negative number of bits. However, the negative number of allocated bits are not reasonable. And the rate control error will be large by simply setting those negative values to zeros. Therefore, it is not accurate enough to apply the bit allocation scheme for non-scalable coding to the EL directly.

There are some R-D analysis work specially for scalable coding. The main focus is the decision of the quantization strategy in the EL. They did the R-D optimization by the Lagrangian multiplier optimization method [50, 108] or by employing the perceptual characteristics of the human visual system [20]. None of them gave us a practical scheme to control the bitrate of different layers. In another study [109], the rate control algorithm for the SNR scalable coder did not provide any analytical model. Therefore, the problem of rate control for scalable coding still needs to be investigated and given a better solution.

5.2 Some Useful Characteristics of EL Compression

To achieve a robust and accurate rate control scheme, there is a need to find an effective rate model based on the analysis of coding behaviors. In this section, some important characteristics regarding the R-D modeling in EL compression are presented.

5.2.1 Relationship Between R_E and r_Q

It is BL quantization errors that the EL is going to encode, so the EL source data should have much dependance on the BL quantization stepsize (Q_B). Thus, there should be a strong relationship between Q_B and the EL quantization stepsize (Q_E) for a given bitrate for the EL. The relationship between r_Q and the EL bitrate is investigated in the following, whereby r_Q denotes the ratio between Q_B and Q_E as given in (5.1).

$$r_Q = \frac{Q_B}{Q_E} \quad (5.1)$$

In the case of uniformly quantizing a given source, the distortion depends on both the source distribution and quantization stepsize. However in high bitrate coding, if distortion is measured by MSE, it can be formulated by a function of only quantization stepsize as given by (5.2) [16, 59]. Thus, the distortion of the BL (D_B) can be approximately expressed by (5.3) in high bitrate coding, where α_B can be understood to a parameter dependent on two factors. One is the relationship between the quantization parameter (QP) and the exact quantization stepsize (q). The other factor is for making up for the model mismatch. In a similar way, the distortion of the EL can be approximately expressed by (5.4). In mathematics, the variance of the EL difference frame pixels (σ_E^2) is the distortion of the BL measured by MSE. Using (5.3), σ_E^2 can be expressed by (5.5).

$$D = \frac{q^2}{12} \quad (5.2)$$

$$D_B(Q_B) = \alpha_B \cdot \frac{Q_B^2}{12} \quad (5.3)$$

$$D_E(Q_E) = \alpha_E \cdot \frac{Q_E^2}{12} \quad (5.4)$$

$$\sigma_E^2 = \alpha_B \cdot \frac{Q_B^2}{12} \quad (5.5)$$

The rate-distortion function in information theory is given by (5.6) [16], where σ^2 is the variance of the data source and the factor ϵ^2 is dependent on the probability density

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

function of data source. When it is applied to video coding, the type of adopted entropy coding also has an impact on the R-D performance and it can be considered as one component of ϵ^2 .

$$D(R) = \epsilon^2 \sigma^2 2^{-2R} \quad (5.6)$$

With the (5.3), (5.5) and (5.6), a group of formulas (5.7) can be obtained for R-D modeling. The approximate rate model of the EL can then be derived as (5.8). The rate function of $\frac{Q_B}{Q_E}$ is plotted in Fig 5.3.

$$\begin{cases} D_E(R_E) &= \epsilon_E^2 \sigma_E^2 2^{-2R_E} \\ \sigma_E^2 &= \alpha_B \frac{Q_B^2}{12} \\ D_E &= \alpha_E \frac{Q_E^2}{12} \end{cases} \quad (5.7)$$

$$R_E(Q_E) = \log_2 \frac{Q_B}{Q_E} + \frac{1}{2} \log_2 \left(\epsilon_E^2 \cdot \frac{\alpha_B}{\alpha_E} \right) \quad (5.8)$$

The distortion of the BL can be reduced only when $Q_E < Q_B$, and Q_E should be an integer larger than 1. Therefore, r_Q is in the range $[1, Q_B]$. In practice, r_Q varies in a small range, because either too large or too small quantization parameters are seldom used. Thus, the rate function in (5.8) can be simplified as (5.9).

$$R_E(Q_E) \propto (r_Q - 1) \quad (5.9)$$

To verify this relationship between the EL rate and r_Q in (5.9), Mobile & Calendar, Boating and Stefan sequences are tested. In the experiment, all the blocks in EL were coded in a non-intra mode due to their differential nature. It should be noted that in case of information loss, blocks in enhancement-layer can be intra- coded for better recovery from loss. The rate of all those coefficients in the EL exclusive of header and syntax, denoted by R_E^0 , is to be investigated under different quantization strategies. The BL is

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

Figure 5.3: Plot of the function R_E of $\frac{Q_B}{Q_E}$

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

encoded with a fixed Q_B equal to 20¹. One frame of those test sequences in the EL is encoded with different Q_E , or equivalently to say different r_Q . R_E^0 versus different r_Q is plotted in Fig 5.4, where R_E is computed by the actual bits used in the frame and it is measured by bit per pixel (bpp). Fig 5.4 shows that R_E has an approximate linear relationship with respect to $(r_Q - 1)$. And the correlation coefficient, given by (5.10), is used to estimate such a relationship, where $Cov(x, y)$ is the covariance and $Var(x)$ and $Var(y)$ is the variance. For different sequences shown in Fig 5.4, $CorCoef(R^E, r_Q)$ is 0.9998, 0.9981 and 0.9994 respectively, which implies the linear relationship between them. Then the aim of the EL rate control can be re-focused on looking for a suitable r_Q instead of traditional looking for a suitable Q_E .

$$CorCoef(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)} \cdot \sqrt{Var(y)}} \quad (5.10)$$

Figure 5.4: Relationship between bitrate (R_E^0) in terms of bit per pixel (bpp) and the ratio of quantization parameters (r_Q) in different test sequences, where the horizontal axis is r_Q and the vertical axis is R_E^0 .

¹mquant denotes the quantization scale for a MB. It is used in the TM5 rate control for MPEG-2 video coding and the final value of mquant is clipped to the range [1,31].

5.2.2 Relationship Between R_E and Non-zero Percentage Among Quantized Coefficients

The ρ -domain rate control algorithm for non-scalable coders has been proposed based on the approximate linear relationship between the ultimate bitrate and the percentage of zeros among quantized transform coefficients [35, 95]. In this work, the relationship between the rate and non-zeros percentage among the quantized coefficients in the EL (P_E^{nz}) is investigated to take the advantage of its little dependances on source statistics. Examples from different test sequences are shown in Fig 5.5, where the BL is encoded with a fixed Q_B equal to 23. It can be seen that R_E^0 monotonically increases with respect to P_E^{nz} . However, the slope of the curve is not constant in all instances. (5.11) is first used to model the relationship between R_E^0 and P_E^{nz} , where K_E^R is the slope and its characteristics are discussed in the following.

$$R_E^0 = K_E^R \cdot P_E^{nz} \quad (5.11)$$

On closer observation of Fig 5.5, it can be seen that the value of the slope K_E^R in the range of small P_E^{nz} differs from that in the range of large P_E^{nz} . In most EL coding scenarios, P_E^{nz} takes a small value at the frame level, e.g. 0.05 or so. And its value varies significantly at the macroblock level where P_E^{nz} is not constant either. This motivates us to investigate the slope K_E^R versus P_E^{nz} . In the study of non-scalable coding [35, 95], the variation of K_E^R is small and K_E^R is taken as a constant at the frame-level. However in EL coding, K_E^R varies much under different coding scenarios. Examples of K_E^R versus P_E^{nz} is shown in Fig 5.6, and correspondingly K_R versus R_E^0 is shown in Fig 5.7 when the BL is encoded with a fixed Q_B equal to 23. It can be seen that K_E^R is very much constant only in the range of larger P_{NZ} , which corresponds to a very high bitrate and beyond the bitrate used in a typical scenario of video coding. K_E^R also exhibits a steep slope for the range of smaller P_E^{nz} , which corresponds to the lower bitrate range. The variation of

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

5.5.a: Boating

5.5.b: Mobile & Calendar

5.5.c: Flower Garden

5.5.d: Stefan

Figure 5.5: R_E^0 (bpp) versus P_E^{nz} in different test sequences

K_E^R is an obstacle to developing a simple rate control strategy. From (5.11), K_E^R can be understood as a statistical parameter, and its value indicates the average number of bits used for every non-zero coefficients, because in a hybrid DCT & entropy coder, bits are mainly assigned for coding non-zero coefficients. When non-zero coefficients occur with a very small probability, it does not match the optimal case that entropy techniques are designed to handle, and K_E^R will show a steep increase in the range of lowest P_E^{nz} . P_E^{nz} usually has a small probability at the frame level, and at the same time, K_E^R and P_E^{nz} varies greatly in different MB. It is not easy to establish an accurate and simple model to describe the relationship between them. Therefore, it will be treated as different constant in different classes by a MB-classification strategy, which will be described in section Section 5.3.

5.2.3 Relationship between r_Q and P_E^{nz}

Another important observation is the strong correlation between r_Q and P_E^{nz} . With the same test sequences and encoding strategy in the previous sections, Fig 5.8 shows the relationship from collected data.

The correlation coefficients between r_Q and P_E^{nz} for those sampling data from different sequence are 0.988, 0.984, 0.970, and 0.979 respectively. Therefore, an approximate linear expression in (5.12) can be used to model the relationship, where K_E^Q is the slope and is a statistical parameter.

$$r_Q = 1 + K_E^Q \cdot P_E^{nz} \quad (5.12)$$

5.3 Optimum Bit Allocation for Enhancement-layer Coding

In this section, the distortion model for EL coding is investigated, and then a practical optimal-bit-allocation (OBA) scheme at the macroblock (MB) level is proposed.

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

5.6.a: Boating

5.6.b: Mobile & Calendar

5.6.c: Flower Garden

5.6.d: Stefan

Figure 5.6: K_E^R versus P_E^{nz} in different sequences

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

5.7.a: Boating

5.7.b: Mobile & Calendar

5.7.c: Flower Garden

5.7.d: Stefan

Figure 5.7: K_E^R versus R_E^0 (bpp) in different sequences

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

5.8.a: Boating

5.8.b: Mobile & Calendar

5.8.c: Flower Garden

5.8.d: Stefan

Figure 5.8: P_E^{nz} versus r_Q in different sequences

5.3.1 Theoretical Model for Optimum Bit Allocation

Without accurate distortion and rate model, OBA cannot be carried out [30]. Besides those relationship related to R-Q modeling in the previous section, an accurate distortion model is necessary to perform OBA for the EL coder.

The related ρ -domain distortion model (2.31) has been developed for non-scalable coding. That motivates us to investigate the relationship between distortion and P_E^{nz} and to see whether it is suitable for EL coding. Let $D_{E0} = \frac{D_E}{\sigma_E^2} \doteq \frac{D_E}{D_B}$ be the normalized distortion, where σ_E^2 is the variance of EL source and is also equal to the distortion of BL measured by MSE. The plots of D_0 versus P_{NZ} are shown in Fig 5.9 where the BL is encoded in a fixed scheme and the EL is encoded with different quantizer. It can be observed that the exponential function can still be used to model the distortion in the EL. Therefore, the EL distortion is modeled by (5.13), where α_E is a statistical parameter. Note that the model parameter α_E takes a larger value compared with its value in non-scalable coding and P_E^{nz} is usually small, e.g. about 0.05 at the frame level of EL coding.

$$D_E(P_E^{nz}) = \sigma_E^2 e^{-\alpha_E \cdot P_E^{nz}} \quad (5.13)$$

OBA assigns bits to each data source in order to minimize the overall distortion and achieve the best quality. This is usually solved by the method of Lagrangian optimization. With the rate in (5.11) and the distortion model (5.13) for the EL coding, the OBA problem can be solved [110]. Let $\{S_i | 1 \leq i \leq L\}$ be the input source and R_T be the target bit number. The problem can be formulated as the group of functions in (5.14), where N_i is the size of source S_i ; and the optimal solution for S_i is given by (5.15).

$$\begin{cases} D_{Ei} &= N_i \sigma_{Ei}^2 \exp(-\alpha_{Ei} \cdot P_{Ei}^{nz}) \\ R_{Ei} &= N_i K_{Ei}^R P_{Ei}^{nz} \\ F &= \min_{P_{Ei}^{nz}} \left(\sum_{i=1}^L N_i \sigma_{Ei}^2 \cdot \exp(-\alpha_{Ei} \cdot P_{Ei}^{nz}) + \lambda \cdot \left[\sum_{i=1}^L N_i K_{Ei}^R \cdot P_{Ei}^{nz} - R_T \right] \right) \end{cases} \quad (5.14)$$

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

5.9.a: Boating

5.9.b: Mobile & Calendar

5.9.c: Stefan

Figure 5.9: Normalized distortion curves of different frames. Normalized distortion D_{E0} versus the percentage of non-zero coefficients P_E^{nz}

$$R_{E_i} = \frac{\xi_i N_i (R_T - \sum_{j=1}^L \xi_j N_j \ln \frac{\sigma_{E_j}^2}{\xi_j})}{\sum_{j=1}^L \xi_j N_j} + \xi_i N_i \ln \frac{\sigma_{E_i}^2}{\xi_i} \quad (\xi_i = \frac{K_{E_i}^R}{\alpha_{E_i}}) \quad (5.15)$$

5.3.2 Challenges for Practical OBA

The above OBA scheme is a solution in mathematics. When it is applied to the EL coding, there are some constraints that should be taken into account.

5.3.2.1 Negative Number of Allocated Bits

First of all, one practical constraint cannot be neglected: $R_{E_i} \geq 0$. It is meaningless to assign a negative number of bits to one data source. However, those errors cannot be ignored if negative solutions of (5.15) are simply modified to zeros.

The value of variance of the EL MB (referring to DCT residues) depends on the complexity of the MB in raw video (referring to source DCT coefficient in the BL) and the adopted quantizer in the BL. Hence, variation of BL MB complexity and variation of the BL quantizer from MB to MB usually result in variance variation of EL MBs. The ratio between variance of two EL MBs may be large. As it is a factor in (5.15), quite a lot of MB will be assigned a negative number of bits. For example, the variance distribution and the corresponding assigned number of bits at the MB level from one frame of Mobile & Calendar sequence is shown in Fig 5.10, where both the BL and the EL are encoded at a target bitrate 3Mbit/s. Fig 5.10.a gives the variance of all MB in the EL, which is sorted by variance value. Fig 5.10.b is the assigned number of bits for all MB according to (5.15). It is easy to find that if the allocated negative number of bits are simply set to be zero, there will be a large mismatch between the actual bitrate and the target bitrate.

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

Figure 5.10: Example of the MB variance distribution and the corresponding allocated bit number with the MB-level OBA

5.3.2.2 Parameters K_E^R and α_E

If OBA is to be done according to (5.15), there is one uncertain key factor, namely $\xi_i = \frac{K_{Ei}^R}{\alpha_{Ei}}$. The characteristics of K_E^R have been discussed in Section 5.2.2 and they are also shown in Fig 5.6. Moreover, α_E varies with different bitrate, one example is shown in Fig 5.11 when a frame in Mobile & Calendar is encoded at different bitrate. Thus, both K_E^R and α_E are not constant in the EL coding and therefore difficult to be formulated. That cause the difficulty in optimal bit allocation and the solution of the EL quantization parameter. Even if any mathematical tool can be used to estimate them approximately, the high computational complexity and latency threat to the stability of the rate control algorithm will be introduced. In the following paragraph, the scheme to deal with such challenges are presented.

5.3.3 Practical OBA Schemes in MB-level

To overcome those challenges, some substitute methods are used to achieve the OBA in the EL coding.

5.3.3.1 Re-optimization scheme

To deal with the problem of negative allocated number of bits, the re-optimization scheme is employed, which takes the following two points into account. First, there is one distinct advantage of EL coding that MB can be skipped arbitrarily and frequently without dramatically decreasing the picture quality. That is because BL has provided the basic quality, and if any MB in the EL cannot be decoded correctly, the result is that visual quality of the corresponding part cannot be enhanced. Second, according to the principle of OBA such as (5.15), those MB with smaller variance are allocated less bits, which indicates their less importance.

In the proposed scheme, OBA is calculated according to (5.15) first. Those MB with negative allocated solutions will be skipped. That means those MB are assigned zero bit.

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

Figure 5.11: Plot of α_E versus coding bitrate using Mobile & Calendar

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

However, from Fig 5.10, it is seen that quite a few MB are allocated a negative number of bits. There will be a large mismatch between target bitrate and actual bitrate, if negative numbers are simply set to be zeros. To decrease this mismatch, a re-optimization scheme is employed by reassigning the target number of bits among those MB with positive allocated number of bits in the first optimization. In this way, the distribution of allocated number of bits will be smooth. In the following step, MB with negative solutions in the second optimization are also skipped. At the same time, those MB with smallest variance among those with positive allocated bit number are skipped to make up for the mismatch.

5.3.3.2 MB Classification

Both K_E^R and α_E vary much for different source or coding bitrate. To deal with the problem, all the MB in one frame are classified by the measure of variance ($\{\sigma_{Ei}^2 | 1 \leq i \leq n\}$) and the mean of variance ($\bar{\sigma}_E^2 = \frac{1}{L} \sum_{i=1}^L \sigma_{Ei}^2$). For lower computational complexity and implementation cost, all the MB are classified into 4 classes, and the classification method is shown as follows:

$$\left\{ \begin{array}{l} C1 : \{MB_i | \sigma_{Ei}^2 \leq \frac{\bar{\sigma}_E^2}{2}\} \\ C2 : \{MB_i | \frac{\bar{\sigma}_E^2}{2} < \sigma_{Ei}^2 \leq \bar{\sigma}_E^2\} \\ C3 : \{MB_i | \bar{\sigma}_E^2 < \sigma_{Ei}^2 \leq 2\bar{\sigma}_E^2\} \\ C4 : \{MB_i | \sigma_{Ei}^2 > 2\bar{\sigma}_E^2\} \end{array} \right. \quad (5.16)$$

It should be noted that the solution of OBA depends on the MB variance very much. Thus, MB of class $C1$ are usually allocated negative number of bits and skipped. After classification, each class can be looked on as an independent input source, in which all the MB can be assumed to have similar mathematical properties and coding behaviors. And MB belonging to the same class can be treated as having the same K_E^R and α_E in

that class level, which can be estimated by the following rewritten expressions of (5.11) and (5.13).

$$K_{Ei}^R = \frac{R_{Ei}^0}{P_{Ei}^{nz}} \quad (5.17)$$

$$\alpha_{Ei} = \frac{1}{P_{Ei}^{nz}} \cdot \ln \frac{\sigma_{Ei}^2}{D_{Ei}} \quad (5.18)$$

5.4 Rate Control

In this section, the rate control scheme is described, which can be used for a two-layered MPEG-2 SNR encoder.

5.4.1 Rate Model

As given by (5.19), the EL rate R_E is composed of bits for texture information (R_E^0) and header and syntax bits, denoted by C_E . Those bits for header and syntax information are usually considered as a relatively constant number.

$$R_E = R_E^0 + C_E \quad (5.19)$$

According to the analysis, rate model with respect to Q_E is derived. First, Combining (5.1) and (5.12), (5.20) can be obtained. Then the rate control model in (5.21) can be derived by substituting (5.20) into (5.11), where $Q_B > Q_E$.

$$P_E^{nz} = \frac{r_Q - 1}{K_E^Q} \quad (5.20)$$

$$R_E(Q_E) = K_E^R \cdot \frac{Q_B/Q_E - 1}{K_E^Q} + C_E \quad (5.21)$$

5.4.2 Frame-level Rate Control

At the frame-level level, a target number of bits is assigned for the current frame to encode. (5.22) is used for selecting the target number of bits for the frame, where R denotes the bitrate, F is the frame rate, and B_f in (5.23) is a buffer feedback factor [40], B is the current buffer level and B_s is the buffer size, set as $4 \cdot \frac{R}{F}$.

$$T = \max\left(\frac{R}{F} \cdot B_f, \frac{R}{4F}\right) \quad (5.22)$$

$$B_f = \frac{B + 2 \cdot (B_s - B)}{2B + (B_s - B)} \quad (5.23)$$

5.4.3 MB-level Rate Control

At the MB level, rate control selects the Q_E for each MB in a frame. The following is a step-by-step description of this method, and a flowchart is shown in Fig 5.12.

Step 1 : Initialization

Step1.1 Computation of parameters

First, compute MB variance, σ_{Ei}^2 ; second, classify MB into different classes, *Class j*, by (5.16); third, compute the parameters K_{Ej}^R , α_{Ej} and K_{Ej}^Q for each *Class j* according to the previous frame, and for the first frame, default values are chosen according to experience; fourth, let K_{Ei}^R , α_{Ei} and K_{Ei}^Q for different MB be equal to K_{Ej}^R , α_{Ej} and K_{Ej}^Q of the class they belong to. In practice, MB in *Class 1* and part of *Class 2* are usually skipped after OBA, and some parameters may not be possible to estimate and they are set to the default constant accordingly.

Step 1.2 Optimum bit allocation

With σ_{Ei}^2 , K_{Ei}^R and α_{Ei} , OBA is performed according to (5.15) and the proposed re-optimization scheme in Section 5.3.3. The target number of bits for each MB can be computed.

Figure 5.12: Flowchart of rate control at MB-level

Step 2 Compute Q_E for i th Macroblock

First, check the allocated number of bit R_{Ei}^0 for i th MB, if $R_{Ei}^0 < 0$, then skip the i th MB, else compute P_{Ei}^{nz} by (5.24).

$$P_{Ei}^{nz} = \frac{R_{Ei}^0}{K_{Ei}^{er}} \quad (5.24)$$

Second, compute r_Q by (5.12), so the desired Q_{Ei} can obtained and then encode the MB.

Step 3 Update Model parameters for each class

After encoding the i th MB belonging to *Class j*, compute the K_{Ej}^R by (5.17) and K_{Ej}^Q by (5.12) for *Class j* according to the previous encoded MB belonging to *Class j* ; then update the K_{Ej}^R and K_{Ej}^Q .

5.5 Experimental Results

The proposed rate control algorithm is applied on an SNR scalable encoder, which is based on the standard MPEG-2 coder [111]. In the experiments, the BL is encoded either in VBR mode or in CBR mode by the TM5 rate control, and at the same time the EL is encoded at a given bitrate. The EL is controlled by the proposed rate control algorithm and the TM5 rate control respectively.

The raw sequences are as follows: (a) Boating (720×576 4 : 2 : 0), 100 frames, (b) Mobile & Calendar (720×576 4 : 2 : 0), 100 frames, shown in Appendix D. In the experiments, the length of GOP is 15, and the distance between I frame and P frame is 3.

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

Table 5.1: Performance comparison when the BL is encoded in VBR model

Video sequences	BL bitrate bit/s(VBR)	BL PSNR	EL bitrate bit/s	BL&EL average PSNR		
				TM5	Proposed	Gain
Boating	3.9M (QP =23)	29.974	2M	30.968	31.488	0.520
			3M	31.474	32.332	0.858
			4M	31.968	33.095	1.127
Mobile & Calendar	4.3M (QP =23)	27.097	2M	27.755	27.979	0.224
			3M	28.070	28.485	0.415
			4M	28.427	29.022	0.595

Table 5.2: Performance comparison when the BL is encoded in CBR model

Video sequences	BL bitrate bit/s(CBR)	BL PSNR	EL bitrate bit/s	BL&EL average PSNR		
				TM5	Proposed	Gain
Boating	3M	27.975	2M	29.502	30.339	0.837
			3M	30.130	31.371	1.241
			4M	30.654	32.265	1.611
Mobile & Calendar	3M	28.314	2M	28.879	29.733	0.854
			3M	29.224	30.459	1.235
			4M	29.622	31.013	1.391

Table 5.1 & 5.2 show the PSNR performance comparison between the proposed algorithm and TM5 rate control. In Table 5.1, the EL is encoded at different target bitrates while the BL is encoded in the VBR mode. Each frame of the BL is encoded at a fixed QP equal to 23. Table 5.2 shows the case that the EL is encoded at different target bitrate while the BL is encoded in the CBR mode. The BL is controlled by TM5 rate control. They both show that the proposed work can achieve a higher average PSNR gain compared with the TM5 rate control. Examples of PSNR comparison for each frame are plotted in Fig 5.13 and Fig 5.14.

The comparison of distortion in the MB-level is shown in Fig 5.15, where the distortion is measured by MSE. One frame in the Mobile & Calendar is taken. After BL coding, the EL is encoded at a given rate budget controlled by the proposed rate control and the TM5 rate control respectively. In the Fig 5.15, the distortion of all MB are plotted, which are from the BL picture, the reconstructed picture with the EL controlled by TM5

rate control and the reconstructed picture with the EL controlled by TM5 rate control. All the MB are sorted by distortion in an ascending order. It is shown that the maximal distortion of sorted reconstructed MB by the proposed algorithm is much smaller than that by the TM5. It implies that the proposed algorithm can achieve smoother picture quality than TM5.

5.6 Summary

In this chapter, rate control for high bitrate SNR scalable video coder is investigated. Some limitations of those rate control algorithms for non-scalable coding are addressed when they are applied to the SNR scalable coding. Based on the characteristics of EL compression, a novel rate control is proposed for the EL coding. The proposed rate control are based on the following points. First, according to quantization theory and rate distortion theory, the linear relationship between the bitrate and r_Q is found. Second, the linear rate model (in ρ -domain rate control), proposed for non-scalable coding is investigated in EL coding. It is found that the slope of rate model and model parameters of distortion models vary significantly when compared with non-scalable coding conditions. The proposed scheme has dealt with such a problem. Third, because rate control parameters are not stable for robust video compression, a MB classification method is proposed to perform EL optimal-bit-allocation. Fourth, a re-optimization scheme for EL optimal-bit-allocation is proposed to decrease the rate control error. In addition, other related useful relationships are integrated to achieve the EL rate control framework. Experiments results suggest that the rate control algorithm performs well and shows an improvement when it is compared with the classical TM5 rate control.

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

5.13.a: Boating

5.13.b: Mobile & Calendar

Figure 5.13: PSNR performance comparison between TM5 and the proposed algorithm for EL coding, when the BL is encoded in the VBR mode and the EL is encoded at 3Mbit/s.

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

5.14.a: Boating

5.14.b: Mobile & Calendar

Figure 5.14: PSNR performance comparison between TM5 and the proposed algorithm for EL coding when the BL and the EL are encoded at 3Mbit/s respectively.

CHAPTER 5. RATE CONTROL FOR CONVENTIONAL SNR SCALABLE CODING

Figure 5.15: Comparison of the distortion MSE of the reconstructed frame in MB-level, where distortion is measured by MSE

Chapter 6

R-D Analysis and Optimal Bit Allocation for FGS Coding

The fine-granularity-scalability (FGS) video coding technique [22], proposed as a new flexible and simple framework for scalable coding, has been incorporated into MPEG-4 as an amendment for the streaming video profile. The major difference between the FGS and conventional SNR scalability is the adoption of the bitplane (BP) coding technique for DCT coefficients instead of the conventional quantization of DCT coefficients. For that reason, the FGS can provide continuous video quality. When the FGS video bitstream is sent over networks, if all the BP information of the FGS-layer is transmitted from the server to the client, which is unnecessary, the bitrate will reach quite a high level and will exhaust network resources greatly. As a matter of fact, video bitstreams are often transmitted in a bandwidth-constrained environment. Therefore, truncating the FGS bitstream at a given bitrate budget is the main concern in the study of rate adaptation for FGS-based video streaming applications. The truncation of the FGS bitstream is essentially bit allocation in the FGS-layer. In this chapter, the issue of bit allocation for FGS coding is investigated. The problem of bit allocation for FGS coding is first addressed by illustrating the drawbacks of existing bit allocation schemes. Subsequently, an optimum bit allocation scheme is proposed based on the analysis of R-D with respect to binary coefficients in different BP.

6.1 Problems Description of Bit Allocation for FGS Coding

Most of existing bit allocation schemes [51, 52, 112] for FGS coding follow the solution of optimal bit allocation in the classical R-D theory. The optimal bit allocation in R-D theory is presented first, and then some limitations of existing bit allocation schemes are highlighted.

6.1.1 Optimal Bit Allocation in R-D Theory

The objective of optimal bit allocation is to achieve the best video quality for the given bitrate constraint. In other words, optimal bit allocation is to minimize the distortion, subject to a constraint R_C on the bitrate R , which can be expressed by (6.1). The optimal solution of (6.1) can be obtained by some optimization techniques combined with known R-D models [88]. For example, the Lagrangian multiplier method [113], which is one of the most popular optimization techniques, is given in (6.2).

$$\min\{D\}, \text{ subject to } R < R_C \quad (6.1)$$

$$J = D + \lambda \cdot (R - R_C) \quad (6.2)$$

The classical solution of optimal bit allocation in R-D theory is briefly reviewed as follows. Suppose there is a coding system with input sources $\{S_i | 1 \leq i \leq N\}$. D_i and R_i denote the distortion and the rate respectively which are obtained after encoding the source S_i . The distortion (D_i) is typically measured by mean-squared-error (MSE). In transform coding, the rate-distortion function for the source S_i can be explicitly written as (6.3) [16], where R_i denotes the average bits per pixel, σ_i^2 is the signal variance of source S_i , and ϵ_i^2 is a constant dependant on the probability distribution of the input source (S_i).

$$D_i = \epsilon_i^2 \sigma_i^2 2^{-2R_i} \quad (6.3)$$

Using the R-D model (6.3), the problem of optimal bit allocation can be formulated by the Lagrangian multiplier in (6.4), and its optimal solution is given in (6.5) where R_T is the target rate. The optimal solution (6.5) has provided the principle for bit allocation among multiple sources. However, it requires the knowledge of both the variance (σ_i^2) and the statistics of the source (ϵ_i^2).

$$J(\lambda) = \sum_{i=1}^N D_i(R_i) + \lambda \cdot \left(\sum_{i=1}^N R_i - R_T \right) \quad (6.4)$$

$$R_i = \frac{R_T}{N} + \frac{1}{2} \log_2 \frac{\epsilon_i^2 \sigma_i^2}{\left(\prod_{j=1}^N \epsilon_j^2 \sigma_j^2 \right)^{\frac{1}{N}}} \quad (6.5)$$

6.1.2 Drawbacks of Existing Bit Allocation Schemes for FGS Coding

The uniform bit allocation scheme is the simplest way among all bit allocation schemes for FGS coding. It allocates an average number of bitrate to each input source, which can be expressed by (6.6) where R_T is the target bitrate, N is the number of sources. Such an approach has a low coding efficiency, as it does not adopt any optimization technique.

$$R_i = \frac{R_T}{N} \quad (6.6)$$

More efficient bit allocation schemes [51, 52, 112] have been proposed based on the R-D optimization techniques. Optimal bit allocation should be performed on the basis of accurate R-D models. According to the method of R-D modeling for FGS coding, the optimal bit allocation schemes can be categorized into two basic approaches. One approach extracts several R-D points empirically according to the collected R-D data

[52], and then it performs bit allocation based on the extracted R-D curve. The empirical approach can not provide deep insights into coding behaviors of FGS coding.

The other approach [51, 52] uses the classical rate-distortion function (6.3) as the R-D model for the FGS-layer coder, and its bit allocation scheme employs the corresponding optimal solution in (6.5). However, there are some drawbacks when the classical optimal-bit-allocation solution is applied to the practical FGS coding. In the optimal solution (6.5), ϵ_i^2 is necessary to calculate the bit allocation result ($\{R_i | 1 \leq i \leq N\}$). However, it is difficult to obtain (or formulate) the value of ϵ_i^2 for the FGS-layer source. The value of ϵ_i^2 depends on the statistics of the FGS-layer source, and for a practical video coding system, it can also be understood to include an additional factor on the efficiency of entropy coding [28]. The probability distribution of the FGS-layer source depends on not only original video pictures but also quantization stepsize of the base-layer according to the analysis for the base-layer DCT residues in Chapter 3. Therefore, the variation of either the video contents or the base-layer quantization stepsize will result in the different value of ϵ_i^2 . However, existing bit allocation schemes for FGS coding simply assume that all ϵ_i^2 are the same and then use the simplified solution in (6.7). Such a simplification indicates that all FGS-layer sources have the same distribution. It can be found that this is not reasonable enough, especially when the base-layer is encoded in a low bitrate or the scenes in raw sequences change a lot. For instance, the values of ϵ_i^2 in the FGS-layer are computed according to (6.3) and they are plotted in Fig 6.1, where the base-layer is coded at 100kbit/s with the VM18 rate control and the FGS-layer is coded at 100kbit/s with uniform bit allocation among frames. The figure shows that it is not accurate to assume all ϵ_i^2 are approximately constant.

$$R_i = \frac{R_T}{N} + \frac{1}{2} \log_2 \frac{\sigma_i^2}{(\prod_{j=1}^N \sigma_j^2)^{\frac{1}{N}}} \quad (6.7)$$

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

6.1.a: Carphone

6.1.b: Foreman

Figure 6.1: ϵ_i^2 of different frames in the FGS-layer

Another drawback of existing optimal bit allocation schemes is that a large control error is probably introduced. There is one constraint ($R_i \geq 0$) in practical coding. Even though the variation of ϵ_i^2 (which is just analyzed) is ignored and bit allocation can be performed according to the simplified (6.7), it probably occurs that $R_i < 0$ for a certain percentage of the total frames. It can be seen from (6.7) that the first item is related to R_T and the second item is related to σ_i^2 . R_T and σ_i^2 are independent, so there will be many negative R_i in the optimal solution due to either a small R_T or widely varying σ_i^2 . Consequently, a large rate control error will be generated if R_i with negative values are set to be zero as mentioned in [52]. For example, variance of FGS-layer frames and the corresponding bit allocation result calculated by (6.7) are plotted in Fig 6.2, where the base-layer is encoded at 100 kbit/s and the target bitrate for the FGS-layer is also 100kbit/s. Fig 6.2 shows that the existing optimal-bit-allocation solution is not robust for FGS coding. In addition, the bit allocation result will be worse if the target bitrate is further reduced or σ_i^2 varies in a wider range.

Besides the existing bit allocation schemes which are based on the classical optimal-bit-allocation solution, there are some other related researches in FGS coding. The addition of two Laplacian probability distributions was proposed in [49] to model the statistics of FGS sources. However, it is difficult to develop a rate model based on its statistical model and then perform bit allocation. Another work in [114] suggested that the truncation of FGS bitstream in the bitplane order is a better choice than that in the traditional raster-scan order. However, it did not give any R-D model and did not provide a practical bit allocation framework among multiple frames. To overcome the drawbacks of existing bit allocation schemes for FGS coding, a robust and efficient bit allocation framework is proposed in the following.

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

6.2.a: frame variance (σ_i^2) in Carphone

6.2.b: R_i (bpp) for different frames in Carphone

6.2.c: frame variance (σ_i^2) in Foreman

6.2.d: R_i (bpp) for different frames in Foreman

Figure 6.2: Allocated number of bits for different frames (R_i) by existing bit allocation schemes

Table 6.1: Definitions of notations for the FGS analysis

NOTATION	DEFINITION
i	frame index
j	bitplane level index, corresponding to the stepsize of 2^j
k	position index for both DCT and binary coefficients
x	magnitude of DCT coefficient
b	binary coefficient
N	the size of the data source
J	the set of bitplane levels to be truncated (to be decoded at the client side)
$P1$	the percentage of non-zero binary coefficient

6.2 Proposed R-D Analysis for FGS Coding

The R-D performance of a video coder depends totally on the source statistics and the coder design. As the source statistics in the FGS-layer are complex and the FGS coder adopts the BP coding of DCT coefficients, the R-D performance of FGS-coder is different from that of general video coders. Accurate R-D analysis for FGS coding should take these characteristics of FGS coding into account. The FGS coder can be understood to encode binary representations of DCT coefficients instead of the DCT coefficients themselves. Therefore, it should be useful to study the effect of binary coefficients on the final R-D performance. For convenience, the definitions of notations to be used in the following part of this chapter are listed in Table.6.1.

6.2.1 Linear Rate Model for FGS Coding

As introduced in Section 2.1.2.2, the FGS coder decomposes DCT residues of the base-layer into the binary representation first. The binary decomposition can be expressed by (6.8), where x_k denotes the magnitude of the k th DCT coefficient and $b_k(j)$ is the binary coefficient. The j th BP is composed of all the binary coefficients ($\{b_k(j) \mid 1 \leq k \leq N\}$) corresponding to the stepsize 2^j . If value of the BP level (j) is larger, the BP j is more

significant. The FGS coder encodes binary coefficients from the most significant bitplane (MSB plane) to the least significant bitplane (LSB plane). Blocks in each BP, every of which consists of 8×8 binary coefficients, are encoded in the conventional way by entropy coding techniques such as run length coding and variable length coding. The entropy-coding tables in the standards are designed by the utilization of the statistical properties of non-zero binary coefficients (NZBC). Therefore, the relationship between the rate and the NZBC is to be investigated.

$$x_k = b_k(0) \cdot 2^0 + b_k(1) \cdot 2^1 + \cdots + b_k(j) \cdot 2^j + \cdots \quad (6.8)$$

The NZBC distribution across different BP is shown in Fig 6.3, where the horizontal axis is the BP level (j) and the vertical axis is the percentage of non-zero binary coefficients ($P1$). The corresponding rate of different BP is shown in Fig 6.4. It can be seen that the percentage of NZBC in the BP j ($P1(j)$) decreases gradually when the BP level (j) increases, and so does the bitrate of the BP j ($R(j)$). This indicates there is a correlation between $P1(j)$ and $R(j)$. In addition, it is observed that the value of $P1(j)$ in Fig 6.3 and the rate ($R(j)$) in Fig 6.4 is quite different from BP to BP. The entropy-coding tables in the FGS standard are designed for different BP [22]. For example, the FGS standard provides four VLC tables for MSB plane, MSB-1 plane, MSB-2 plane and other BP respectively. Therefore, the relationship between the rate and the percentage of the non-zero binary coefficient ($P1$) is studied at the BP level.

The actual relationship between the percentage of NZBC ($P1(j)$) and the rate ($R(j)$) in different BP is shown in Fig 6.5. It is likely that the whole BP is not truncated entirely in the FGS server due to the limited bandwidth. In such cases, the $P1(j)$ stands for the percentage of the truncated NZBC in one BP, and accordingly the $R(j)$ stands for the bitrate of the BP j which is not truncated entirely. The correlation coefficients of the

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

Figure 6.3: Percentage of non-zero binary coefficients ($P1$) in different BP

Figure 6.4: Rate (bpp) of different BP

four set of data in Fig 6.5 are 1.000, 0.999, 0.999 and 0.998 respectively. It suggests that there is a linear relationship between the rate and the percentage of NZBC in individual BP, so the linear rate model of $P1$ is given as follows. The percentage of NZBC in the BP j ($P1(j)$) can be calculated by (6.9). The rate model for one BP is given by (6.10), and the rate model for the whole frame is given by (6.11), where $\theta(j)$ is a statistical slope in the BP j , N_{bp} is the size of the BP, and the capital letter J denotes the set of all the truncated BP j . It should be noted that if only a portion of binary coefficients in the BP j are truncated, $P1(j)$ denotes the percentage of the truncated NZBC in the BP j instead of the percentage of all the NZBC in the BP j .

$$P1(j) = \frac{\sum_{k=1}^{N_{BP}} b_k(j)}{N_{bp}} \quad (6.9)$$

$$R(j) = \theta(j) \cdot P1(j) \cdot N_{bp} \quad (6.10)$$

$$R = \sum_{j \in J} \theta(j) \cdot P1(j) \cdot N_{bp} \quad (6.11)$$

6.2.2 Distortion Analysis for FGS Coding

The rate distribution across different BP and the linear rate model have been presented. According to the linear relationship between the rate and the percentage of NZBC, the aim of R-D analysis for FGS coding can be re-focused on minimizing the distortion at a given number of NZBC rather than minimizing the distortion at a given bitrate constraint. In the following, the optimal strategy of truncating NZBC is studied by minimizing the overall distortion at a given number of NZBC.

The distortion is measured by MSE as usual. The overall distortion can be expressed by binary coefficients in all the BP. The magnitude of source DCT coefficients in the FGS-layer is denoted by $X = \{x_k | 1 \leq k \leq N\}$, and the binary coefficients decomposed from the source DCT coefficient (x_k) are denoted by $\{b_k(j) | j = 0, 1, 2, \dots\}$. Suppose all

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

Figure 6.5: The rate R (bit) versus the percentage of NZBC ($P1$) in different BP from one frame in Foreman

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

the binary coefficients ($b_k(j)$), whose j belongs to J_k , are truncated and decoded at the client side. In other words, $b_k(J_k)$ is defined as the whole set of the $b_k(j)$ to be truncated. For example, if $\{b_k(0), b_k(4), b_k(5)\}$ are truncated and sent to the client side, J_k is the set $\{0, 4, 5\}$. The distortion can then be expressed by (6.12). It can be found that the distortion changes only when $b_k(j) = 1$, which corresponds to one NZBC. It indicates that the strategy of truncating NZBC is the key to minimize the overall distortion.

$$D = \sum_{k=1}^N (x_k - \sum_{j \in J_k} b_k(j) \cdot 2^j)^2 \quad (6.12)$$

Based on (6.12), the following two conclusions can be drawn on the optimal strategy of selecting NZBC to truncate (proof: see Appendix C). First, the larger the BP level (j) where the truncated NZBC is located, the smaller the distortion is. In other words, the NZBC in the larger BP should be truncated first. In the case of no frame-delay video coding such as real-time video streaming, a target number of bits on the average are allocated at the frame-level, and the truncation strategy should be from most-significant-bitplane to least-significant-bitplane till the allocated bits are used up. In another case of bit allocation among multiple frames such as long-delay video streaming, the larger BP among all frames should be given a higher priority during truncation. The second conclusion is that the NZBC decomposed from x_k with a larger residue value of x_k should have a higher priority to be truncated. What this implies is that in the same BP j , the NZBC decomposed from the DCT coefficient with a larger residue should be truncated first, if the target number of bits are not sufficient for the whole BP.

The above R-D analysis is based on the relationship between R-D and $P1$ (or to say equivalently, the distribution of NZBC). The advantage of the proposed method is little dependence on the probability distribution of the source in the FGS-layer. The rate model is accurate because it is extracted from the actual coding process. And the distortion

model in (6.12) is the actual distortion, which is more accurate than the traditional one estimated by rate-distortion function in theory.

6.3 Bit Allocation Scheme

In this section, our bit allocation scheme for FGS coding is addressed on the basis of the proposed rate model and distortion analysis.

6.3.1 Estimation of θ in the Rate Model

The linear rate model is given in (6.10). The estimation of the parameter θ is important to the accuracy of the rate model and therefore the bit allocation. Actual values of $\theta(j)$ in different BP are shown in Fig 6.6, where experimental data are from 300 frames in the Foreman sequence.

It can be observed that the smaller the BP level is, where $P1$ is larger and varies little from frame to frame, more constant θ is. In the smaller BP level such as BP 0, BP 1, BP 2 and BP 3, the value of θ is quite constant. However, its value varies much in the larger BP such as BP 4 and BP 5. The variation of $\theta(j)$ in the larger BP will not affect much the result of the overall bit allocation, as only a small portion of bits are spent on the large BP level, which can be seen from Fig 6.4.

The value of θ should be obtained prior to truncation and only can be estimated according to the previous experimental data. The θ can be understood as the average number of bits spent on each NZBC statistically. Thus its value should be related to the distribution of NZBC. It can be observed from Fig 6.6 that the value of θ becomes small when $P1$ increases. θ is therefore approximately modeled by the monotonic function of $P1$ in (6.13), where A and B are the coefficients. Suppose that m couples of experimental data are collected; according to the linear regression theorem [115], the optimum estimation of θ is given in (6.16).

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

Figure 6.6: θ versus $P1$ in different BP from 300 frames of Foreman

$$\theta = F(P1) = A + B \cdot P1 \quad (6.13)$$

$$\hat{B} = \frac{m \sum_{n=1}^m P1_n \theta_n - (\sum_{n=1}^m P1_n)(\sum_{n=1}^m \theta_n)}{m \sum_{n=1}^m x_n^2 - (\sum_{n=1}^m x_n)^2} \quad (6.14)$$

$$\hat{A} = \frac{1}{m} \sum_{n=1}^m \theta_n - \frac{\hat{B}}{m} \sum_{n=1}^m P1_n \quad (6.15)$$

$$\hat{\theta} = \hat{A} + \hat{B} \cdot P1 \quad (6.16)$$

6.3.2 Bit Allocation Algorithm

Given a target bitrate R_T , our bit allocation algorithm is presented as follows. Set a group of frames (GoF) with the length F ($F \geq 1$). The group of frames is the bit allocation unit. Given the frame rate (f_s), the target number of bits for the unit (T_{GoF}) can be

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

calculated by (6.17). The bit allocation task is to select NZBC at the rate constraint T_{GoF} , as described by (6.18) where i is the frame index and j is the BP index.

$$T_{GOP} = R_T \cdot \frac{F}{f_s} \quad (6.17)$$

$$R_{GoF} = \sum_{i=1}^F \sum_{j \in J_i} P1(i, j) \cdot \theta(i, j) \cdot N_{bp}, \quad (6.18)$$

subject to $R_{GoF} \leq T_{GoF}$

According to the conclusion in Section 6.2.2 on the optimal selection of NZBC, the truncation of binary coefficients in the group of frames should start from the largest BP level to the smallest BP until the allocated number of bits are used up. In such a strategy, the last BP level (L) to be truncated in the group of frames can be obtained by the solution of the inequality in (6.19), where H denotes the largest BP level in the group of frames. The allocated number of bits for frame i ($R(i)$) can be computed by (6.20), where H_i is the largest BP level of frame i , and $R(i, L)$ denotes the allocated number of bits for the BP level L of frame i . In the following, the method to compute $R(i, L)$ is presented separately, as it is possible that there is insufficient number of bits available to encode the whole smallest BP.

$$R_{GoF}(L) = \sum_{i=1}^F \sum_{j=L}^H P1(i, j) \cdot \theta(i, j) \cdot N_{bp} \quad (6.19)$$

$$\text{subject to } R_{GoF}(L+1) < T_{GoF} \leq R_{GoF}(L)$$

$$R_i = \sum_{j=L+1}^{H_i} P1(i, j) \cdot \theta(i, j) \cdot N + R(i, L) \quad (6.20)$$

Once the smallest BP level L is obtained according to (6.19), the remaining number of bits for the BP level L ($R_{GoP}(L)$) can be calculated by (6.21). Then the last step of bit

allocation is to allocate $R_{GoP}(L)$ to different frames if the whole BP L cannot be truncated due to insufficient bits left. This is a problem of selecting NZBC in the same BP level essentially. According to the second conclusion in Section 6.2.2, the NZBC decomposed from the DCT coefficient with a larger residue will be given a higher priority to truncate. However, this optimal selection requires skipping some blocks where the NZBC with smaller residue is located, so it will introduce some extra overhead to indicate the block-skipping. For that reason, optimal selection in the same BP level will not increase the picture quality much when compared with truncation in a common order. Therefore, the last BP in each frame is truncated in a common order, and the target bits for the BP L in individual frames are allocated according to the number of NZBC in the last truncated BP, as given by (6.22). Once $R(i, L)$ is obtained, the number of bits assigned to individual frames can be calculated according to (6.20).

$$R_{GoF}(L) = T_{GOP} - R_{GoF}(L + 1) \quad (6.21)$$

$$R(i, L) = R_{GoP}(L) \cdot \frac{P1(i, L)}{\sum_{f=1}^F P(f, L)} \quad (6.22)$$

6.4 Experimental Results

An MPEG-4-based FGS coder is used. Carphone, Foreman, News and Stefan sequences (qcif resolution, 4:2:0) are tested. The proposed bit allocation scheme is applied on the video coder to allocate a target number of bitrate to all the frames in the sequence. It is compared with the uniform bit allocation. The uniform bit allocation scheme allocates an average number of bits to every frame, and it truncates the FGS-layer from the most significant bitplane to the less significant bitplane in individual frames. Table.6.2 gives the PSNR comparison between the proposed bit allocation scheme and the uniform bit allocation scheme. In the experiments, the encoding frame rate is 10 frame/s for both the

Table 6.2: PSNR comparison between the uniform bit allocation (uniform) and the proposed bit allocation (proposed)

Video sequences	base-layer		FGS-layer			
	bitrate (bit/s)	PSNR (db)	bitrate (bit/s)	average PSNR (db)		
				uniform	optimal	gain
Carphone	48k	32.791	100k	35.549	35.797	0.248
Foreman	64k	32.334	100k	34.869	35.245	0.376
News	48k	32.234	100k	35.513	35.707	0.194
Stefan	150k	29.163	150k	32.083	32.246	0.163

base-layer and the FGS-layer. The base-layer is encoded at a constant bitrate controlled by the VM18 rate control, and the FGS-layer is truncated by the proposed bit allocation scheme and the uniform bit allocation scheme respectively. It is seen from Table.6.2 that the proposed bit allocation scheme can achieve a higher average PSNR. Furthermore, the PSNR comparison for individual frames is shown in Fig 6.7. From the figure, it can be observed that the picture quality by the proposed bit allocation scheme appears to be smooth. For frames whose base-layer is with very low PSNR comparatively, a gain of up to 3dB can be achieved compared with uniform bit allocation. In addition, it is worth noting that one important advantage of the proposed approach over traditional ones is nearly no control error. Such a point has been illustrated in Fig 6.2.

6.5 Summary

This chapter investigated the problem of bit allocation for FGS coding. The drawbacks of traditional bit allocation schemes are first illustrated. Traditional bit allocation schemes are not accurate because they did not take into account neither the variation of the FGS-source statistics nor the characteristics of bitplane coding. To overcome these drawbacks, the relationship between R-D and NZBC is studied. A linear rate model, based on the linear relationship between the bitrate of the truncated BP and the number of truncated NZBC, is proposed. The advantage of the rate model is minimally source-dependent.

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

Figure 6.7: PSNR Comparison between the proposed work (red solid line) and uniform bit allocation (black dotted line)

Subsequently, the relationship between the distortion and NZBC is analyzed. From the theoretical derivation, an optimal strategy of truncating the FGS bitstream is presented to minimize the overall distortion. A bit allocation scheme is then proposed based on the proposed rate model and distortion analysis. When it is applied to the MPEG-4 FGS coder, experimental results suggest that the proposed bit allocation scheme can achieve smoother and higher picture quality than the uniform bit allocation. In addition, the

CHAPTER 6. R-D ANALYSIS AND OPTIMAL BIT ALLOCATION FOR FGS CODING

proposed bit allocation scheme is much robust, as it will not allocate a negative number of bits to any frame and therefore will not cause the potential control error.

Chapter 7

Conclusions and Future Work

This chapter concludes this dissertation. Our main contributions to the field of R-D analysis for DCT-based video coding are presented. In the final part of this chapter, some future work is discussed.

7.1 Conclusion

The objective of this dissertation has been set to develop an accurate R-D modeling framework and more efficient control algorithms for DCT-based video coding including non-scalable video coding and SNR scalable video coding. To accomplish the objective, the task has been divided into the following four parts.

- modeling the distribution of DCT residues
- R-D analysis and control for non-scalable video coding
- R-D analysis and control for QP-based SNR scalable video coding
- R-D analysis and bit allocation for FGS coding (bitplane-coding based SNR scalable video coding)

An analytical model is first proposed to describe the distribution of DCT residues. It is derived based on well-known statistical models for video source and the analysis of the

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

quantization process in video compression. Extensive experimental results suggest that the proposed distribution model is much closer to the actual distribution of DCT residues than commonly-used statistical models such as uniform, Gaussian and Laplacian distribution models. Furthermore, the distribution range of DCT residues is definitely limited by the adopted quantization stepsize and it is bounded. In terms of the distribution range, the proposed model is more mathematically correct than other existing models such as the Gaussian model and the addition of two Laplacian model whose distribution range is unbounded. Besides the accuracy, it has the advantage over traditional models that it can quantitatively model the distribution of DCT residues with respect to the source variance and with respect to the quantization stepsize. This enables the proposed model to accurately predict the distribution of DCT residues prior to actual video coding.

Based on the proposed distribution model of DCT residues, a distortion model is developed. The proposed distortion model, which is a function of source variance and the quantization stepsize, can adapt well to the variation of both the video source and the quantization strategy. It can estimate the distortion accurately in the full range of the quantization parameter. Moreover, a quality control algorithm is developed based on the proposed distortion model. It is applied to encode the raw video at a defined fidelity level. Experimental results suggest that the proposed distortion model outperforms the uniform distortion model and adapts well to the variation of video contents. Combining the proposed distortion model with the classical rate-distortion function, the rate model is thus obtained. Subsequently, a rate control algorithm is developed based on the rate model. It is then applied on an MPEG-4 video coder to control video coding at a constant bitrate. Compared with the standardized rate control algorithm [40], the proposed rate control algorithm can achieve higher coding efficiency.

Besides the non-scalable coding, R-D analysis and robust control for SNR scalable coding has also been investigated. To cope with much-varying statistics of the

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

enhancement-layer source and take the characteristics of enhancement-layer coding into account, more source-independent R-D optimization algorithms have been proposed. According to the technique of dealing with DCT coefficients in the enhancement-layer, SNR scalable coders are categorized into conventional QP-based coders and bitplane-coding-based FGS coders. In the study of the conventional QP-based SNR coder, an approximate linear relationship between the bitrate and the ratio of quantization stepsize of different layers is first formulated. Then a rate model, between the bitrate and the ratio of the quantization parameter in the base-layer to the enhancement-layer, is developed.

In the study of FGS coding, a robust and efficient bit allocation scheme is proposed to truncate the FGS bitstream. We study the relationship between R-D and the non-zero binary coefficients (NZBC) in different bitplanes. A linear rate model, which is based on the linear relationship and the number of truncated NZBC, is developed for FGS-coding. According to mathematical analysis of the relationship between the distortion and the NZBC, the optimal strategy of truncating NZBC is given to minimize the overall distortion. Based on the proposed R-D analysis with respect to NZBC, an optimal bit allocation algorithm is developed. Experimental results show the proposed bit allocation algorithm can achieve smooth and high video quality. And another advantage of the proposed bit allocation algorithm over traditional ones is no control error.

7.2 Future Work

Some ideas about future extensions of our work are discussed in the following.

- **Video quality control at a given bitrate**

In Section 4.2, a quality control algorithm has been presented based on our distortion model. It is designed to control video coding at a defined level of visual fidelity. Experimental results have proved that the proposed distortion model outperforms the widely-used uniform distortion model. In the design of the quality

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

control algorithm, the bitrate constraint has not been taken into account. However, some applications consider the visual quality in a highlighted place, but the coding bitrate constraint is still an important factor that cannot be ignored in the control of video coding. For example, if a video sequence is to be encoded and stored in the disk, some scenes of interest may be highlighted, while the total size of encoded video requires to be maintained remains the same. Therefore, it is worth extending our quality control algorithm to control, smooth and selectively highlight the video quality at a given bitrate constraint.

- **Rate-control for video streaming over Internet**

The rate control for DCT-based video coding, including single-layer coding and two-layer coding, has been investigated at a constant-bitrate mode. The proposed rate control algorithm mainly concentrates on source coding at the sender side, so it employs a simple network model that the available bandwidth is constant. As a matter of fact, a practical system for video transmission over Internet is complicated. Transmitted video has to suffer from the impairments caused by fluctuations of channel conditions, transmission delay and packet loss [72]. Ignoring each of these factors may result in the unsatisfactory video quality at the receiver side. Buffer techniques can reduce the impairments of the fluctuations of channel conditions to some extent. The buffer size depends on the available bandwidth and potential transmission delay. The internet protocol can also provide some useful information for adaptive control. For example, the realtime control protocol [116] can give some feedback information about the channel condition, and then the sender can adjust the sending rate to avoid network congestion [117]. Thus, the source coding bitrate, which is the rate constraint for maximizing compressed video quality, should be dynamically adjusted according to these buffer constraints and network

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

feedback information. In the case of live video streaming which re-transmission is not applicable for, the robustness of coded video needs to be improved to cope with the packet loss. Error resilient techniques can much improve the received video quality in presence of information loss [70]. For example, forward error correction is to add redundant information in coded video bitstreams [118], so it can protect original information from packet loss. Based on the source rate model proposed in this dissertation, it is promising to consider these factors jointly and develop a rate control framework for an end-to-end video transmission system.

- **Rate-control for SNR scalable video streaming over network with scalable quality of service**

Current Internet provides the best-effort service only. The quality of transmitted video cannot be guaranteed by the existing Internet, because the video stream is inherently variable bitrate and the Internet is an unpredictable time-varying channel. One promising solution is to allow the network to provide different quality of service levels. There are two typical approaches, the integration service [119] and differentiated service [120]. These internet protocols can provide scalable quality of service. For example, differentiated service protocols can provide two classes of service: 1) a premium service, which provides low loss and low delay service; and 2) an assured service, which provides better than best-effort service but without guarantee. The layered video bitstreams can be mapped to different classes well. In live video transmission, the base-layer can be mapped to the class with the best service, as it contains the most vital visual information; and the enhancement-layer will be given lower priority. Therefore, future work can be concentrated on rate control of layered video streaming over the quality-of-service guaranteed network.

- **Bit allocation for progressive FGS coding**

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

A bit allocation scheme for FGS coding has been presented in Chapter 6. It is quite robust and efficient. The FGS coder does not employ any temporal dependence among consecutive frames in the FGS-layer. To improve the coding efficiency, a progressive FGS (PFGS) coding framework has been proposed to use reference frames with higher quality reference frames compared with the FGS coder [121]. Namely, bitplane information in the FGS-layer of one frame is used in the prediction of the next frame. However, when those bitplane information used in the prediction are not truncated, they will not be available for decoding the next frame. As a result, the employment of the temporal dependence introduces error propagation in this scenario. The error propagation should be considered in the bit allocation for PFGS coding. The leaky prediction technique [92, 102] can give us some ideas to cope with error propagation in the scalable coding. It scales the reference frame by a leaky factor. Then the error, which is caused by the absence of information in the reference frame, decays as an exponential function of the leaky factor in the temporal direction. Combining the leaky prediction technique and the proposed R-D analysis, a bit allocation scheme for PFGS coding is worthy of and capable of being developed.

Appendix

Appendix A. Brief Review of Kolmogorov-Smirnov Test for the Goodness-of-fit

A well-known test for goodness-of-fit is the Kolmogorov-Smirnov test (KS-test) [103]. It is used to test whether or not the data samples are consistent with a specified distribution function. A brief description of the KS test is as follows. Suppose that the N sample observations are arranged in an increasing order as a set of data $\mathbf{Z} = \{z_1, z_2, \dots, z_N\}$. The empirical distribution function $G(z)$ for all those samples is defined in (A.1). The KS test tells the goodness-of-fit based on the differences between the empirical distribution function $G(z)$ and the specified distribution function $F(z)$. The KS test statistic t_{ks} is defined in (A.2), which is the maximum difference between $G(z_i)$ and $F(z_i)$ at the sample point z_i . A small value of the t_{ks} indicates a good fit.

$$G(z) = \begin{cases} 0 & \text{if } z < z_1 \\ \frac{i}{N} & \text{if } z_i \leq z < z_{i+1}, \quad i = 1, 2, \dots, N-1 \\ 1 & \text{if } z \geq z_N \end{cases} \quad (\text{A.1})$$

$$t_{ks} = \max_{i=1,2,\dots,N} |G(z_i) - F(z_i)| \quad (\text{A.2})$$

APPENDIX

Appendix B. Derivation: PDF of Quantization Error when a Laplacian Source is quantized by a uniform quantizer

Problem Description:

If a Laplacian source X is uniformly quantized, the probability density function (PDF) of quantization error is to be derived.

The Laplacian PDF is given by (A.3), where λ_x is the model parameter and $\frac{2}{\lambda_x^2}$ is equal to source variance (σ_x^2). The characteristics function of uniform threshold quantizer is defined as (A.4), where X is input, q is the quantization stepsize and t_0 ($t_0 \geq 0$) is the dead-zone. If t_0 is equal to zero, the quantizer has no dead-zone, which can be rewritten in (A.5). The staircase curves of uniform quantizer are plotted in Fig 7.1.

$$f^L(x) = \frac{\lambda_x}{2} \exp(-\lambda_x \cdot |x|) \quad (\text{A.3})$$

$$\text{UTQ}(X) = \begin{cases} 0 & |X| \leq t_0 \\ q \cdot \text{Round}[\frac{X-t_0}{q}] & X > t_0 \\ q \cdot \text{Round}[\frac{X+t_0}{q}] & X < -t_0 \end{cases} \quad (\text{A.4})$$

$$\text{UTQ}_0(X) = q \cdot \text{Round}[\frac{X}{q}] \quad (\text{A.5})$$

Derivation:

Let $E = X - \text{UTQ}(X)$ represent the quantization error. It can be easily concluded that E is distributed over the range $[-\frac{q}{2} - t_0, \frac{q}{2} + t_0]$. The cumulative distribution function (CDF) of E is given by (A.6).

APPENDIX

Figure 7.1: Uniform quantizer

$$\begin{aligned}
 F_e(e) = & P \left\{ -\frac{q}{2} - t_0 \leq X \leq e \right\} & e \in \left[-\frac{q}{2} - t_0, \frac{q}{2} + t_0 \right) \\
 & + \sum_{k \geq 1} P \left\{ -\frac{q}{2} + k \cdot q + t_0 \leq X \leq e + k \cdot q + t_0 \right\} & e \in \left[-\frac{q}{2}, \frac{q}{2} \right) \\
 & + \sum_{k \leq -1} P \left\{ -\frac{q}{2} + k \cdot q - t_0 \leq X \leq e + k \cdot q - t_0 \right\} & e \in \left[-\frac{q}{2}, \frac{q}{2} \right)
 \end{aligned} \quad (\text{A.6})$$

Combining the PDF of source X (A.3) with (A.6), the CDF of E is written as (A.7):

$$\begin{aligned}
 F_e(e) = & \int_{-\frac{q}{2}-t_0}^e f^L(x) dx & e \in \left[-\frac{q}{2} - t_0, \frac{q}{2} + t_0 \right) \\
 & + \sum_{k \geq 1} \int_{-\frac{q}{2}+k \cdot q+t_0}^{e+k \cdot q+t_0} f^L(x) dx & e \in \left[-\frac{q}{2}, \frac{q}{2} \right) \\
 & + \sum_{k \leq -1} \int_{-\frac{q}{2}+k \cdot q-t_0}^{e+k \cdot q-t_0} f^L(x) dx & e \in \left[-\frac{q}{2}, \frac{q}{2} \right)
 \end{aligned} \quad (\text{A.7})$$

The PDF of E ($f_e(e)$) can be calculated by the derivative of the CDF with respect to e as follows:

$$f_e(e) = \left(\int_{-\frac{q}{2}-t_0}^e f^L(x) dx \right)' + \left(\sum_{k \geq 1} \int_{-\frac{q}{2}+k \cdot q+t_0}^{e+k \cdot q+t_0} f^L(x) dx + \sum_{k \leq -1} \int_{-\frac{q}{2}+k \cdot q-t_0}^{e+k \cdot q-t_0} f^L(x) dx \right)' \quad (\text{A.8})$$

APPENDIX

(i) The first item of the PDF in (A.8):

$$\left(\int_{-\frac{q}{2}-t_0}^e f^L(x) dx \right)' = \frac{\lambda_x}{2} \exp(-\lambda_x |e|) \quad e \in [-\frac{q}{2} - t_0, \frac{q}{2} + t_0] \quad (\text{A.9})$$

(ii) The second item of the PDF in (A.8):

$$\begin{aligned} & \left(\sum_{k \geq 1} \int_{-\frac{q}{2}+k \cdot q+t_0}^{e+k \cdot q+t_0} f^L(x) dx + \sum_{k \leq -1} \int_{-\frac{q}{2}+k \cdot q-t_0}^{e+k \cdot q-t_0} f^L(x) dx \right)' \\ &= \sum_{k=1}^{+\infty} \frac{\lambda_x}{2} \exp(-\lambda_x |e + k \cdot q + t_0|) + \sum_{k=-\infty}^{-1} \frac{\lambda_x}{2} \exp(-\lambda_x |e + k \cdot q - t_0|) \\ &\because t_0 > 0, e \in [-\frac{q}{2}, \frac{q}{2}] \\ &= \sum_{k=1}^{+\infty} \frac{\lambda_x}{2} \exp[-\lambda_x (e + k \cdot q + t_0)] + \sum_{k=-\infty}^{-1} \frac{\lambda_x}{2} \exp[\lambda_x (e + k \cdot q - t_0)] \\ &= \frac{\lambda_x \cdot \exp(-\lambda_x t_0)}{2} \left\{ [\exp(-\lambda_x e) + \exp(\lambda_x e)] \sum_{k=1}^{+\infty} \exp(-\lambda_x q \cdot k) \right\} \\ &= \frac{\lambda_x \exp(-\lambda_x t_0 - \lambda_x q)}{2 [1 - \exp(-\lambda_x q)]} [\exp(-\lambda_x e) + \exp(\lambda_x e)] \\ &= \frac{\lambda_x \exp(-\lambda_x t_0 - \lambda_x q)}{2 [1 - \exp(-\lambda_x q)]} [\exp(-\lambda_x |e|) + \exp(\lambda_x |e|)] \quad e \in [-\frac{q}{2}, \frac{q}{2}] \quad (\text{A.10}) \end{aligned}$$

(iii) Combining (A.9) with (A.10), the PDF can be concluded as follows.

$$f_e(e) = \begin{cases} \frac{\lambda_x}{2} \exp(-\lambda_x |e|) & e \in [-\frac{q}{2} - t_0, -\frac{q}{2}] \cup [\frac{q}{2}, \frac{q}{2} + t_0] \\ \frac{\lambda_x \{ \exp(\lambda_x |e| - \lambda_x t_0 - \lambda_x q) + \exp(-\lambda_x |e|) \cdot [1 - \exp(-\lambda_x q) + \exp(-\lambda_x q - \lambda_x t_0)] \}}{2 [1 - \exp(-\lambda_x q)]} & e \in [-\frac{q}{2}, \frac{q}{2}] \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.11})$$

If the uniform quantizer has no dead-zone, namely $t_0 = 0$, the PDF of quantization error can be rewritten as:

$$f_e(e) = \frac{\lambda_x}{2 \cdot [1 - \exp(-\lambda_x q)]} [\exp(\lambda_x |e| - \lambda_x q) + \exp(-\lambda_x |e|)] \quad e \in [-\frac{q}{2}, \frac{q}{2}] \quad (\text{A.12})$$

APPENDIX

Appendix C. Proof: the optimal strategy of truncating NZBC

Problem Description:

The magnitude of source DCT coefficients in the FGS-layer ($X = \{x_k | 1 \leq k \leq N\}$) can be represented by binary coefficients ($b_k(j)$), as given by (A.13). $b_k(J_k)$ is the whole set of all truncated binary coefficients decomposed from x_k , where J_k can be defined as $\forall j \in \text{the set } J_k, b_k(j) \text{ is truncated}$. The binary coefficients decomposed from all the DCT coefficients (X) are assumed to be truncated in an arbitrary strategy $J = \{J_k | 1 \leq k \leq N\}$ first. Suppose there are two arbitrary NZBC which are not truncated, $b_{k1}(j1) \neq 0$ and $b_{k2}(j2) \neq 0$. If one more NZBC is to be truncated, which one should be selected for the minimal distortion?

$$x_k = b_k(0) \cdot 2^0 + b_k(1) \cdot 2^1 + \dots + b_k(j) \cdot 2^j + \dots \quad (\text{A.13})$$

Analysis:

According to the distortion expression (6.12), the distortion is given in (A.14), where $D_{j1,k1}$ is the distortion if only $b_{k1}(j1)$ is truncated and $D_{j2,k2}$ is the distortion if only $b_{k2}(j2)$ is truncated. In the following, the optimal strategy is given in three steps by comparing $D_{j1,k1}$ with $D_{j2,k2}$.

$$\left\{ \begin{array}{l} D_{j1,k1} = \sum_{k=0, k \neq k1}^N (x_k - \sum_{j \in J_k} b_k(j) \cdot 2^j)^2 + (x_{k1} - \sum_{j \in \{J_{k1}+j1\}} b_{k1}(j) \cdot 2^j)^2 \\ D_{j2,k2} = \sum_{k=0, k \neq k2}^N (x_k - \sum_{j \in J_k} b_k(j) \cdot 2^j)^2 + (x_{k2} - \sum_{j \in \{J_{k2}+j2\}} b_{k2}(j) \cdot 2^j)^2 \end{array} \right. \quad (\text{A.14})$$

(i) First, when $k1 = k2$ and $j1 \neq j2$

APPENDIX

$$\Delta D = D_{j_1, k_1} - D_{j_2, k_2} = \{2 \cdot (x_{k_1} - \sum_{j \in \{J_{k_1}\}} b_{k_1}(j) \cdot 2^j) - 2^{j_2} - 2^{j_1}\} (2^{j_2} - 2^{j_1}) \quad (\text{A.15})$$

If $j_1 > j_2$, then $\Delta D < 0$. It proves that the NZBC, which is located in the larger BP level, should be chosen on this condition.

(ii) Second, when $k_1 \neq k_2$ and $j_1 \neq j_2$.

Suppose $j_1 > j_2$, then (A.16) can be obtained. In addition, we assume that j_1 and j_2 satisfy the conclusion in step (i), which means that j_1 is the largest one in the set $\{j \mid b_{k_1}(j) \text{ is not truncated}\}$ and j_2 is the largest in the set $\{j \mid b_{k_2}(j) \text{ is not truncated}\}$. Then we can obtain (A.17).

$$2^{j_1} \geq 2 \cdot 2^{j_2} \quad (\text{A.16})$$

$$2^{j_2} > x_{k_2} - \sum_{j \in \{J_{k_2} + j_2\}} b_{k_2}(j) \cdot 2^j \quad (\text{A.17})$$

$$\begin{aligned} \Delta D = D_{j_1, k_1} - D_{j_2, k_2} &= 2^{2 \cdot j_2} + 2 \cdot 2^{j_2} \cdot (x_{k_2} - \sum_{j \in \{J_{k_2} + j_2\}} b_{k_2}(j) \cdot 2^j) \\ &\quad - \{2^{2 \cdot j_1} + 2 \cdot 2^{j_1} \cdot (x_{k_1} - \sum_{j \in \{J_{k_1} + j_1\}} b_{k_1}(j) \cdot 2^j)\} \end{aligned}$$

By combining (A.16) with (A.17), we can derive $\Delta D < 0$. It proves that we should select the one located in the higher BP level.

(iii) Third, when $k_1 \neq k_2$ and $j_1 = j_2$.

$$\Delta D = D_{j_1, k_1} - D_{j_2, k_2} = 2^{j_1+1} \cdot \{(x_{k_2} - \sum_{j \in J_{k_2}} b_{k_2}(j) \cdot 2^j) - (x_{k_1} - \sum_{j \in J_{k_1}} b_{k_1}(j) \cdot 2^j)\}$$

When $x_{k_1} - \sum_{j \in J_{k_1}} b_{k_1}(j) \cdot 2^j > x_{k_2} - \sum_{j \in J_{k_2}} b_{k_2}(j) \cdot 2^j$, $\Delta D < 0$, namely we should choose the one with a larger residue on this condition.

APPENDIX

In summary, the optimal strategy of selecting NZBC to truncate can be concluded as the following two points. First, the one which is located in the larger BP should be selected; second, for those NZBC in the same BP, the one having a larger residue should be selected.

APPENDIX

Appendix D. Sample of Test Video Sequences

The Sample picture of test video sequences used in this dissertation is shown as follows.

Carphone, 176×144

Foreman, 176×144

News, 176×144

Stefan, 176×144

Flower Garden, 352 × 240

Boating, 720×576

Mobile & Calendar, 720×576

References

- [1] T. Sikora, “MPEG digital video-coding standards,” *IEEE Signal Processing Magazine*, vol. 14, pp. 82–100, Sept. 1997.
- [2] J. Gemmell, H. M. Vin, D. D. Kandlur, P. V. Rangan, and L. A. Rowe, “Multimedia storage servers: A tutorial,” *IEEE Computer*, vol. 28, no. 5, pp. 40–49, 1995.
- [3] D. Wu, Y. T. Hou, W. Zhu, Y.-Q. Zhang, and J. M. Peha, “Streaming video over the Internet: approaches and directions,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 282–300, Mar. 2001.
- [4] G. Conklin, G. Greenbaum, K. Lillevold, A. Lippman, and Y. Reznik, “Video coding for streaming media delivery on the Internet,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 269–281, Mar. 2001.
- [5] ITU-T Recommendation H.261, “Video codec for audiovisual services at $p \times 64$ kbits,” Mar. 1993.
- [6] Information Technology, ISO/IEC 11172-2, “Coding of moving pictures and associated audio for digital storage media at up to about 1.5mbit/s: Video,” 1993.
- [7] Information Technology, ISO/IEC 13818-2, “Generic coding of moving pictures and associated audio information: Video. first edition,” May 1996.
- [8] D. LeGall, “MPEG: A video compression standard for multimedia application,” *Communications of the ACM*, vol. 34, pp. 46–58, Apr. 1991.
- [9] ITU-T Recommendation H.263 Version 1, “Video coding for low bitrate communication,” Mar. 1996.

REFERENCES

-
- [10] ITU-T Recommendation H.263 Version 2, “Video coding for low bitrate communication,” Sept. 1997.
 - [11] G. Cote, B. Erol, M. Gallant, and F. Kossentini, “H.263+: video coding at low bit rates,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 849 – 866, Nov. 1998.
 - [12] ISO/IEC 14496-2 MPEG-4 Visual Version 1 , “Coding of audio-visual objects - part 2: Visual,” Apr. 1999.
 - [13] T. Sikora, “The MPEG-4 video standard verification model,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 19 –31, Feb. 1997.
 - [14] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560 –576, July 2003.
 - [15] V. K. Goyal, “Theoretical foundations of transform coding,” *IEEE Signal Processing Magazine*, vol. 18, pp. 9–21, Sept. 2001.
 - [16] N. S. Jayant and P. Noll, *Digital Coding of Waveforms Principles and Applications to Speech and Video*. Englewood Cliff, New Jersey: Prentice-Hall, 1984.
 - [17] R. M. Gray and D. L. Neuhoff, “Quantization,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2325–2383, Oct. 1998.
 - [18] M. Ghanbari, “Two-layer coding of video signals for VBR networks,” *IEEE Journal on Selected Areas in Communications*, vol. 7, pp. 771–781, June 1989.
 - [19] H. Gharavi and M. H. Partovi, “Multilevel video coding and distribution architectures for emerging broadband digital networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 459 – 469, Oct. 1996.
 - [20] B. Lee, K. Park, and J. Hwang, “H.263-based SNR scalable video codec,” *IEEE Transactions on Consumer Electronics*, vol. 43, pp. 614–622, Aug. 1997.

REFERENCES

-
- [21] T. Gardos, "Efficient receiver-driven layered video multicast using h.263+ snr scalability," in *International Conference on Image Processing*, vol. 3, June 1998, pp. 32–35.
 - [22] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 301–317, Mar. 2001.
 - [23] H.Radha, M. van der Schaar, and Y.chen, "The MPEG-4 fine-grained scalable video coding method for multimedia streaming over IP," *IEEE Transactions on Multimedia*, vol. 3, pp. 53–68, Mar. 2001.
 - [24] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *Institute of Radio Engineers, International Convention Record, Part 4*, vol. 7, pp. 142 – 163, 1959.
 - [25] T. Berger, *Rate-distortion theory: A mathematical basis for data compression*. Englewood Cliff, New Jersey: Prentice-Hall, 1971.
 - [26] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, vol. 5, pp. 676–683, Sept. 1968.
 - [27] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 246 –250, Feb. 1997.
 - [28] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application: I. fundamental theory," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 287 –298, Apr. 1997.
 - [29] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 172 –185, Feb. 1999.
 - [30] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 12 –20, Feb. 1996.

REFERENCES

-
- [31] L.-J. Lin and A. Ortega, “Bit-rate control using piecewise approximated rate-distortion characteristics,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 446–459, Aug. 1998.
 - [32] L. Wang, “Rate control for MPEG video coding,” *Signal Processing: Image Communications*, vol. 15, pp. 493–511, 2000.
 - [33] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, “Scalable rate control for MPEG-4 video,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 878–894, Sept. 2000.
 - [34] Z. He and S. K. Mitra, “A unified rate-distortion analysis framework for transform coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 1221–1236, Dec. 2001.
 - [35] ———, “A linear source model and a unified rate control algorithm for DCT video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 970–982, Nov. 2002.
 - [36] K. Ngan, T. Meier, and Z. Chen, “Improved single-video-object rate control for MPEG-4,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 385–393, May 2003.
 - [37] F. Pan, Z. Li, K. Lim, and G. Feng, “A study of MPEG-4 rate control scheme and its improvements,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 440–446, May 2003.
 - [38] “MPEG-2 Video Test Model 5,” ISO/IEC JTC1/SC29/WG11, MPEG93/457, Apr. 1993.
 - [39] Video Coding Experts Group, ITU, “Video Codec Test Model, Version 10(TMN10),” Apr. 1998.
 - [40] “MPEG-4 Video Verification Model 18.0,” ISO/IEC JTC1/SC29/WG11, N3908, Jan. 2001.

REFERENCES

-
- [41] R. Reininger and J. Gibson, "Distributions of the two-dimensional DCT coefficients for images," *IEEE Transactions on Communications*, vol. 31, pp. 835–839, June 1983.
 - [42] F. Muller, "Distribution shape of two-dimensional DCT coefficients of natural images," *Electron. Lett.*, vol. 29, pp. 1935–1936, Oct. 1993.
 - [43] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, pp. 1661–1665, Oct. 2000.
 - [44] F. Bellifemine, A. Capellino, A. Chimienti, R. Picco, and R. Ponti, "Statistical analysis of the 2D-DCT coefficients of the differential signal of images," *Signal Processing: image communication*, vol. 4(6), pp. 477–488, 1992.
 - [45] G. Yovanof and S. Liu, "Statistical analysis of the DCT coefficients and their quantization error," in *Conference Record of the Thirtieth Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 1996, pp. 601 – 605.
 - [46] M. Robertson and R. Stevenson, "DCT quantization noise in compressed images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 27–38, Jan. 2005.
 - [47] J. Xie, L.-T. Chia, and B.-S. Lee, "An improved distortion model for rate control of DCT-based video coding," in *IEEE international Conference on Multimedia Modeling*, Jan. 2006, p. to appear.
 - [48] Y. Yan and K. Challapali, "A novel distortion-quantization model and its application in low bitrate video communications," in *International Conference on Image Processing*, vol. 1, Sept. 2002, pp. 81–84.
 - [49] M. Dai, D. Loguinov, and H. Radha, "Statistical analysis and distortion modeling of MPEG-4 FGS," in *International Conference on Image Processing*, vol. 3, Sept. 2003, pp. III– 301–304.

REFERENCES

-
- [50] D. Wilson and M. Ghanbari, "Optimization of MPEG-2 SNR scalable codecs," *IEEE Transactions on Image Processing*, vol. 8, pp. 1435–1438, Oct. 1999.
 - [51] Q. Wang, Z. Xiong, F. Wu, and S. Li, "Optimal rate allocation for progressive fine granularity scalable video coding," *IEEE Signal Processing Letters*, vol. 9, pp. 33–39, Feb. 2002.
 - [52] X. M. Zhang, A. Vetro, Y. Q. Shi, and H. Sun, "Constant quality constrained rate allocation for FGS-coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 121–130, Feb. 2003.
 - [53] J. Xie, L.-T. Chia, and B.-S. Lee, "Optimum bit allocation for FGS coding," *IEEE Electron. Lett.*, vol. 40, pp. 983–984, Aug. 2004.
 - [54] M. J. and S. D., "The effects of a visual fidelity criterion of the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, pp. 525–536, July 1974.
 - [55] S. R. Smoot and L. A. Rowe, "Study of DCT coefficients distributions," *SPIE Symposium on in Human Vision and Electronic Imaging*, vol. 2657, pp. 403–411, Jan. 1996.
 - [56] "Coding of Audio-visual objects, Part-2 Visual, Amendment 4: Streaming video profile," ISO/IEC 14496-2/FPDAM, July 2000.
 - [57] F. Ling, Q. Li, and H. Sun, "Bitplane coding of DCT coefficients for image and video compression," in *SPIE Visual Communications and Image Processing (VCIP)*, San Jose, CA, Jan. 1999, pp. 25–27.
 - [58] M. Dai, D. Loguinov, and H. Radha, "Rate-distortion modeling of scalable video coders," in *ICIP*, vol. 2, Oct. 2004, pp. 1093 – 1096.
 - [59] R. M. Gray, *Source coding theory*. Norwell, MA, USA: Kluwer Academic Publishers, 1990.
 - [60] N. Mohsenian, R. Rajagopalan, and C. A. Gonzales, "Single-pass constant- and variable-bit-rate MPEG-2 video compression," *IBM J. RES. DEVELOP*, vol. 43, pp. 489 – 509, July 1999.

REFERENCES

-
- [61] D. Bagni, B. Biffi, and R. Ramalho, “A constant-quality, single-pass VBR control for DVD recorders,” *IEEE Transactions on Consumer Electronics*, vol. 49, pp. 653 – 662, Aug. 2003.
 - [62] J. Chen, U.-V. Koc, and K. J. R. Liu, *Design of Digital Video Coding Systems*. New York: Marcel Dekker, 2002.
 - [63] B. Girod, “The efficiency of motion-compensating prediction for hybrid coding of video sequences,” *IEEE Journal on Selected Areas in Communications*, vol. 5, pp. 1140 – 1154, Aug. 1987.
 - [64] —, “Efficiency analysis of multihypothesis motion-compensated prediction for video coding,” *IEEE Transactions on Image Processing*, vol. 9, pp. 173 – 183, Feb. 2000.
 - [65] N. Ahmed, T. Natarajan, and R. Rao, “Discrete cosine transform,” *IEEE Transactions on Computers*, vol. C-23, pp. 90–93, Jan. 1974.
 - [66] W. Li and Y.-Q. Zhang, “Vector-based signal processing and quantization for image and video compression,” *Proceedings of the IEEE*, vol. 83, pp. 317 – 335, Feb. 1995.
 - [67] N. Jayant, J. Johnston, and R. Safranek, “Signal compression based on models of human perception,” *Proceedings of the IEEE*, vol. 81, pp. 1385 – 1422, Oct. 1993.
 - [68] M. Ghanbari, *Video coding an introduction to standard codecs*. London, United Kingdom: The institution of electrical engineers, 1999.
 - [69] Ohm, J.-R., “Advances in scalable video coding,” *Proceedings of the IEEE*, vol. 93, pp. 42 – 56, Jan. 2005.
 - [70] Y. Wang and Q.-F. Zhu, “Error control and concealment for video communication: a review,” *Proceedings of the IEEE*, vol. 86, pp. 947–997, May 1998.
 - [71] M. Gallant and F. Kossentini, “Rate-distortion optimized layered coding with unequal error protection for robust internet video,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 357–372, Mar. 2001.

REFERENCES

-
- [72] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Digital video processing and communication*. Prentice Hall, 2001.
 - [73] R. Aravind, M. Civanlar, and A. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 426–435, Oct. 1996.
 - [74] M.-T. Sun and A. R. Reibman, *Compressed Video over Networks*. New York: Marcel Dekker, 2001.
 - [75] J. Arnold, M. Frater, and Y. Wang, "Efficient drift-free signal-to-noise ratio scalability," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 70–82, Feb. 2000.
 - [76] R. A.R., B. L., and B. A., "Scalable video coding with managed drift," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 131–140, Feb. 2003.
 - [77] M. van der Schaar and H. Radha, "A hybrid temporal-SNR fine-granular scalability for Internet video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 318–330, Mar. 2001.
 - [78] M. van der Schaar and Y.-T. Lin, "Content-based selective enhancement for streaming video," in *International Conference on Image Processing*, vol. 2, Oct. 2001, pp. 977–980.
 - [79] Y. He, S. Yang, and Y. Zhong, "Block-based fine granularity scalable video coding for content-aware streaming," in *International Conference on Image Processing*, vol. 2, Sept. 2002, pp. II. 45–48.
 - [80] W. Li, "Frequency weighting for FGS," ISO/IEC JTC1/SC29/WG11, MPEG99/M5489, Dec. 1999.
 - [81] S. Peng and van der Schaar M., "Adaptive frequency weighting for Fine-Granularity-Scalability," in *Visual Communications and Image Processing*, vol. 4671, Oct. 2002, pp. 840–849.

REFERENCES

-
- [82] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: John Wiley, 1991.
 - [83] T. Berger and J. D. Gibson, “Lossy source coding,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2693–2723, Oct. 1998.
 - [84] H.-M. Hang and J.-J. Chen, “Source model for transform video coder and its application: II. variable frame rate coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 299–311, Apr. 1997.
 - [85] H. Song and C.-C. Kuo, “Rate control for low-bit-rate video via variable-encoding frame rates,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 512–521, Apr. 2001.
 - [86] E. Reed and J. Lim, “Optimal multidimensional bit-rate control for video communication,” *IEEE Transactions on Image Processing*, vol. 11, pp. 873–885, Aug. 2002.
 - [87] K. Miyata and A. Taguchi, “A novel MPEG-4 rate control method with spatial resolution conversion for low bit-rate coding,” in *International Symposium on Circuits and Systems (ISCAS)*, vol. 5, May 2005, pp. 4538–4541.
 - [88] G. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, Nov. 1998.
 - [89] B. Tao, B. Dickinson, and H. Peterson, “Adaptive model-driven bit allocation for MPEG video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 147–157, Feb. 2000.
 - [90] R. Neff and A. Zakhor, “Matching pursuit video coding .I. dictionary approximation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 13–16, Jan. 2002.
 - [91] —, “Matching-pursuit video coding .II. operational models for rate and distortion,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 27–39, Jan. 2002.

REFERENCES

-
- [92] H.-C. Hunag, C.-N. Wang, and T. Chiang, “A robust fine granularity scalability using trellis-based predictive leak,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 372–384, June 2002.
 - [93] H. M. Kim, “Adaptive rate control using nonlinear regression,” *IEEE Trans. Circuits Syst. Video Techn.*, vol. 13, no. 5, pp. 432–439, 2003.
 - [94] J.-C. Tsai, “Rate control for low-delay video using a dynamic rate table,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, pp. 133 – 137, Jan. 2005.
 - [95] Z. He, Y. K. Kim, and S. K. Mitra, “Low-delay rate control for DCT video coding via ρ -domain source modeling,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 928 –940, Aug. 2001.
 - [96] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. New York: McGraw-Hill, 1979.
 - [97] Z. He, “ ρ -domain rate control analysis and rate control for visual coding and communication,” Ph.D. dissertation, School of Electrical and Computer Engineer, University of California, Santa Barbara, June 2001.
 - [98] P. H. Westerink, R. Rajagopalan, and C. A. Gonzales, “Two-pass MPEG-2 variable-bit-rate encoding,” *IBM J. RES. DEVELOP*, vol. 43, pp. 489 – 509, July 1999.
 - [99] L. Overmeire, F. Verdicchio, J. Barbarie, P. Schelkens, and L. Nachtergaele, “Constant quality rate-control for video encoding based on activity segmentation,” in *International Conference on Image Processing*, vol. 1, Sept. 2002, pp. 81–84.
 - [100] B. Xie and W. Zeng, “Sequence-based rate control for constant quality video,” in *International Conference on Image Processing*, vol. 1, Sept. 2002, pp. 77–80.
 - [101] M. Jiang, X. Yi, and N. Ling, “Frame layer bit allocation scheme for constant quality video,” in *International Conference on Multimedia and Expo*, vol. 2, June 2004, pp. 1055–1058.

REFERENCES

-
- [102] M. Ghanbari and V. Seferidis, “Efficient H.261-based two-layer video codecs for ATM networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 171 – 175, Apr. 1995.
 - [103] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. New York: Marcel Dekker, 2003.
 - [104] T. K. Moon, “The expectation-maximization algorithm,” *IEEE Signal Processing Magazine*, vol. 13, pp. 47 – 60, Nov. 1996.
 - [105] G. Triantafyllidis, D. Tzovaras, and M. Strintzis, “Blocking artifact detection and reduction in compressed data,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 877 – 890, Oct. 2002.
 - [106] A. Sripad and D. Snyder, “A necessary and sufficient condition for quantization errors to be uniform and white,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, pp. 442–448, Oct. 1977.
 - [107] J. Liao and J. Villasenor, “adaptive intra block update for robust transmission of H.263,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 30 – 35, Feb. 2000.
 - [108] M. Gallant and F. Kossentini, “Efficient scalable DCT-based video coding,” in *Proceeding of the IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 2, Alberta, Canada, May 1999.
 - [109] E. Miloslavsky and A. Zakhori, “Rate control for layered video compression using matching pursuits,” in *International Conference on Image Processing*, vol. 2, 1999, pp. 357–361.
 - [110] Z. He and S. K. Mitra, “Optimum bit allocation and accurate rate control for video coding via ρ -domain source modeling,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 840 – 849, Oct. 2002.
 - [111] “<http://www.mpeg.org/mpeg/mssg/#source>.”

REFERENCES

-
- [112] L. Zhao, J. W. Kim, and C.-C. Kuo, “Constant quality rate control for streaming MPEG-4 FGS video,” in *IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 4, Mar. 2002, pp. IV.544–547.
 - [113] Y. Shoham and A. Gersho, “Efficient bit allocation for an arbitrary set of quantizers,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 1445–1453, Sept. 1988.
 - [114] S. Parthasarathy and H. Radha, “Optimal rate control methods for fine granularity scalable video,” in *International Conference on Image Processing*, vol. 2, Sept. 2003, pp. III– 805–809.
 - [115] D.C. Montgomery and E.A. Peck, *Introduction to linear regression analysis*. John Wiley and Sons, 1992.
 - [116] H. Schulzrinne and S. Casner, “RTP Profile for Audio and Video Conferences with Minimal Control,” July 2003, RFC 3551.
 - [117] Floyd, S. and Fall, K., “Promoting the use of end-to-end congestion control in the internet,” *IEEE/ACM Transactions on Networking*, vol. 4, pp. 458–472, Aug. 1999.
 - [118] P. Frossard, “Fec performance in multimedia streaming,” *IEEE Communications Letters*, vol. 5, pp. 122 – 124, Mar. 2001.
 - [119] J. Wroclawski, “The Use of RSVP with IETF Integrated Services,” Sept. 1997, RFC 2210.
 - [120] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An Architecture for Differentiated Services,” Dec. 1998, RFC 2475.
 - [121] F. Wu, S. Li, and Y.-Q. Zhang, “A framework for efficient progressive fine granularity scalable video coding,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 332–344, Mar. 2001.

List of Publication

- 1) Jun Xie, Liang-Tien Chia, Bu-Sung Lee, "Optimum bit allocation for FGS video coding" *Proceedings - International Conference on Image Processing (ICIP)*, Pages 2295-2298, 2004
- 2) Jun Xie, Liang-Tien Chia, Bu-Sung Lee, "Optimal bit allocation for FGS coding", *Electronics Letters*, 40(16), Pages 983-984, 2004
- 3) Jun Xie, Liang-Tien Chia, Bu-Sung Lee, "Enhancement layer rate control for high bitrate SNR scalable video coding", *Journal of Visual Communication and Image Representation*, 16 (2), Pages 159-179, 2005
- 4) Jun Xie, Liang-Tien Chia, Bu-Sung Lee, "An Improved Distortion Model for rate control of DCT-based Video Coding", *12th IEEE International Conference on MultiMedia Modelling*, 2006