Efficient rate-distortion optimization techniques for H.264

Yu, Hongtao

2009

Yu, H. (2009). Efficient rate-distortion optimization techniques for H.264. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/19274>

<https://doi.org/10.32657/10356/19274>

Efficient Rate-Distortion Optimization Techniques for H.264

Yu Hongtao

School of Electrical & Electronic Engineering

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

2009

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

.....

Date

.....

Yu Hongtao

Acknowledgements

I would like to thank Nanyang Technological University (NTU) for granting me the scholarship and opportunity to do the research of this thesis. I would also like to thank Software Engineering Lab, NTU, and Institute for Infocomm Research (I²R) for the provision of the necessary working facilities during my Ph.D. study on this thesis.

I would like to express my sincere appreciation and respect to Associate Professor Lin Zhiping, my supervisor, for his extensive guidance, invaluable support and constant encouragement on my research work. I appreciate him not for only trying his best to cultivate my academic and research capabilities but also enthusiastically encouraging and helping me to achieve life goal.

I would also like to express my sincere appreciation to Dr. Pan Feng, former scientist at I²R and informal co-supervisor of my Ph.D. thesis, who shared his professional knowledge and ideas with me, provided me with many useful references, and valuable advices for my research. I appreciate the discussions with him. His intelligence and open-mindedness made research work enjoyable.

I would like to extend my gratitude to staffs and research students at I²R and Software Engineering Lab. Dr. Wu Dajun supported me throughout my research work. I would not forget the scene that we discussed and worked together till midnight. Dr. Wu Shiqian, Dr. Li Zhengguo and Dr. Lim Keng Pang gave me good research suggestions. Mr. Su Xinrong, Mr. Sun Yin, Mr. Kenny Choo Tsu Wei, and Mr. Chin Zhengyang shared their expertise on programming and helped me on subjective evaluation of video sequences. Dr. Zheng Jinghong, Miss Zhou Nina and Miss Liu Wen also did me favor in their own ways.

I would also like to thank Dr. Tan Ru San from National Heart Centre of Singapore for his kind evaluation of medical video sequences.

Last but not least, I must owe my deepest gratitude to my family members for their constant support and encouragement rendered during this period.

Summary

This thesis addresses the rate-distortion optimization (RDO) techniques for H.264. The research of the current project aims at improving the coding efficiency of video sequences. It proposes several efficient RDO techniques for H.264. These techniques include rate control, perceptual bit allocation and fast mode decision.

To solve the rate control problem in H.264, we propose a motion complexity measure and a minimum distortion based mode decision process. The motion complexity measure can represent the complexity of a video frame's motion contents effectively and is used to estimate the frame's target bit in frame level bit allocation. Different from other existing H.264 rate control methods, quantization parameter (QP) of a macroblock is accurately estimated based on the minimum distortion after full-search of all coding modes. According to the motion complexity measure and minimum distortion mode decision process, an efficient rate control scheme is proposed.

To solve the perceptual bit allocation problem in H.264, we propose a Lagrangian multiplier update algorithm in mode decision. The Lagrangian multiplier update algorithm decides the macroblock coding mode according to the perceptual characteristics of video contents, so that more bits are allocated to pixels which are perceptually more important and fewer bits are allocated to pixels which are perceptually less important.

To solve the fast mode decision problem in H.264, we propose a scalable fast RDO algorithm based on the occurring probability of different coding modes. We provide a priority queue in which the coding modes are placed according to their popularities. The statistical properties of macroblock coding modes are analyzed to determine the order of coding modes in the priority queue so that the most probable mode will be checked first, followed by the second most probable mode, and so forth. The process will be terminated as soon as the computed rate-distortion (RD) cost is below a threshold that is content adaptive and is also dependent on the RD cost of the previous macroblocks. By adjusting the threshold we achieve a good tradeoff between time saving and peak-signal-to-noise ratio (PSNR).

Contents

Acknowledgements	iii
Summary	iv
List of Figures	ix
List of Tables	xii
List of Abbreviations	xiv
1 Introduction	1
1.1 Motivation	1
1.2 Objectives	4
1.3 Major Contributions of the Thesis	4
1.4 Organization of the Thesis	6
2 Background	8
2.1 H.264 Video Coding Standard	8
2.2 Rate-Distortion Optimization	10
2.3 H.264 Rate Control	15
2.4 H.264 Perceptual Bit Allocation	19
2.5 H.264 Fast Mode Decision	22

3	Content Adaptive Rate Control Scheme for H.264	28
3.1	Introduction	28
3.2	Mode Decision Process Based upon Minimum Distortion	29
3.3	Content Adaptive Motion Complexity Measure	32
3.4	Minimum Distortion Based QP Estimation Rate Control	38
3.5	Experimental Results	43
3.6	Summary	58
4	Perceptual Bit Allocation Scheme for H.264	60
4.1	Introduction	60
4.2	Perceptual Mode Decision	62
4.3	Perceptual Bit Allocation Algorithm	67
4.4	Experimental Results	67
4.4.1	Objective Quality Evaluation	69
4.4.2	Subjective Quality Evaluation	73
4.5	Region of Interest Bit Allocation Applications for Medical Sequences	74
4.6	Summary	81
5	Scalable Fast Mode Decision Scheme for H.264	82
5.1	Introduction	82
5.2	Statistical Study on Macroblock Coding Modes in H.264	85
5.2.1	Probability Distribution of Macroblock Coding Modes	85
5.2.2	Mean Value and Standard Deviation of RD Cost	89

5.2.3	Correlation Coefficient of RD Cost	91
5.3	Priority Mechanism and Early Termination Strategy	92
5.3.1	Priority Mechanism	92
5.3.2	Early Termination Strategy	94
5.4	Fast Mode Decision Algorithm	95
5.5	Experimental Results	97
5.6	Summary	107
6	Conclusion and Recommendations	108
6.1	Conclusion	108
6.2	Recommendations	110
	Author's Publications	112
	Bibliography	114

List of Figures

Figure 1.1	H.264 encoder	2
Figure 2.1	RD curve	11
Figure 2.2	Formulation process of coefficient bit	12
Figure 2.3	CBR and VBR	14
Figure 2.4	Encoder output buffer and decoder input buffer	14
Figure 2.5	Rate controller	14
Figure 2.6	Macroblock coding modes in H.264	16
Figure 2.7	Parallel mode decision scheme of RDO	23
Figure 2.8	Modified encoder with fast mode decision	24
Figure 3.1	Mode decision process of the H.264 rate control algorithms	29
Figure 3.2	Minimum distortion based mode decision process	30
Figure 3.3	Flow chart of minimum distortion based mode decision process	32
Figure 3.4	Bit allocation of sequence “Foreman”	33
Figure 3.5	Two consecutive frames with high motion in sequence “Foreman” ...	33
Figure 3.6	Motion complexity and normalized bits of sequence “Foreman”	35
Figure 3.7	Normalized bits and motion complexity	37
Figure 3.8	Flow chart of proposed rate control scheme	39
Figure 3.9	Buffer model	39
Figure 3.10	RD performance comparison between JVT-H014 and proposed	

	method	52
Figure 3.11	PSNR vs. frame number	53
Figure 3.12	Two consecutive frames with camera motion in sequence “Coastguard”	54
Figure 3.13	PSNR vs. frame number for very high and very low motion sequences	55
Figure 4.1	Mode migration	62
Figure 4.2	The 32nd frame of sequence “Echocardiography”	65
Figure 4.3	The 44th frame of sequence “CT”	66
Figure 4.4	RD performance comparison between JVT-H014 and proposed scheme	71
Figure 4.5	PSNR vs. frame number	73
Figure 4.6	The 3rd frame of sequence “LV”	76
Figure 4.7	PSNR vs. frame number inside ROI	80
Figure 4.8	PSNR vs. QP inside ROI	81
Figure 5.1	Sequential mode decision scheme of RDO	84
Figure 5.2	Probability distribution of macroblock coding modes for different sequences	88
Figure 5.3	Mean of RD cost vs. macroblock number for sequence “Foreman” ...	90
Figure 5.4	Standard deviation of RD cost vs. macroblock number for sequence “Foreman”	90
Figure 5.5	Correlation coefficient of RD cost	92
Figure 5.6	Priority mechanism	93

Figure 5.7	Flowchart of proposed fast RDO algorithm	96
Figure 5.8	The mode distribution for the 45th frame of sequence “Foreman”	100
Figure 5.9	The 45th frame of sequence “Foreman”	101
Figure 5.10	Average PSNR difference and average time difference	103
Figure 5.11	Average PSNR difference and time difference for sequence “Carphone”	105
Figure 5.12	Average PSNR difference and time difference for sequence “Container”	106
Figure 5.13	Average PSNR difference and time difference for sequence “Stefan”	107

List of Tables

Table 2.1	Quantization step sizes in H.264 codec	13
Table 3.1	Similarity between bits and motion complexity	38
Table 3.2	Test conditions	44
Table 3.3	Specifications of the test sequences	44
Table 3.4	Target bit rate	45
Table 3.5	Results achieved by JVT-H014 and proposed rate control scheme ...	47
Table 3.6	Average PSNR difference	50
Table 3.7	Average bit rate difference	51
Table 3.8	Average bit rate mismatch ratio	56
Table 4.1	Test conditions	68
Table 4.2	Specifications of the test sequences	68
Table 4.3	Target bit rate	68
Table 4.4	Results of JVT-H014 and proposed scheme	69
Table 4.5	Average PSNR difference over complete range of bit rate between JVT-H014 and the proposed perceptual bit allocation scheme	70
Table 4.6	Average bit rate difference over complete range of PSNR between JVT-H014 and the proposed perceptual bit allocation scheme	70
Table 4.7	Subjective score for sequence “CT”	74
Table 4.8	Subjective score for sequence “Echocardiography”	74

Table 4.9	Subjective score of JVT-H014 and the proposed bit allocation scheme	75
Table 4.10	Target bit rate	77
Table 4.11	PSNR and bit rate	78
Table 4.12	Average PSNR difference over complete range of bit rate and average bit rate difference over complete range of PSNR	79
Table 4.13	Average PSNR inside ROI (dB)	79
Table 5.1	Mean value and standard deviation of RD cost for sequence “Foreman”	89
Table 5.2	Test conditions	98
Table 5.3	Experimental results, $\alpha = 0.3$	100
Table 5.4	Performance details for sequence “Foreman”	100
Table 5.5.	Experimental results, $\alpha = 0.5$	102
Table 5.6	Comparison between Grecos’s algorithm and proposed algorithm, $\alpha = 0.3$	104
Table 5.7	Comparison between Grecos’s algorithm and proposed algorithm, $\alpha = 0.5$	105
Table 5.8	Comparison between Grecos’s algorithm and proposed algorithm, $\alpha = 0.7$	106

List of Abbreviations

AVC	Advanced Video Coding
CBR	Constant Bit Rate
CGS	Coarse Gain Scalability
CIF	Common Intermediate Format
CODEC	Encoder/Decoder
DSCQS	Double Stimulus Continuous Quality Scale
GOP	Group of Pictures
HVS	Human Vision System
IEC	International Electrotechnology Commission
ISO	International Standards Organization
ITU	International Telecommunications Union
ITU-R	ITU – Radiocommunication Sector
ITU-T	ITU – Telecommunication Standardization Sector
JVT	Joint Video Team
LMU	Lagrangian Multiplier Update
LV	Left Ventricle
MAD	Mean Absolute Difference
MCBE	Motion Complexity Based Bit Estimation
MDQE	Minimum Distortion Based QP Estimation

MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
MQE	MAD Based QP Estimation
MRI	Magnetic Resonance Imaging
NP	Nondeterministic Polynomial
PSNR	Peak Signal-to-Noise Ratio
QCIF	Quarter Common Intermediate Format
QP	Quantization Parameter
QPU	QP Update
RD	Rate-Distortion
RDO	Rate-Distortion Optimization
ROI	Region of Interest
SAD	Sum of Absolute Differences
SATD	Sum of Absolute Transform Differences
SNR	Signal-to-Noise Ratio
SSD	Sum of Squared Differences
SVC	Scalable Video Coding
TM	Test Model
VBR	Variable Bit Rate
VCEG	Video Coding Experts Group

Chapter 1

Introduction

1.1 Motivation

H.264 [1] is the state of the art video coding standard developed by the Joint Video Team (JVT) consisting of experts from ITU-T's video coding experts group (VCEG) and ISO/IEC's moving picture experts group (MPEG). It has achieved a significant improvement in rate-distortion (RD) efficiency compared with the existing standards. It has been recognized as the most powerful video coding standard since the beginning of video coding [2][3].

H.264 employs a well-known hybrid block-based coding structure, as illustrated in Figure 1.1. Specifically, since the quantization part introduces permanent information loss to video data, the hybrid coding scheme leads to a lossy compression, whose performance is characterized by the RD function of the source [4]. As shown in Figure 1.1, modules of transformation/quantization, entropy coding, inverse quantization/inverse transformation, motion estimation/compensation all contribute to the RD function and there is no easy way to quantitatively separate their contributions. Therefore, the fundamental tradeoff in the design of H.264 which has such hybrid coding structure is its overall RD performance, based on which many optimization methods, broadly referred to

as RD optimization (RDO) methods, have been developed and widely used in H.264 video compression applications [5].

Figure 1.1. H.264 encoder

As in other video coding standards, rate control remains as an open issue for H.264 standard. In video coding, rate control is a technique that moderates the bit rate of video transmission over a channel of limited bandwidth and controls the bit allocation within the video sequence in order to maximize its overall visual quality. Rate control is implemented by allocating bits to transform coefficients, motion vectors, modes, and so on. In H.264, the Lagrangian multiplier method is used for mode selection. However, utilizing the Lagrangian multiplier method also makes rate control a difficult task because the quantization parameter (QP) is involved in the rate and distortion calculation. That is, the RD cost of a macroblock is associated with the QP decided by the rate control, while RDO should be performed firstly to obtain necessary source information for determining QP in the rate control. This is a contradiction and dilemma. For this reason, it is difficult to adopt existing rate control algorithms for previous video coding standards directly for

H.264. This has motivated many researchers to propose new rate control algorithms for H.264.

With standard rate control methods, all macroblocks are treated as they have the same importance and accordingly bits are allocated to the macroblocks with the same importance. However, human vision is sensitive to movement of well structured objects but more tolerant to large distortion in regions with randomly moving structures. Since the random structure region contains many random stimuli (small edges) that cover up coding noises, it is treated as unimportant region. In bit allocation, more bits are allocated to the important region satisfying some target visual quality while allowing the quality of unimportant region to gracefully degrade. Such bit allocation technique is called perceptual bit allocation. It aims at providing preferential treatment for the region of the video frame that is deemed to be of higher perceptual importance to the viewer. The difficulty of implementing perceptual bit allocation comes firstly from the detection of perceptual region that is an extra task compared with standard rate control methods. Thus, how to implement perceptual bit allocation method is a more challenging task than those standard rate control methods.

In H.264 encoder, RDO techniques are used to get the best coding efficiency in terms of maximizing video quality and minimizing allocated data bits. The idea of RDO is implemented by the RD cost computation and mode decision, as illustrated in Figure 1. The encoder experiments on all possible macroblock coding modes to get the block segmentation, reference frame, motion vector, quantization level, and so on. After entropy coding, the bits allocated to a certain macroblock coding mode is determined. The distortion between current frame and reference frame is also obtained after motion compensation. RD cost with regard to a certain mode is computed using the information

of rate, distortion and a Lagrangian multiplier. The computation of rate and distortion contributes to the overall RDO in achieving much better coding efficiency while compared with other video coding standards. On the other hand, such computation increases the expense of computational complexity and makes real time video coding using H.264 a difficult problem. Therefore, algorithms that reduce computational complexity of H.264 while at the same time maintaining the coded bit rate and reconstructed video quality are indispensable for real time implementation of H.264.

1.2 Objectives

We aim at developing efficient RDO techniques for H.264. Our objectives are:

- (1) Develop new and efficient rate control methods: These methods should improve the performance of existing H.264 rate control methods and produce high quality decoded video at a given target bit rate.
- (2) Develop new and efficient perceptual bit allocation methods: These methods should improve the performance of existing H.264 rate control methods in the sense of perceptual video quality.
- (3) Develop new and efficient fast mode decision methods: These methods should improve the performance of existing H.264 mode decision methods with respect to encoding time saving with negligible loss of coding efficiency.

1.3 Major Contributions of the Thesis

We focus on developing new methods that can improve the RD performance of existing H.264 methods. Our contributions are as follows:

- (1) To solve the rate control problem, we aim to achieve consistent video quality at scene changes and high motions, and estimate QP accurately. Though the consumed bit is a good measure to represent the complexity of a frame's motion contents, it is difficult to obtain the actual consumed bit before entropy coding. Thus a simple and effective measure, called motion complexity, is proposed to represent the complexity of motion contents. Motion complexity is the ratio between the predicted bit of the current frame and the average consumed bit of the previously encoded frames. It is used in the frame-level bit estimation. Since the QP estimation is included in the mode decision process, a new mode decision process is proposed. Different from other existing H.264 rate control methods, QP is accurately estimated after the full-search of all coding modes, rather than performing it before motion estimation. After motion estimation, the minimum distortion has been achieved and is used in the QP model to estimate QP in macroblock-level QP estimation. Based on the minimum-distortion mode decision process, a rate control scheme is proposed. Experimental results show that the proposed rate control scheme achieves better performance than the standard H.264 rate control proposal for low motion sequences in general, and for high motion sequences at scene changes and high motions.

- (2) To solve the perceptual bit allocation problem, we propose a Lagrangian multiplier update algorithm. The Lagrangian multiplier update algorithm decides the macroblock coding mode according to the perceptual characteristics of video contents, so that more bits are allocated to the pixels with perceptual importance and fewer bits are allocated to the pixels without perceptual importance. Experimental results show that the proposed perceptual bit allocation schemes can achieve better video quality for both the standard testing videos and medical videos.
- (3) To solve the fast mode decision problem, we aim at not only the saving of encoding time but also the scalability of time saving and coding efficiency. We provide a priority queue which places the coding modes. The statistical properties of macroblocks are analyzed to determine the order of coding modes in the mode decision priority queue such that the most probable mode will be checked first, followed by the second most probable mode, and so forth. The process will be terminated as soon as the computed RD cost is below a threshold that is content adaptive and is also dependent on the RD cost of the previous macroblocks. By adjusting the threshold we can choose a good tradeoff between time saving and peak-signal-to-noise ratio (PSNR). Experimental results show that the proposed scalable fast mode decision scheme can drastically reduce the encoding time up to 50% with negligible loss of coding efficiency.

1.4 Organization of the Thesis

The rest of this thesis is organized as follows. In Chapter 2, a literature review is presented on H.264, RDO, rate control methods for H.264, perceptual bit allocation methods for H.264, and fast mode decision methods for H.264. In Chapter 3, we propose a motion complexity measure to represent the complexity of motion contents. We also propose a minimum distortion based mode decision process. Based on the motion complexity measure and minimum distortion based mode decision process, we propose an efficient rate control method for H.264. In Chapter 4, we propose a perceptual mode decision algorithm. Based on the perceptual mode decision algorithm, we propose an efficient perceptual bit allocation method for H.264. In Chapter 5, we first study the statistical properties of macroblock coding modes in H.264. Based on the occurring probability of different coding modes, we then propose an efficient scalable fast mode decision for H.264. In Chapter 6, the contributions are summarized and the future work is outlined.

Chapter 2

Background

2.1 H.264 Video Coding Standard

Since the late 1990s after the well known video coding standards like H.263 and MPEG-4 developed, the ITU-T VCEG has started putting long-term effort to further development on a new standard for low bit rate visual communications. This long-term effort led to the draft H.26L standard, which provided significantly higher video compression efficiency than existing ITU-T standards. In 2001, the ISO MPEG recognized the potential benefits of H.26L and formed JVT which included experts from MPEG and VCEG. JVT's main task was to develop the draft H.26L model into a complete international standard. This new standard is called MPEG-4 Part 10 or advanced video coding (AVC), and by its ITU-T document number, H.264. In 2003, H.264 was ratified as recommendation by ITU-T, and AVC became international standard by ISO in the same year. However for consistency and simplicity, throughout this thesis we call it H.264.

H.264 came about as a result of an MPEG requirement for advanced video coding tools. It has a narrower scope than MPEG-4 Visual and was designed primarily to support efficient and robust coding and transport of rectangular video frames. Its original aim was to provide similar functionality to the existing video coding standards such as H.263+

and MPEG-4 Visual (Simple Profile) but with significantly better compression performance and improved support for reliable transmission. Target applications include two-way video communication (video conferencing or video telephony), television broadcasting, video streaming over packet networks, digital storage media, telemedicine systems, and digital cinema, etc. Support for robust transmission over networks is built in and the standard is designed to facilitate implementation on as wide a range of processor platforms as possible. The standard is specified in more detail than many existing video coding standards in an attempt to minimize the possibility for misinterpretation by developers [6]. It is not a fundamentally different method, but rather a significant refinement of well-established methods.

H.264 has a number of advantages that distinguish it from existing standards, while at the same time, sharing common features with existing standards. The following features are some of the key advantages of H.264 [7]:

- (1) Up to 50% in bit rate savings;
- (2) High quality video;
- (3) Error resilience;
- (4) Network friendliness.

Simulation results [8]-[11] have shown that H.264 has achieved substantial superiority of video quality over that achieved by H.263, MPEG-2, and MPEG-4. The JVT reference model software has achieved up to 50% in bit rate saving compared with existing most optimized H.263 or MPEG-4 codec. This means that H.264 offers significantly higher quality levels with the same bit rates. As for new features such as advanced inter and intra prediction, H.264 supports a number of error resilient techniques

that facilitate the codec to cope with different channel situations. These characteristics of H.264 make it an ideal codec for applications with very limited channel capacity and extremely error prone channels such as mobile systems and video telephony. Due to its high compression efficiency, the codec can be used for coding of high quality video at lower rates. H.264 is expected to replace previous standards like H.263 and MPEG-2 Visual quite soon [12]. Therefore, H.264 will be a serious contender for a variety of next generation multimedia applications. Moreover, another important fact is that H.264 is a public and open standard. Every manufacturer can build encoders and decoders in a competitive market. This will bring prices down quickly, making this technology affordable to everybody. There is no proprietary format, as on the Internet today, which is of utmost importance for the multimedia and communication community.

2.2 Rate-Distortion Optimization

Rate-distortion (RD) theory is a branch of information theory, which gives theoretical performance bounds for lossy compression schemes. The basic problem in RD theory is: what is the minimum bit rate required to encode a source in order to achieve a certain distortion? The RD theory provides a RD function $R(D)$ which has the following properties [13]:

- (1) For any given distortion D , it is possible to design a coding scheme which can obtain average distortion arbitrarily close to D with rate arbitrarily close to $R(D)$.

- (2) There is no practical compression system that can achieve distortion D or below D at a bit rate below $R(D)$.

Figure 2.1 shows the curve of a theoretic RD function which is called RD curve [4]. The RD curve is convex, continuous and strictly decreasing. However in practice, a real RD curve may not be strictly decreasing.

Figure 2.1. RD curve

By adding efficient coding options that are related to RD to a video codec, the overall RD performance may increase. The optimization task is to choose, for each image region, the most efficient coded representation (segmentation, prediction modes, motion vectors, quantization levels, etc.) in the RD sense. Such task is called rate-distortion optimization (RDO) in video coding. This task is complicated by the fact that the various coding options show varying efficiency at different bit rates and with different scene contents.

Rate control plays a crucial role in RDO. On the one hand, it regulates the coded bit stream to satisfy certain given conditions. On the other hand, it enhances the quality of

coded video. Without rate control any video coding scheme would be practically useless in many applications because the client buffer may often undergo buffer overflow or underflow. Therefore, every video coding standard has its own rate control technique.

The aim of rate control is to achieve a target bit rate of the coded bit stream. The coded bit stream includes bits for macroblock coding modes, motion vectors, QPs, residual coefficients, and so on. The bits for macroblock modes, motion vectors, QPs and so on are called header bits. The bits for residual coefficients are called coefficient bits. Header bits are controlled by selecting macroblock modes. When a macroblock mode is selected, header bits of the macroblock usually vary in a limited range. The control of coefficient bits is more accurate, flexible, and efficient than the control of header bits. Thus the control of coefficient bits is the main task in rate control. The following figure illustrates the formulation process of coefficient bit in the encoder of most of the standards.

Figure 2.2. Formulation process of coefficient bit

In Figure 2.2, the bit variation between residual coefficient and coefficient bit occurs at quantization module which is usually a scalar quantizer. The basic quantization operation is as follows:

$$Y = \text{round} \left(\frac{X}{Q_{Step}} \right) \quad (2.1)$$

where X is the transform coefficient, Q_{Step} is the quantization step size, and Y is the quantized coefficient. In real time implementation, there is a look-up table that lists the quantization step sizes and their corresponding indices, say QPs, for quantization. Table 2.1 is an example of such a table that is employed in H.264 [6]. According to Shannon's RD theory [4], the wide range of quantization step sizes makes it possible for an encoder to accurately and flexibly control the trade-off between bit rate and video quality. By selecting QP from the table, we can get its corresponding quantization step size and use it for quantization. Therefore, the implementation of rate control has been converted from controlling target bit to selecting suitable QP.

Table 2.1. Quantization step sizes in H.264 codec

QP	0	1	2	3	4	5	6	7	8	9	10	11	12	...
Q_{Step}	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.375	1.625	1.75	2	2.25	2.5	...
QP	...	18	...	24	...	30	...	36	...	42	...	48	...	51
Q_{Step}	...	5	...	10	...	20	...	40	...	80	...	160	...	224

From the bit rate point of view, there are two kinds of rate control: constant bit rate (CBR) control and variable bit rate (VBR) control. Figure 2.3 illustrates the video quality represented by PSNR and bit rate of CBR and VBR respectively. In this thesis, we only consider CBR control. In CBR, the variable bit rate produced from encoder can be smoothed by buffering the encoded bit stream prior to transmission. Figure 2.4 shows a typical arrangement of CBR control.

The simplest way of rate control is to feedback output buffer parameters and current bit rate to a rate control module, namely rate controller, to determine QP that is used for

quantization. Figure 2.5 illustrates the relationship among rate controller, encoder and output buffer.

Figure 2.3. CBR and VBR

Figure 2.4. Encoder output buffer and decoder input buffer

Figure 2.5. Rate controller

2.3 H.264 Rate Control

Rate control techniques have been studied very intensively for various standards and applications. Compared with other existing video coding standards, the uniqueness of H.264 rate control comes from the following aspects:

- (1) Macroblock coding mode;
- (2) RDO.

Similar to other video coding standards, only the bit stream syntax and the decoding procedure are specified in H.264, and the encoder is flexible to have different implementations. The encoder must select certain parameters such as macroblock modes, motion vectors, QP and residual data for each macroblock among several choices. For interframes in H.264, there are more macroblock modes than in previous standards. These modes are: SKIP, INTER 16×16 , INTER 16×8 , INTER 8×16 , INTER 8×8 , INTRA 16×16 , INTRA 4×4 . In INTER 8×8 mode, each block can be further divided independently into 8×8 , 8×4 , 4×8 , or 4×4 sub-partitions. The more number of modes is one of the novel features of H.264 so that the temporal and spatial details in a macroblock are best presented. Figure 2.6 shows the different macroblock modes in H.264. This large amount of choices makes optimization of the encoder very crucial to achieve good RD efficiency. In other words, a poor design of encoder can generate even lower quality bit streams than a simple traditional video encoder. To select the best mode, RDO is employed so that for each macroblock, all the macroblock modes are tried and the one that leads to the least RD cost is selected. This is to achieve the best trade-off of the rate

and distortion performance. In the JVT reference model software, the Lagrangian multiplier method is applied to achieve RDO.

Figure 2.6. Macroblock coding modes in H.264

The Lagrangian multiplier method is based on converting a constraint optimization problem to an unconstrained one [4]. The constraint optimization problem is to minimize the distortion D with a constraint that the rate R should be less than a rate constraint R_c , which can be expressed in the following equation:

$$\min D: R < R_c \quad (2.2)$$

Using the Lagrangian multiplier method, the above problem is converted to minimize the Lagrangian cost J , with λ being the Lagrangian multiplier:

$$\min J: J = D + \lambda \times R \quad (2.3)$$

The selection of encoding parameters for every macroblock i in a picture determines its rate R_i and distortion D_i , and the sum of these values generates R and D . Assuming that a picture is composed of regular macroblocks, then the optimization of (2.3) is simplified to separately minimizing the cost of each macroblock:

$$\min J_i : J_i = D_i + \lambda_i \times R_i \quad (2.4)$$

In H.264, motion estimation and mode selection have been optimized using the Lagrangian method. The motion estimation process finds a motion vector that minimizes motion vector cost J_{MV} which is calculated by using the bits needed to send each motion vector R_{MV} and the relating sum of absolute differences (SAD) SAD_{MV} :

$$J_{MV} = SAD_{MV} + \lambda_{Motion} \times R_{MV} \quad (2.5)$$

After determining the optimal motion vectors for inter modes, the Lagrangian cost of each mode J_M is calculated by using the required number of bits to send all macroblock information R_M and the sum of squared differences (SSD) between the original and the reconstructed pixels SSD_M :

$$J_M = SSD_M + \lambda_{Mode} \times R_M \quad (2.6)$$

The mode that has the minimum J_M among all the others is selected as the optimum one.

In the above optimization process, the value of λ is calculated using empirical equations [11]:

$$\lambda_{Mode} = 0.85 \times 2^{(QP-12)/3} \quad (2.7)$$

$$\lambda_{Motion} = \sqrt{\lambda_{Mode}} \quad (2.8)$$

Therefore rate control is conducted via controlling for instance the QP and adjusting the Lagrangian multipliers accordingly using Equation (2.7) and (2.8). Unfortunately, employing RDO in mode decision makes rate control a difficult task. This is because that in order to perform RDO for a macroblock, a QP should be firstly determined for this macroblock depending on the motion compensated residual information after motion compensation. However, such residual information is only available after performing

RDO. This is typically a dilemma and we call it QP dilemma, which imposes a big problem in H.264 rate control. For this reason, it is difficult to adopt existing rate control algorithms of other video coding standards directly to H.264 video coding.

As in other video coding standards, rate control remains an open issue for H.264 standard. This has motivated a lot of researchers focusing on this issue and some achievements have been obtained in the following areas.

- (1) Bit-QP model: The purpose of bit-QP model is to represent the relationship among bits, distortion and QP, and to estimate QP by bit-QP model. In [14], a linear bit-QP model was proposed. In [15], a logarithmic model was proposed. Kim and Ho [16] employed a logarithmic model. Wang and Leou [17] proposed both a quadratic model and a logarithmic model to compute QP and selected QP with the nearer one in a quantization array. Tsai and Leou [18] proposed a quadratic model, a logarithmic model and an exponential model to compute QP separately, and selected QP with the nearest one in a quantization array.
- (2) RD model: The purpose of RD model is to represent the relationship between rate and distortion, and to estimate bits by using RD model. Milani *et al.* [19] employed a model that is called ρ -model deriving from [20] to represent the RD relationship, where ρ is the percentage of null quantized coefficients in each block. Shin *et al.* [21] also proposed a similar ρ -model.
- (3) Bit estimation: Navakitkanok *et al.* [22] proposed a bit estimation scheme to maintain the buffer fullness slowly changing. Jiang *et al.* [23]-[28] introduced

MAD ratio as a measure for global frame encoding complexity, and allocated bits according to MAD ratio to improve video quality at high motions.

- (4) QP estimation: Li *et al.* [29]-[32] used a linear model to predict MAD and QP. Their proposal was accepted as the H.264 proposal. Ma *et al.* [33]-[37] predicted QP by a scheme which was derived from MPEG-2 test model 5 (TM5) [38].

2.4 H.264 Perceptual Bit Allocation

The problem of rate control is identical to the problem of optimal bit allocation. We are given a set of M source signals $S_1, S_2, \dots, S_i, \dots, S_M$ such that each source signal S_i has a set of N_i possible reconstructed signals. Quantizer Q_{ij} produces the j th reconstructed signal of source signal S_i with distortion D_{ij} and rate R_{ij} . If we use quantizer Q_{ij} on S_i , we set indicator variable $X_{ij}=1$ and otherwise set $X_{ij}=0$. Given a rate budget R , the goal of rate control is to minimize the overall distortion D . Thus the rate control problem can be formulated as the following equations:

$$\text{minimize} \quad D = \sum_{i=1}^M \sum_{j=1}^{N_i} X_{ij} D_{ij} \quad (2.9)$$

$$\text{subject to} \quad \sum_{i=1}^M \sum_{j=1}^{N_i} X_{ij} R_{ij} \leq R \quad (2.10)$$

$$\sum_{j=1}^{N_i} X_{ij} = 1, \quad i = 1, \dots, M \quad (2.11)$$

$$X_{ij} \in \{0, 1\}, \quad \forall i, j \quad (2.12)$$

The value of the set of X_{ij} gives a solution to the rate control problem: the reconstructed signal is produced by quantizer Q_{ij} if and only if $X_{ij}=1$. As shown in [39], the rate control problem has the equivalent formulation with the multiple-choice knapsack problem. Since the multiple-choice knapsack problem is well-known to be NP-hard [40], the rate control problem is also NP-hard. Thus it is unlikely that it is optimally solvable in polynomial time, and our task is to present a near optimal solution for the problem.

Human vision is sensitive to movement of well structured objects while tolerant to large distortion in moving areas with random structure (texture wise and motion wise) [41]. This implies that structured motion region is the perceptually important region. The goal of perceptual bit allocation is to achieve the best perceptual video quality at a given bit rate through a proper bit allocation process.

Depending on the continuity of interested image regions, perceptual bit allocation can be classified as global perceptual bit allocation and region of interest (ROI) bit allocation. The difference between ROI and global perceptual contents is the continuity and number of interested regions. If a ROI is manually assigned, its number is usually equal to one. If a ROI is obtained by automatically detecting the boundary of an object, it is usually the region surrounded by the boundary. For global perceptual contents, they are mostly random and the number of perceptual contents is much more than ROI. Simply speaking, perceptually important region in global perceptual bit allocation is all around the image while it is only a small part of the image in ROI bit allocation.

- (1) Global perceptual bit allocation: In the existing H.264 rate control methods, mean absolute difference (MAD) is selected to estimate QP and further to

decide coding modes. However, it is demonstrated that the minimum MAD does not translate into minimum perceptual distortion [42]. Simple computational distortion measures, such as the MAD, cannot reliably predict the difference of the perceived quality of one image with another [43]. Thus a better approach is to use the perceptual characteristics of the video contents. Sengupta *et al.* segmented an image into foreground part and background part [44]. For different part of the image, they designed different bit allocation strategy. In [45], a video coding approach performing adaptive RDO guided by perceptual hints was proposed. The key idea is to adaptively adjust the Lagrange multipliers of the RDO coder control module based on visual attention analysis. Tang *et al.* proposed a perceptual bit allocation scheme by adopting a visual distortion sensitivity model [46]. The proposed automatic distortion sensitivity analysis process analyzes both the motion and texture structures in the video sequences in order to evaluate the tolerable perceptual distortions and allocate fewer bits to regions permitting large perceptual distortions for rate reduction. In [47], a Lagrangian optimized rate control algorithm was proposed for the H.264 video encoder. The algorithm controls the bit rate by adjusting the Lagrangian multiplier adaptively for every picture and specifying the QP for every MB. The success of the above perceptual algorithms depends largely on the good estimation of visual features and thus a more accurate visual important map needs to be established.

- (2) ROI bit allocation: ROI is a selected subset of image pixels identified for a particular purpose. It can be determined by manually assigning a geographic

shaped region in the image or automatically detecting the boundary of an object. Pietrowcew *et al.* proposed a ROI bit allocation algorithm based on the ρ -domain model [48]. The pixels inside ROI are allocated bits according to their zero quantized coefficients. Agraftotis *et al.* designed a priority factor to specify the importance of a macroblock according to whether the macroblock belongs to ROI or not [49]. Similar to Agraftotis's method, Liu *et al.* proposed a method that allocates bits by a weighting factor incorporating the macroblock's importance [50]. The weighting factor is related to the macroblock's MAD which is used to detect ROI. All the above methods allocate more bits inside ROI.

2.5 H.264 Fast Mode Decision

In the JVT reference model software, in order to choose the best macroblock coding modes, H.264/AVC makes use of full search RDO, which is computationally expensive, and a Lagrangian multiplier method is used to achieve RDO. Figure 1.1 shows the procedure to achieve RDO using full search scheme and the position of mode decision in the encoder. It can be seen that mode decision is performed after motion estimation/compensation and entropy coding. The mode decision result controls the actual output stream to be transmitted. The detailed steps of this exhaustively full search RDO are as follows.

- (1) Perform motion estimation for all the inter modes.
- (2) Compute RD cost of all the coding modes including SKIP mode, inter modes and intra modes. The RD cost of each mode J is calculated by using the

number of bits R consumed by this macroblock and the SSD between the original and the reconstructed pixels, as shown in Equation (2.6).

- (3) The coding mode that has the minimum J is selected as the best coding mode for this macroblock.

It can be seen from the above steps that the mode decision strategy of the full search RDO scheme is “parallel” such that RD costs of all coding modes for an MB must be calculated before a decision can be made, as shown in Figure 2.7. However, it is not necessary to test all the coding modes if the mode can be decided earlier by using the local content information of the video. For example, if the video objects contain many detailed texture or high motion, the probability of coding it using small partition such as INTER8×8 mode is much higher than that of using a larger partition, and vice versa. Once a decision is made prior to testing all the modes, processing time will be saved. Such mode decision method that does not need to test all the modes is called fast mode decision.

Figure 2.7. Parallel mode decision scheme of RDO

It should be noted that the decision result of fast mode decision should be feedback to motion estimation and motion compensation to stop further processing. Thus compared

with Figure 1.1, the encoder structure with fast mode decision must be modified. Such modification of encoder structure is illustrated in Figure 2.8.

Figure 2.8. Modified encoder with fast mode decision

Depending on the types of mode to be checked, fast mode decision method can be classified as fast SKIP mode decision, fast intra mode decision, fast inter mode decision, and fast hybrid mode decision.

- (1) Fast SKIP mode decision: Fast SKIP mode decision method is used to check SKIP mode. If the mode is determined to be SKIP mode, no other mode testing is needed. Jeon and Lee proposed a early SKIP mode decision method by checking the motion vector and transform coefficients whether they are all equal to zero [51]. Choi *et al.* [52] provided a similar method like the above. Wei and Ngan proposed an algorithm to calculate the transform coefficient cost according to its level [53]. The SKIP mode decision is made by comparing the coefficient cost with a threshold. Saha *et al.* proposed a fast

SKIP mode decision method based on transform domain properties and spatio-temporal correlation in video sequences [54]. If the number of zeros in the sum of absolute transform differences (SATD) of a block and its neighboring blocks is smaller than a predefined threshold, SKIP mode is determined. All the above fast SKIP mode methods satisfy the following conditions:

- the block size of best mode is 16×16 ;
- the motion vector is (0, 0) or the same as its predicted motion vector;
- the transform coefficients are all quantized to zero.

(2) Fast intra mode decision: Fast intra mode decision is used to reduce the computation time of intra mode decision. Pan *et al.* [55][56] proposed a fast intra mode decision method based on local edge directional information to reduce the amount of calculations in intra prediction. With the use of edge of direction histogram derived from the edge map of the image, only a small number of most possible intra modes are chosen for RDO calculation. Tseng *et al.* proposed a fast Hadamard transform algorithm to reduce computation time for INTRA 4×4 mode decision [57]. Wang *et al.* proposed a regular spatial domain filtering technique to compute the dominant edge strength in order to reduce the selection of all directions [58]. Kim proposed a method of comparing the RD cost of best inter mode of current macroblock and the RD costs of neighboring macroblocks' best modes to decide whether to skip intra mode checking [59]. These fast mode decision methods reduce intra mode decision time by:

- skipping intra mode decision procedure;
- fast transformation;
- reducing the number of intra mode prediction directions;
- early terminating the search time of each intra mode prediction direction.

(3) Fast inter mode decision: Fast inter mode decision is used to reduce the computation time of inter mode decision. Lim and Wu *et al.* [60][61] proposed a fast inter mode decision method that makes use of the spatial homogeneity and the temporal stationarity characteristics of video objects. If the block with larger size is homogeneous, skip the other modes with smaller sizes. Kuo *et al.* proposed a fast mode decision method by fast motion search [62]. With early termination by comparing the estimated bits for residual coding with a predefined threshold, fast motion search is performed and thus the inter mode decision time is reduced. Zhou *et al.* also proposed a fast inter mode decision method by fast motion search [63]. Fast motion search is performed by comparing motion cost with a predefined threshold. Yu *et al.* proposed a hierarchical structure of three levels to implement fast inter mode decision [64]. Early termination of mode decision is triggered at any of the levels by checking the spatial complexity. These fast inter mode decision methods reduce inter mode decision time by:

- fast motion estimation;
- checking the characteristics of video objects;
- checking the spatial correlation between neighboring macroblocks.

(4) Fast hybrid mode decision: Fast hybrid mode decision is the combination of fast SKIP mode, fast intra mode and fast inter mode decisions. Grecos and Yang implemented fast mode decision by checking the RD cost of the subsets of whole modes [65][66]. Yu *et al.* proposed a fast intra mode decision algorithm by checking the transform coefficients and a fast inter mode decision algorithm by checking a complexity measure to identify those low detailed macroblocks [67]. Kim and Kuo proposed a feature based fast intra and inter mode decision method [68]. Features are extracted from the macroblock to form a feature space which is partitioned into several risk regions. The likelihood function of risk is calculated to determine which mode should be discarded. Kim and Kim made several early termination conditions for different modes which are based on block correlation [69]. If one of the conditions is satisfied, the corresponding mode is selected and the other modes are skipped.

Chapter 3

Content Adaptive Rate Control Scheme for H.264

3.1 Introduction

It is well known that the QP dilemma problem makes rate control methods for H.264 quite different from the existing rate control methods for other video coding standards. To solve the QP dilemma, all the existing H.264 rate control methods estimate QP or mean absolute difference (MAD) by using the information of previous macroblocks. For example, Ma *et al.* [33]-[37] proposed a scheme that was derived from MPEG-2 TM5 to estimate QP from neighboring macroblocks. Since the information is estimated and is not directly related to the current macroblock, the estimated QP is not accurate. In [29]-[32], Li *et al.* used a linear model to predict the MAD of current basic unit (a basic unit can be one or more than one macroblock) in the current frame by using a co-sited basic unit in the previous frame, and the QP dilemma was thus resolved. It is noted that in this approach, the target bit is estimated solely based on the buffer fullness, regardless of the frame's content. If the buffer is nearly full, fewer target bits would be allocated to a new frame. This may lead to a drastic drop in PSNR, especially in the case of high motion scenes or scene changes. To improve the video quality at scene changes, Jiang *et al.* [23]-[28] introduced MAD ratio as a measure of motion complexity. In their approach, bit

budget was allocated to frames according to their MAD ratio and the buffer status. However, MAD ratio is not a good way of representing the motion contents, as it can only represent the similarity between the current frame and its reference frame.

In this chapter, we focus on improving the video quality at scene changes and high motions. We estimate QP according to the minimum distortion after performing motion estimation of all the modes. We define a new motion complexity measure to represent the amount of motion contents between two consecutive frames by using the number of bits consumed by the previous frames. Based on motion complexity and full mode motion estimation process, we can estimate the target bit and QP accurately, and thus greatly improve the objective and subjective qualities.

3.2 Mode Decision Process Based upon Minimum Distortion

To solve the QP dilemma, the existing H.264 rate control methods estimate QP or MAD before motion estimation. In [31], a linear model is used to estimate MAD from the MAD of the co-located macroblock in the previous frame, and then the estimated MAD is used to predict QP. In [36], QP is estimated from the neighboring macroblock's activity. The following figure illustrates the mode decision process of the existing H.264 rate control methods.

Figure 3.1. Mode decision process of the H.264 rate control algorithms

To achieve a good video quality, QP should be estimated accurately so that the bits allocated to the individual macroblocks are in accordance with their actual needs. The QP estimation usually utilizes the information of the neighboring macroblocks or the macroblocks at the same place of the previous frame. This estimated QP is then used both in quantizing the motion compensated residual information after motion estimation and calculating the Lagrangian multiplier for RDO. Since the information of current macroblock is not used in QP estimation, the estimated QP is not accurate enough. This will affect the performance of RDO, and consequently the quality of the reconstructed video. To estimate QP accurately, we employ two times of QP estimation. First, we estimate QP with the traditional method and use it in motion estimation. After a full-search motion estimation of all the modes, the minimum MAD of all the modes is obtained and is used to further estimate QP. We call this new mode decision process minimum distortion based mode decision. The minimum distortion based mode decision process is illustrated in Figure 3.2.

Figure 3.2. Minimum distortion based mode decision process

Figure 3.3 shows the flow chart of minimum distortion based mode decision process, which can be summarized step by step as follows:

- (1) Use a pre-determined QP to compute the motion vector Lagrangian multiplier λ_{Motion} in Equation (2.8). Here we determine the QP with the traditional estimation method.

- (2) Use λ_{Motion} to compute the motion vector cost J_{MV} in Equation (2.5).
- (3) Perform motion estimation of all the macroblock modes, find the one which has the minimum distortion.
- (4) Use the minimum distortion to compute QP according to the quadratic model.
- (5) Compute the RD Lagrangian multiplier λ_{Mode} in Equation (2.7).
- (6) Use λ_{Mode} to compute the RD cost J_M in Equation (2.6).
- (7) Select the best mode with the guide of J_M .

Figure 3.3. Flow chart of minimum distortion based mode decision process

In the above minimum distortion based mode decision process, one important step different from all the existing rate control methods is to get the minimum distortion

among all the macroblock modes, which is done after performing all-mode motion estimation. The advantage of this method is that the information of current macroblock can be utilized sufficiently. With this minimum distortion, QP of the macroblock is then estimated. Since this QP estimation utilizes the macroblock's information, QP should be more accurate than those obtained without current macroblock's information. Moreover, since the calculation of updated QP estimation occupies very little time, the cost of performing further QP estimation will be negligible.

3.3 Content Adaptive Motion Complexity Measure

The purpose of rate control is to produce high quality video at a given bit rate by appropriately allocating bits to residual information, motion vectors, and some other parameters. In order to achieve consistent video quality, bits are usually allocated according to the motion complexity. The motion complexity refers to the moving picture contents of two consecutive frames in a video sequence. If the objects in a frame are stationary, or a frame can be well matched to the previous frame with very little difference, then such a frame is a low motion frame. If there is a scene change between two consecutive frames, or a moving object is unable to find a good match in the previous frame, then such a frame is considered as a high motion frame.

Figure 3.4 shows the bit allocation of the sequence “Foreman” with fixed QP=28. From this figure, we can find that the bit allocation is in accordance with the complexity of frame's motion contents. For example, from the 253rd frame to the 254th frame, high motion happens at the region near the man's eyes and hand. Figure 3.5 shows the pictures of the 253rd frame and the 254th frame. The number of bits allocated to the 254th frame

is 39% more than that allocated to the 253rd frame. This shows that the number of allocated bits is a good measure to represent the motion complexity of a frame.

Figure 3.4. Bit allocation of sequence “Foreman”

(a) 253rd frame

(b) 254th frame

Figure 3.5. Two consecutive frames with high motion in sequence “Foreman”

However, it is difficult to obtain the actual bits consumed before entropy encoding. What we can do is only to estimate the target bits allocated to the current frame. Usually there is a difference between the estimated bits and the actually consumed bits. The estimated target bits should represent the motion complexity of current frame. Since there is a temporal correlation between consecutive frames, we can use the actual consumed bits of the previous coded frames to estimate the current frame’s motion complexity, and use it to estimate the target bits. Here we propose a new measure to represent the motion

complexity of a frame. In this measure, motion complexity depends on the bits allocated to the previously encoded frames. Let $B_{p,i}$ be the predicted bits at frame i , B_j ($j=1,2,\dots,i$) be the actual allocated bits to the previously encoded frames, then motion complexity at frame i is:

$$C_i = \frac{B_{p,i}}{\frac{1}{i-1} \sum_{j=1}^{i-1} B_j} \quad (3.1)$$

where $B_{p,i}$ is the linear prediction of the actual previous frame's bits:

$$B_{p,i} = \alpha_i B_{i-1} \quad (3.2)$$

It should be noticed that α_i needs to be updated after bits have actually been allocated. It depends upon the actual consumed bits by the current frame and the previous frame. Because of this reason, the coefficient α_i can only be updated after the second frame. Thus the update of α_i can be expressed as follows:

$$\alpha_{i+1} = \frac{B_i}{B_{i-1}} \times \alpha_i, \quad i \geq 2 \quad (3.3)$$

For the first and second frames, the actual consumed bits cannot be obtained before encoding. Thus $\alpha_1 = \alpha_2$ is assigned a predefined value.

From Equations (3.2) and (3.3), the motion complexity in Equation (3.1) can also be derived based only on the actual consumed bits of the previous frames as follows:

$$C_i = \frac{B_{p,i}}{\frac{1}{i-1} \sum_{j=1}^{i-1} B_j} = \frac{\alpha_i B_{i-1}}{\frac{1}{i-1} \sum_{j=1}^{i-1} B_j} = \frac{\left(\alpha_{i-1} \frac{B_{i-1}}{B_{i-2}} \right) B_{i-1}}{\frac{1}{i-1} \sum_{j=1}^{i-1} B_j} = \dots = \frac{\alpha_2 \frac{B_{i-1}^2}{B_1}}{\frac{1}{i-1} \sum_{j=1}^{i-1} B_j} \quad (3.4)$$

It can be seen from the above new expression of the motion complexity that although the motion complexity is predicted from the previous frames, it is not a simple linear prediction.

From Equations (3.1) and (3.4), it is seen that the computation of motion complexity is simple. It is only related to the actual consumed bits of the previous frames. By employing motion complexity, we can make clear the complexity of current frame's motion contents. This will help us accurately estimate the target bits in rate control.

Figure 3.6. Motion complexity and normalized bits of sequence “Foreman”

We illustrate the motion complexity curve in Figure 3.6. Comparing Figure 3.4 with Figure 3.6(a), we can find that the shape of motion complexity curve is almost the same as the shape of bits curve. This means that motion complexity is also a good measure for the complexity of motion contents. To quantitatively analyze the relationship between bits and motion complexity, we normalize them and combine the normalized results into one figure in Figure 3.6(b). As can be seen from Figure 3.4 and Figure 3.6(a), a point on

the bits curve or motion complexity curve has four types of status: ascent, descent, locally maximum, and locally minimum. For convenience of defining the similarity between bits and motion complexity to be discussed in the following, we assign a unique integer value to each status, that is, we assign 1, 2, 3, and 4 to the status of ascent, descent, locally maximum, and locally minimum, respectively. Assume g_i is the number of allocated bits or motion complexity of the i th frame, then the status of the i th frame on the bits curve or motion complexity curve is defined as follows:

$$s_i = \begin{cases} 1 & \text{ascent, that is, } g_{i-1} \leq g_i \leq g_{i+1} \\ 2 & \text{descent, that is, } g_{i-1} \geq g_i \geq g_{i+1} \\ 3 & \text{locally maximum, that is, } g_{i-1} \leq g_i, g_i \geq g_{i+1} \\ 4 & \text{locally minimum, that is, } g_{i-1} \geq g_i, g_i \leq g_{i+1} \end{cases} \quad (3.5)$$

For each frame, we define two points on the motion complexity curve and the bits curve as similar if they have the same status. Thus for a video sequence, the similarity between bits and motion complexity is defined as the ratio of the number of similar points to the number of total points as follows:

$$r_s = \frac{n_s}{N} = \frac{\sum_{i=1}^N f(s_{M,i} - s_{B,i})}{N} \quad (3.6)$$

where r_s is the similarity, n_s is the number of similar points, N is the number of total points, $s_{M,i} \in \{1, 2, 3, 4\}$ is the value of the status of the i th point on the motion complexity curve, $s_{B,i} \in \{1, 2, 3, 4\}$ is the value of the status of the i th point on the bits curve, f is a function defined as follows:

$$f(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases} \quad (3.7)$$

(a) Sequence “Children”

(b) Sequence “Coastguard”

(c) Sequence “Grandma”

(d) Sequence “Weather”

Figure 3.7. Normalized bits and motion complexity

Using the above defined similarity between bits and motion complexity, we have tested the proposed method over several video sequences with fixed $QP = 28$. The results on these test video sequences are given in Table 3.1. As can be seen from Table 3.1, the similarity for most of the test video sequences is greater than 86%. For the convenience of the reader, we have depicted the normalized bits and motion complexity of video sequence “Children”, “Coastguard”, “Grandma”, and “Weather” in Figure 3.7. As can be seen from Figure 3.7, the shapes of the bits curve and the motion complexity curve are

quite similar for each sequence, including the sequence of “Children” which has the lowest similarity of 79.93%. Therefore it is reasonable to conclude that the motion complexity is a good measure to represent the allocated bits.

Table 3.1. Similarity between bits and motion complexity

Sequence	Similarity (%)
Carphone	91.30
Children	79.93
Coastguard	97.66
Football	88.37
Foreman	97.66
Grandma	90.97
News	86.63
Silent	87.63
Suzie	89.26
Weather	88.29

3.4 Minimum Distortion Based QP Estimation Rate Control

Based on the minimum distortion based mode decision process and motion complexity measure, we propose a rate control scheme. This scheme includes two levels of rate control: the frame-level rate control and the macroblock-level rate control. At the frame-level rate control, we estimate the target bits of current frame according to motion complexity. The first QP of a frame is estimated from the previous frame. At the macroblock-level rate control, we estimate QP in the mode decision process, according to

the minimum distortion after all-mode motion estimation. The target bits of current macroblock is estimated equally. The flow chart of the scheme is shown in Figure 3.8.

Figure 3.8. Flow chart of proposed rate control scheme

At the encoder, all the encoded bits are put into an output buffer for transmission. The output buffer fullness should be considered in target bits estimation so that the buffer is maintained neither overflow nor underflow. The output buffer can be illustrated with a fluid model as shown in Figure 3.9.

Figure 3.9. Buffer model

In this model, R_b is the bit rate (suppose it is a constant), R_f is the frame rate, F_i is the buffer fullness at time i , B_i is the bits of frame i . Then at time $i+1$, the buffer fullness is:

$$F_{i+1} = F_i + B_i - \frac{R_b}{R_f} \quad (3.8)$$

In order to avoid buffer overflow, B_i should be determined according to the following constraint:

$$B_i \leq F_s - F_i + \frac{R_b}{R_f} \quad (3.9)$$

where F_s is the buffer size. This buffer model will be used in estimating target bits.

At the frame-level rate control, target bits estimation includes the total number of bits of a group of pictures (GOP), the remaining bits of a GOP, and the target bits of the current frame. The total number of bits of a GOP is determined by the following equation:

$$B = N \times \frac{R_b}{R_f} \quad (3.10)$$

where B is the total number of bits of a GOP, N is the number of frames in a GOP. The bits of a GOP when encoding frame i is:

$$B_{R,i} = \begin{cases} B & i = 1 \\ B_{R,i-1} - B_{i-1} & i = 2, 3, \dots, N \end{cases} \quad (3.11)$$

The bits consumed by the first frame, that is, the I (intra) frame, will not be estimated but is obtained actually according to the initial QP. The bit allocation for other frames will be estimated according to their motion complexity, the remaining bits of the GOP, buffer size, buffer fullness, bit rate, and frame rate. The number of target bits of the i th frame is:

$$B_{T,i} = \frac{C_i \times B_{R,i} + \beta_i \times (F_s - F_i + \frac{R_B}{R_F})}{N - i + 1} \quad i = 2, 3, \dots, N \quad (3.12)$$

where C_i is the motion complexity at frame i that is defined in Equation (3.1), and β_i ($0 \leq \beta_i \leq 1$) is a coefficient to control buffer fullness. In order to avoid buffer overflow, $B_{T,i}$ should be determined according to the following constraint:

$$B_{T,i} \leq F_s - F_i + \frac{R_B}{R_F} \quad (3.13)$$

Here motion complexity is employed to estimate target bit. From Equation (3.12) it can be seen that the estimated target bit is proportional to motion complexity. Thus this scheme ensures that more bits will be allocated to high motion frames, and vice versa. It is worth pointing out here that although the predicted bits at frame i , that is, $B_{p,i}$ is the linear prediction of the actual consumed bits of the previous frame in Equation (3.2), the number of target bits $B_{T,i}$ is not simply linear predictable from the previous frames. The reason is the same as we have revealed in Equation (3.4).

At the macroblock-level rate control, the number of target bits of the macroblock is estimated according to the total number of bits allocated to the frame. Suppose the distribution of coefficient bits and header bits is steady in a frame, then we have:

$$\frac{b_{C,j}}{b_{H,j}} = \frac{\sum_{k=1}^{j-1} b_{C,k}}{\sum_{k=1}^{j-1} b_{H,k}} \quad (3.14)$$

where $b_{C,j}$ is the coefficient bits of macroblock j , $b_{H,j}$ is the header bits of macroblock j .

Let n be the number of macroblocks in a frame, then the total number of bits of macroblock j can be estimated as:

$$b_{C,j} + b_{H,j} = \frac{B_{T,i} - \sum_{k=1}^{j-1} b_{C,k} - \sum_{k=1}^{j-1} b_{H,k}}{n - j + 1} \quad (3.15)$$

Combining Equation (3.14) and (3.15), we obtain the estimated coefficient bits of macroblock j as follows:

$$b_{C,j} = \frac{\sum_{k=1}^{j-1} b_{C,k} \times \left(B_{T,i} - \sum_{k=1}^{j-1} b_{C,k} - \sum_{k=1}^{j-1} b_{H,k} \right)}{(n - j + 1) \times \left(\sum_{k=1}^{j-1} b_{C,k} + \sum_{k=1}^{j-1} b_{H,k} \right)} \quad (3.16)$$

In the mode decision process, QP is computed according to the minimum distortion among all the possible modes. Assuming that the source statistics is Laplacian distribution, a quadratic QP model which is derived from [70][71] is employed to estimate current macroblock's QP, that is:

$$\frac{b_{C,j}}{MAD_j} = \frac{X_{1,j}}{Q_j} + \frac{X_{2,j}}{Q_j^2} \quad (3.17)$$

where MAD_j is the MAD of the macroblock j , Q_j is the quantization step of the macroblock j which can be looked up according to Table 2.1, $X_{1,j}$ and $X_{2,j}$ are two coefficients of the quadratic model. Denote $\frac{b_{C,j}}{MAD_j}$ as R_j . That is:

$$R_j = \frac{b_{C,j}}{MAD_j} \quad (3.18)$$

Then we can calculate the coefficients $X_{1,j}$ and $X_{2,j}$ of the current macroblock according to the actual allocated bits, MAD and quantization step of the previous frames as shown as follows:

$$X_{2,j} = \frac{(j-1) \times \sum_{k=1}^{j-1} R_k - \left(\sum_{k=1}^{j-1} \frac{1}{Q_k} \right) \times \sum_{k=1}^{j-1} (Q_k R_k)}{(j-1) \times \sum_{k=1}^{j-1} \frac{1}{Q_k^2} - \left(\sum_{k=1}^{j-1} \frac{1}{Q_k} \right)^2} \quad (3.19)$$

$$X_{1,j} = \frac{\sum_{k=1}^{j-1} \left(Q_k R_k - \frac{X_{2,j}}{Q_k} \right)}{j-1} \quad (3.20)$$

By replacing MAD with the minimum distortion achieved from the proposed minimum distortion based mode decision process, we will obtain the estimated quantization step. Finally the estimated QP of the current macroblock can also be obtained by looking up Table 2.1.

3.5 Experimental Results

To evaluate the performance of our rate control scheme, we compare it with JVT-H014 [31] which is the existing H.264 rate control proposal. Since JVT-H014 is based on H.264 reference model version JM6.1e [72], we implement our rate control scheme with the same model.

The conditions of the experiment are listed in Table 3.2. The specifications of the test sequences are listed in Table 3.3. In our experiments, we only consider the features enabled in the Main Profile of H.264. In the experiments, in order to determine the target bit rate, we firstly encode the test sequences by using a fixed QP to find out the actual bits it consumes, then we use it as the target bit rate in the rate control scheme. For every video sequence, we select four QPs which are 28, 32, 36 and 40 respectively. For each

QP, the target bit rate is determined by encoding the sequence using the fixed QP without rate control. Table 3.4 lists the target bit rate for different video sequences. We implement the encoder with both JVT-H014 rate control proposal and our rate control scheme respectively.

Table 3.2. Test conditions

Frame type	IPPP...
Frame rate	30 fps
Slice mode	OFF
RDO	ON
Hardmard	OFF
Search range	32
Restrict search range	No restriction
Symbol mode	CABAC
Partition mode	No data partition
Out File mode	Annex B

Table 3.3. Specifications of the test sequences

Sequence	Frame size	Number of frames	Format
Carphone	QCIF	300	4:2:0 YUV
Children	QCIF	300	4:2:0 YUV
Coastguard	QCIF	300	4:2:0 YUV
Football	QCIF	130	4:2:0 YUV
Foreman	QCIF	300	4:2:0 YUV
Grandma	QCIF	300	4:2:0 YUV
News	QCIF	300	4:2:0 YUV
Silent	QCIF	300	4:2:0 YUV
Suzie	QCIF	150	4:2:0 YUV
Weather	QCIF	300	4:2:0 YUV

Table 3.4. Target bit rate

Sequence	Target bit rate (kbps)	Initial QP	Sequence	Target bit rate (kbps)	Initial QP
Carphone	23.34	40	Children	49.65	40
	42.22	36		85.39	36
	79.08	32		143.68	32
	147.56	28		227.19	28
Coastguard	21.11	40	Football	126.57	40
	41.74	36		215.31	36
	88.35	32		362.10	32
	192.27	28		582.74	28
Foreman	27.04	40	Grandma	5.55	40
	43.85	36		8.73	36
	74.26	32		15.48	32
	135.13	28		31.21	28
News	15.51	40	Silent	15.60	40
	25.51	36		27.13	36
	42.30	32		46.69	32
	71.15	28		79.72	28
Suzie	15.03	40	Weather	18.33	40
	23.73	36		28.82	36
	39.54	32		47.11	32
	72.41	28		78.40	28

The video quality should be evaluated in terms of the differences between the original pictures and the reconstructed pictures. Like most of the video compression methods, we employ the peak signal-to-noise ratio (PSNR) to evaluate the video quality. It is defined as the follows:

$$PSNR = 10 \times \log \frac{255^2}{MSE} \quad (3.21)$$

where 255 is the maximum intensity value (that is, the peak value) of a pixel for an 8-bit picture, MSE is the mean square error (MSE) between the original picture and the reconstructed picture. In the case of a picture $f(m, n)$ with dimension $M \times N$ and its corresponding reconstructed picture $\hat{f}(m, n)$, where (m, n) denotes the pixel position in a picture, the MSE is defined as the follows:

$$MSE = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N [\hat{f}(m, n) - f(m, n)]^2 \quad (3.22)$$

Table 3.5 shows the results achieved by JVT-H014 proposal and our proposed rate control scheme. To see the difference between JVT-H014 and our proposed scheme, we use PSNR gain and bit rate saving in the table. PSNR gain is the PSNR difference in dB between our proposed scheme and JVT-H014. Bit rate saving is the bit rate difference in percentage between our proposed scheme and JVT-H014.

As observed from Table 3.5, our scheme is able to improve both the PSNR and bit rate in most cases. In particular, our scheme performs better for low motion sequences such as “Grandma”, “Silent” and “Weather”. PSNR gain is obtained for all the target bit rates for low motion sequences. Absolute average PSNR gain of 0.27 dB, 0.11 dB and 0.22 dB are achieved for sequence “Grandma”, “Silent” and “Weather” respectively. Although sometimes the bit rate saving is negative, for example at 31.21 kbps for “Grandma”, these low motion sequences still get bit rate saving in most cases and thus are overall better in bit rate saving. For high motion sequences “Children” and “Coastguard”, our scheme performs better in PSNR gain for all the target bit rates.

Table 3.5. Results achieved by JVT-H014 and proposed rate control scheme

Sequence	Target bit rate (kbps)	Initial QP	PSNR (dB)		PSNR gain (dB)	Bit rate (kbps)		Bit rate saving (%)
			H014	Proposed		H014	Proposed	
Carphone	23.34	40	28.91	28.90	-0.01	23.38	23.36	0.09
	42.22	36	31.36	31.29	-0.07	42.22	42.24	-0.05
	79.08	32	34.01	33.97	-0.04	79.05	79.06	-0.01
	147.56	28	36.85	36.77	-0.08	147.51	147.54	-0.02
Children	49.65	40	25.92	26.17	0.25	49.76	49.79	-0.06
	85.39	36	28.96	29.04	0.08	85.49	85.52	-0.04
	143.68	32	32.31	32.35	0.04	143.79	143.79	0.00
	227.19	28	35.65	35.77	0.12	227.33	227.47	-0.06
Coastguard	21.11	40	26.08	26.19	0.11	21.15	21.14	0.05
	41.74	36	28.36	28.36	0.00	41.79	41.77	0.05
	88.35	32	30.96	30.97	0.01	88.40	88.39	0.01
	192.27	28	34.05	34.06	0.01	192.34	193.05	-0.37
Football	126.57	40	27.45	27.35	-0.10	126.26	126.61	-0.28
	215.31	36	29.68	29.54	-0.14	214.81	215.14	-0.15
	362.10	32	32.17	32.07	-0.10	361.85	361.48	0.10
	582.74	28	34.93	34.85	-0.08	582.34	582.39	-0.01
Foreman	27.04	40	28.13	28.09	-0.04	27.13	27.16	-0.11
	43.85	36	30.58	30.55	-0.03	43.89	43.86	0.07
	74.26	32	32.91	32.97	0.06	74.25	74.27	-0.03
	135.13	28	35.57	35.56	-0.01	135.16	135.14	0.01
Grandma	5.55	40	29.90	30.21	0.31	5.71	5.78	-1.23
	8.73	36	31.80	31.99	0.19	8.79	8.78	0.11
	15.48	32	33.95	34.37	0.42	15.50	15.49	0.06
	31.21	28	37.03	37.19	0.16	31.22	31.26	-0.13

Table 3.5. Results achieved by JVT-H014 and proposed rate control scheme (Continued)

Sequence	Target bit rate (kbps)	Initial QP	PSNR (dB)		PSNR gain (dB)	Bit rate (kbps)		Bit rate saving (%)
			H014	Proposed		H014	Proposed	
News	15.51	40	28.16	28.20	0.04	15.58	15.58	0.00
	25.51	36	31.03	30.94	-0.09	25.64	25.61	0.12
	42.30	32	33.86	33.95	0.09	42.43	42.38	0.12
	71.15	28	36.91	36.93	0.02	71.27	71.30	-0.04
Silent	15.60	40	27.96	28.00	0.04	15.64	15.47	1.09
	27.13	36	30.67	30.89	0.22	27.16	27.17	-0.04
	46.69	32	33.43	33.50	0.07	46.70	46.70	0.00
	79.72	28	36.47	36.57	0.10	79.75	79.74	0.01
Suzie	15.03	40	30.11	29.90	-0.11	15.09	15.04	0.33
	23.73	36	32.48	32.33	-0.15	23.81	23.69	0.50
	39.54	32	34.52	34.44	-0.08	39.58	39.63	-0.13
	72.41	28	36.90	36.86	-0.04	72.47	72.54	-0.01
Weather	18.33	40	26.31	26.59	0.28	18.70	18.73	-0.16
	28.82	36	29.17	29.21	0.04	28.93	28.92	0.03
	47.11	32	32.53	32.84	0.31	47.29	47.27	0.04
	78.40	28	36.21	36.45	0.24	78.55	78.60	-0.06

To see the overall performance of our proposed scheme for all the test sequences, we employ average PSNR difference and average bit rate difference as performance measure [73]. The two measures were provided by G. Bjontegaard to evaluate the global RD performance of two different video compression methods [73]. For each compression method, four QPs are required. For each QP, one compression method is performed on the same video sequence, then the PSNR and bit rate are finally obtained as simulation results. Thus for each method, there are four data (pair of PSNR and bit rate) points and a

RD curve can be plotted through these four data points. Deriving from [73], PSNR and bit rate can be interpolated according to the follows respectively:

$$PSNR = a + b \times bit + c \times bit^2 + d \times bit^3 \quad (3.23)$$

$$bit = a + b \times PSNR + c \times PSNR^2 + d \times PSNR^3 \quad (3.24)$$

Since there are two different methods, two different curves are therefore plotted and the average difference between the two curves is calculated according to the algorithm in [73]. Based on the variable which is bit rate or PSNR, the average difference can also be represented as the average PSNR difference over complete range of bit rate or the average bit rate difference over complete range of PSNR. In other words, the average PSNR difference over complete range of bit rate is the average difference between two RD curves where the variable is bit rate and the function is PSNR, and the average bit rate difference over complete range of PSNR is the average difference between two RD curves where the variable is PSNR and the function is bit rate. These measures are often used to compare RD performance between two different methods, which are recommended by JVT Test Model Ad Hoc Group [74]. The source code of implementing the calculation of the two measures is listed in [75]. The results are shown in Table 3.6 and 3.7.

As can be seen from Table 3.6 and 3.7, we can find that the average performance of our scheme is better than JVT-H014 for most of the test video sequences. Comparing with Table 3.5, although the PSNR gain and bit rate saving of sequence “Foreman” do not perform well at some target bit rates, the average PSNR difference and bit rate difference are good. It should be noticed that for sequences “Carphone”, “Football”, and “Suzie”, our scheme performs not as good as JVT-H014. “Football” is the sequence that

has the highest motion among all the test sequences, while “Suzie” is the sequence that has the lowest motion among all the test sequences. This shows that our rate control scheme needs to be further improved for very high or very low motion sequences.

Table 3.6. Average PSNR difference

Sequence	Average PSNR difference over complete range of bit rate (dB)
Carphone	-0.053
Children	0.091
Coastguard	0.017
Football	-0.103
Foreman	0.011
Grandma	0.313
News	0.014
Silent	0.135
Suzie	-0.104
Weather	0.207

Figure 3.10 shows the RD performance of the test sequences. It can be seen that our proposed method achieves better RD performance than that of JVT-H014.

Table 3.7. Average bit rate difference

Sequence	Average bit rate difference over complete range of PSNR (%)
Carphone	1.21
Children	-1.40
Coastguard	-0.52
Football	2.09
Foreman	-0.20
Grandma	-7.45
News	-0.27
Silent	-2.49
Suzie	2.39
Weather	-3.01

(a) Sequence “Children”

(b) Sequence "Grandma"

(c) Sequence "Silent"

(d) Sequence "Weather"

Figure 3.10. RD performance comparison between JVT-H014 and proposed method

(a) “Coastguard”, target bit rate = 192.27

(b) “Foreman”, target bit rate = 74.26 kbps,

kbps, initial QP = 28

initial QP = 32

(c) “Weather”, target bit rate = 47.11 kbps,

initial QP = 32

(d) “Grandma”, target bit rate = 5.55 kbps,

initial QP = 40

Figure 3.11. PSNR vs. frame number

We illustrate the PSNR result with frame by frame as shown in Figure 3.11. In Figure 3.11, there are two high motion sequences and two low motion sequences from high bit rate to low bit rate. Observing from Figure 3.11(a), we can see that our scheme

performs better around the 74th frame. Actually in the original sequence, there is a high motion starting from the 71st frame. The original images of the 71st frame and 72nd frame are shown in Figure 3.12. Because of the large camera motion, the motion effect delays to the 74th frame. In Figure 3.11(a), JVT-H014 performs poorly around this frame. But our scheme has obtained a PSNR gain of 1.3% at this frame. This shows our scheme performs better at high motion than JVT-H014 does. In Figure 3.11(b), our scheme has obtained a PSNR gain of 1.1% at the 258th frame compared to JVT-H014. As we have seen in Figure 3.5, a high motion occurs at the 254th frame. Since more intra-block modes are used in encoding this frame, the motion effect delays to the 258th frame. This shows our scheme performs better at very high motion than JVT-H014 does. Figure 3.11(c) and 3.11(d) are the results for low motion sequences “Weather” and “Grandma” at low bit rates. In these cases, our scheme gets better performance than JVT-H014. It should be especially noticed that even at a very low bit rate of 5.55 kbps (as shown in Figure 3.11(d)), our scheme still performs much better than JVT-H014. This shows that for low motion sequences our scheme performs better than JVT-H014 both in sequence level and frame level.

(a) 71st frame

(b) 72nd frame

Figure 3.12. Two consecutive frames with camera motion in sequence “Coastguard”

For the few sequences like “Carphone”, “Football”, and “Suzie” which do not perform as good as JVT-H014, we investigate the PSNR frame by frame for these sequences. We choose sequence “Football” (the highest motion sequence) and sequence “Suzie” (the lowest motion sequence). We consider the cases that have the largest global PSNR difference, that is, QP = 36 for sequence “Football” and QP = 36 for sequence “Suzie”. The results are illustrated in Figure 3.13.

(a) “Football”, target bit rate = 215.31 kbps, (b) “Suzie”, target bit rate = 23.73 kbps,
initial QP = 36 initial QP = 36

Figure 3.13. PSNR vs. frame number for very high and very low motion sequences

It can be seen that in Figure 3.13(a), there is almost no difference between our scheme and JVT-H014. In Figure 3.13(b), after the 110th frame, JVT-H014 performs better than our scheme. Especially from the 147th frame to the 149th frame, there is a large PSNR drop of our scheme. Observing the original video sequence, we find that there is a big motion happening around the 45th frame and stopping around the 100th frame. Our scheme has checked the motion and allocated more bits for the frames so that we achieve better video quality. This leads to the fact that the remaining bits of our

scheme are less than those of JVT-H014. Thus the quality of our scheme is worse than that of JVT-H014 after the high motion stops. This shows that if there is some high motion in the very low motion sequence, since our scheme has allocated the bits with the same importance before the high motion and then more bits after the high motion, the performance will be degraded after the high motion stops. One possible solution of this problem is to allocate fewer bits before the high motion happens. However, if there is no high motion throughout the sequence, the global performance must be degraded. Thus how to allocate bits before high motion happens is an interesting topic to be investigated in the future.

Table 3.8. Average bit rate mismatch ratio

Sequence	Initial QP	Average bit rate mismatch ratio (%)		Difference of average bit rate mismatch ratio (%)
		H014	Proposed	
Carphone	28	6.92	6.20	0.72
	32	9.15	8.79	0.36
	36	22.26	24.70	-2.44
	40	56.13	54.17	1.96
Children	28	13.57	12.95	0.62
	32	16.04	15.12	0.92
	36	20.81	21.50	-0.69
	40	28.69	28.43	0.26
Coastguard	28	5.95	4.93	1.02
	32	10.02	8.81	1.21
	36	16.11	17.83	-1.72
	40	47.33	46.87	0.46

Table 3.8. Average bit rate mismatch ratio (Continued)

Sequence	Initial QP	Average bit rate mismatch ratio (%)		Difference of average bit rate mismatch ratio (%)
		H014	Proposed	
Football	28	5.56	5.37	0.19
	32	6.57	7.18	-0.61
	36	8.04	7.41	0.63
	40	9.98	8.07	1.91
Foreman	28	8.54	8.67	-0.13
	32	12.75	12.84	-0.11
	36	28.97	37.57	-8.60
	40	53.39	51.88	1.51
Grandma	28	18.43	21.45	-3.20
	32	57.30	57.48	-0.18
	36	65.65	60.55	5.10
	40	71.24	67.66	3.58
News	28	17.53	17.51	0.02
	32	20.78	19.94	0.84
	36	28.13	33.09	-4.96
	40	55.34	50.76	4.58
Silent	28	17.21	15.69	1.52
	32	25.88	25.64	0.24
	36	39.49	40.00	-0.51
	40	61.22	64.35	-3.13
Suzie	28	10.90	11.89	-0.99
	32	18.83	20.87	-2.04
	36	50.92	54.23	-3.31
	40	60.53	62.01	-1.48
Weather	28	15.38	14.67	0.71
	32	25.81	25.64	0.17
	36	38.90	43.35	-4.45
	40	69.08	70.92	-1.84

In order to evaluate the accuracy of target bit estimation at frame i , we employ a measure called bit rate mismatch ratio $r_{BM,i}$ which is defined as follows:

$$r_{BM,i} = \frac{B_i - B_{T,i}}{B_{T,i}} \quad (3.25)$$

where B_i is the actual consumed bits of frame i and $B_{T,i}$ is the estimated target bits of frame i . Table 3.8 lists the average bit rate mismatch ratio with JVT-H014 and our proposed method for all the test video sequences. For sequences “Carphone”, “Children”, “Coastguard”, “Football” and “News”, our proposed method performs better than JVT-H014. For sequences “Foreman” and “Suzie”, our proposed method performs not as good as JVT-H014. As for the other three sequences, the overall performances of both methods are about the same. This shows that our method can achieve more accurate target bit estimation for most of the test sequences. At QP = 28, 32 and 40, our proposed method performs better than JVT-H014. At QP = 36, our proposed method performs not as good as JVT-H014. This shows in general our method can achieve more accurate target bit estimation for most of the QPs.

3.6 Summary

In this chapter, we aim at improving the video quality at scene changes and high motions. We have proposed a minimum distortion based mode decision process that estimates QP after the full-search of motion estimation. We have also proposed a simple motion complexity measure to quantize the complexity of a frame’s motion contents. Based on motion complexity and minimum distortion based mode decision, we have presented a

new rate control scheme for H.264. In this scheme, we employ motion complexity to estimate target bit so that the bit allocation is in accordance with the complexity of frame's motion contents, and we estimate QP according to the macroblock's own information. Our experimental results have shown that our rate control scheme is simple and effective. It outperforms the existing H.264 proposal for most of the video sequences at scene changes and high motions.

Chapter 4

Perceptual Bit Allocation Scheme for H.264

4.1 Introduction

The basic idea of perceptual bit allocation is to allocate more bits to the region which is perceptually more important than the region which is perceptually less sensitive. To implement perceptual bit allocation, Tsai and Tang *et al.* [45][46] classified an object into either a structured moving object or a random texture/moving object. In this classification, background is classified into less insignificant object. However, background part is well structured and is also important to human vision [41][43]. Sengupta *et al.* [44] segmented one full image frame into the foreground part and the background part, and allowed the background quality to degrade gracefully within an acceptable minimum value. Ghandi and Ghanbari [47] implemented perceptual bit allocation by adjusting the Lagrangian multiplier and QP, and checked the resulting rate or distortion to decide further processing. If the resulting rate or distortion is acceptable, the encoding of the frame is completed. Otherwise the frame is coded again using another Lagrangian multiplier and QP. The disadvantage of this method is obvious in time consuming.

One of the main applications of perceptual bit allocation is in biomedical video coding which is briefly discussed below.

In recent years, ongoing improvements in technology have had dramatic impacts on the medical healthcare field. From diagnosis to treatment, advanced technologies are playing an ever-increasing role and the area of medical video is no exception. Easy access to medical videos and 3D medical data sets, such as computed tomography (CT), magnetic resonance imaging (MRI), echocardiography and so on, provides doctors with better capability of analyzing and diagnosing patients' conditions. Medical videos require large amounts of memory storage and transmission bandwidth in telemedicine applications. This calls for video compression to reduce significantly the amount of data needed to represent a video.

Unlike in some other video compression applications where quality could be compromised with compression ratio, it is a very challenging task to compress medical videos as it requires both high quality and high compression ratio. On the one hand, for diagnostic purposes, it is essential that the compression process causes no tangible loss of detail and introduces no noticeable artifacts which could be otherwise misinterpreted as being pathological in nature. On the other hand, due to the limitation of storage and transmission and the huge amount of medical videos, high compression ratio is often required. As a result, although some progress in medical video compression has been made in the past decade, much remains to be done in this area.

In this chapter, different from the segmentation methods in [44]-[46], we segment the full frame into three parts: edge, texture, and background. We observe that the edge part and background part are more important to human vision, and thus more bits are allocated to these parts. To implement perceptual bit allocation, we will propose a perceptual mode decision algorithm that updates Lagrangian multiplier according to the

characteristics of video content. Finally, we will combine our proposed perceptual bit allocation algorithm and region of interest (ROI) bit allocation for an application in biomedical video sequence coding.

4.2 Perceptual Mode Decision

Figure 4.1. Mode migration

From the description of H.264 mode decision and RDO in Chapter 2, it is recognized that the selection of Lagrangian multiplier will influence the selection of coding modes. Figure 4.1 illustrates the effect of updating the Lagrangian multiplier. In Figure 4.1, each point on the RD curve represents a coding mode. Assume that mode 1 is the initial case where the RD cost is calculated according to Equation (2.1). If we increase the Lagrangian multiplier and maintain RD cost, rate will be decreased and the coding mode will transfer to mode 2 which means fewer bits are allocated and larger distortion is possible. If we decrease the Lagrangian multiplier and maintain RD cost, rate will be

increased and the coding mode will transfer to mode 3 which means more bits are allocated and less distortion is allowed.

One of the important characteristics of human vision system (HVS) is called edge masking, that is, HVS is more sensitive to errors on a prominent edge in an image while less so to errors next to it. In other words, a dominant edge pattern will obscure the perception of other lower contrast variations in this block [76]. Hence, more bits should be allocated to the dominant edge pattern, while fewer bits should be allocated to the texture pattern. This observation can also be illustrated in Figure 4.1. If the macroblock is texture pattern, fewer bits should be allocated to it and its coding mode is more likely mode 2. On the other hand, if the macroblock is edge pattern, more bits should be allocated to it and its coding mode is more likely mode 3. Therefore, we can consider the bits, the Lagrangian multiplier and the macroblock pattern simultaneously to allocate bits perceptually. That is, we update the Lagrangian multiplier according to the macroblock pattern, and this will change the mode selection, and correspondingly the allocated bits.

Based on the above analysis, a segmentation and mode decision method has been proposed to allocate bits perceptually. The steps of this method are listed as follows.

(1) Calculate the Sobel operators [77]:

$$G_x = I_{x-1,y+1} + 2I_{x,y+1} + I_{x+1,y+1} - I_{x-1,y-1} - 2I_{x,y-1} - I_{x+1,y-1} \quad (4.1)$$

$$G_y = I_{x+1,y-1} + 2I_{x+1,y} + I_{x+1,y+1} - I_{x-1,y-1} - 2I_{x-1,y} - I_{x-1,y+1} \quad (4.2)$$

(2) Calculate the average squared gradients:

$$G_{xx} = \sum_W G_x^2, \quad G_{yy} = \sum_W G_y^2, \quad G_{xy} = \sum_W G_x G_y \quad (4.3)$$

where W is the window of image block.

(3) Calculate the coherence of the squared gradient [78]:

$$Coh = \frac{\sqrt{(G_{xx} - G_{yy})^2 + 4G_{xy}^2}}{G_{xx} + G_{yy}} \quad (4.4)$$

(4) Determine the block patterns according to edge threshold T_{Edge} and texture threshold $T_{Texture}$:

$$Pattern = \begin{cases} Edge, & \text{if } Coh > T_{Edge} \\ Texture, & \text{if } T_{Texture} < Coh \leq T_{Edge} \\ Background, & \text{others} \end{cases} \quad (4.5)$$

(5) Update the Lagrangian multiplier:

$$a_i = \begin{cases} 0.5, & \text{if } Pattern = Edge \\ 1.5, & \text{if } Pattern = Texture \\ 0.5, & \text{if } Pattern = Background \end{cases} \quad (4.6)$$

$$a = \sum_w a_i \quad (4.7)$$

$$\lambda' = a\lambda \quad (4.8)$$

where λ' is the updated Lagrangian multiplier.

The result of the above segmentation process can be shown with an example in Figure 4.2. Figure 4.2(a) is the original image of the 32nd frame of medical video sequence “Echocardiography”. Figure 4.2(b) is its segmentation image. In Figure 4.2(b), the black region is the background. The gray region is the texture which is not important and can be greatly compressed. The white region is the edge which has the most important cardiac information and should be compressed perceptually. It can be seen that most of the important edges and backgrounds can be recognized in the segmentation image. Figure 4.3 is an example of sequence “CT”.

(a) Original image

(b) Segmentation image

Figure 4.2. The 32nd frame of sequence “Echocardiography”

(a) Original image

(b) Segmentation image

Figure 4.3. The 44th frame of sequence “CT”

4.3 Perceptual Bit Allocation Algorithm

We employ our proposed rate control method in Chapter 3 to implement perceptual bit allocation. At the frame level rate control, we estimate target bits and QP of a frame according to the output buffer model and motion complexity measure. At the macroblock level rate control, we estimate QP and calculate the original Lagrangian multiplier. Then we use Equations (4.1) – (4.8) to update the Lagrangian multiplier. The updated Lagrangian multiplier joins the RD cost computation to select the best mode. Since the mode is updated according to the macroblock's perceptual characteristics, more bits are allocated to the edge and background blocks.

4.4 Experimental Results

The evaluation of experimental results includes two parts: objective quality evaluation and subjective quality evaluation. H.264 is implemented with H.264 reference model version JM6.1e. The experimental conditions are the same as in Chapter 3, as shown in Table 4.1. We still compare our proposed method with JVT-H014 [31]. The specifications of the test sequences are listed in Table 4.2. We use the same method as in Chapter 3 to determine target bit rates which are shown in Table 4.3.

Table 4.1. Test conditions

Frame type	IPPP...
Frame rate	30 fps
Slice mode	OFF
RDO	ON
Hardmard	OFF
Search range	32
Restrict search range	No restriction
Symbol mode	CABAC
Partition mode	No data partition
Out File mode	Annex B

Table 4.2. Specifications of the test sequences

Sequence	Frame size	No. of frames	Frame rate (fps)	Format
CT	512×512	60	30	4:2:0 YUV
Echocardiography	640×480	60	30	4:2:0 YUV

Table 4.3. Target bit rate

Sequence	Initial QP	Target bit rate (kbps)
CT	28	438.19
	32	285.03
	36	197.55
	40	145.10
Echocardiography	28	2266.14
	32	1317.29
	36	667.08
	40	308.26

4.4.1 Objective Quality Evaluation

Table 4.4 shows the results achieved by the JVT-H014 rate control algorithm and the proposed perceptual bit allocation scheme. To see the difference between JVT-H014 and the proposed scheme, we use PSNR gain and bit rate saving as performance measures in the table. PSNR gain is the PSNR difference in dB between the proposed scheme and JVT-H014. Bit rate saving is the bit rate difference in percentage between the proposed scheme and JVT-H014.

Table 4.4. Results of JVT-H014 and proposed scheme

Sequence	Target Bit Rate (kbps)	PSNR (dB)		PSNR Gain (dB)	Bit Rate (Kbps)		Bit Rate Saving (%)
		H014	Proposed		H014	Proposed	
CT	438.19	42.89	42.97	0.08	439.45	439.80	-0.08
	285.03	40.28	40.43	0.15	286.37	286.92	-0.19
	197.55	37.43	37.69	0.26	198.00	197.91	0.05
	145.10	35.06	35.33	0.27	146.73	145.79	0.64
Echocardiography	2266.14	38.17	38.16	-0.01	2268.22	2265.24	0.13
	1317.29	35.20	35.27	0.07	1319.49	1318.66	0.06
	667.08	32.43	32.56	0.13	671.47	670.46	0.15
	308.26	29.84	30.00	0.16	311.73	308.45	1.05

As observed from Table 4.4, the proposed scheme is able to improve both the PSNR and bit rate in most cases. As QP increases, or in other words the target bit rate decreases, the PSNR gain of the proposed scheme compared with JVT-H014 increases. This shows the proposed scheme performs better in the low bit rate network environment. In particular, the proposed scheme performs better for high motion sequence “CT”. PSNR

gain is positively obtained for all the target bit rates for sequence “CT”. Although sometimes the bit rate saving is negative, for example at 285.03 kbps, the proposed scheme still gets bit rate saving in most cases and is thus overall better in bit rate saving. For sequence “Echocardiography”, the proposed scheme obtains positive bit rate saving in all the cases and positive PSNR gain in most cases. To see the performance of sequence “Echocardiography” at very high bit rate, we have also done experiments in the case of $QP = 24$ and target bit rate = 3402.04 kbps. In this case, the PSNR and the bit rate of H014 are 41.29 dB and 3403.36 kbps, respectively, while the PSNR and the bit rate of the proposed scheme are 41.26 dB and 3403.38 kbps, respectively, implying that the PSNR gain and the bit rate saving of the proposed scheme are -0.03 dB and 0.00%, respectively. This shows that at very high bit rate for sequence “echocardiography”, H014 performs slightly better than the proposed scheme. Despite of this, the proposed scheme is still overall better than JVT-H014.

Table 4.5. Average PSNR difference over complete range of bit rate between JVT-H014 and the proposed perceptual bit allocation scheme

Sequence	Average PSNR difference (dB)
CT	0.19
Echocardiography	0.11

Table 4.6. Average bit rate difference over complete range of PSNR between JVT-H014 and the proposed perceptual bit allocation scheme

Sequence	Average bit rate difference (%)
CT	-2.65
Echocardiography	-2.66

Tables 4.5 and 4.6 are the results of average PSNR difference and bit rate difference between the proposed rate control scheme and JVT-H014. As can be seen from the tables,

the average performance of the proposed scheme is better than JVT-H014. Comparing with Table 4.4, although sometimes the PSNR gain and bit rate saving do not perform well at some target bit rates, the average PSNR difference and bit rate difference are still good. This shows the proposed scheme outperforms the H.264 rate control algorithm JVT-H014.

(a) Sequence "CT"

(b) Sequence "Echocardiography"

Figure 4.4. RD performance comparison between JVT-H014 and proposed scheme

Figure 4.4 shows the RD performance of the test sequences. It can be seen that our proposed scheme achieves better RD performance than that of JVT-H014.

(a) Sequence "CT", initial QP = 32

(b) Sequence "CT", initial QP = 36

(c) Sequence "Echocardiography", initial QP = 32

(d) Sequence “Echocardiography”, initial QP = 36

Figure 4.5. PSNR vs. frame number

Figure 4.5 shows the PSNR results frame by frame at initial QP = 32 and initial QP = 36. It can be seen from the figure that the proposed scheme achieves higher PSNR than JVT-H014 at most frames. This shows again that the proposed scheme performs better than JVT-H014 in terms of PSNR.

4.4.2 Subjective Quality Evaluation

For subjective quality evaluation, we use the double stimulus continuous quality scale (DSCQS) test method which is described in ITU-R BT.500-11 [79]. The mean opinion score (MOS) scales for observers to vote for the quality after viewing are excellent, good, fair, poor and bad. Correspondingly, the scores are 5, 4, 3, 2 and 1. Ten observers were involved in the subjective quality evaluation. Except one medical doctor, the other observers have no expert knowledge on medical video processing. Table 4.7 lists the scores for sequence “CT”. Table 4.8 lists the scores for sequence “Echocardiography”. Table 4.9 lists the average results. The higher the score gain is, the better the perceptual

quality is achieved. From the Table 4.9, we can see that the subjective score of the proposed scheme is better than that of JVT-H014. This shows that the proposed scheme can achieve better perceptual quality than JVT-H014. It is interesting to see that positive score gains are achieved for all the low bit rate cases for both sequences, agreeing with the objective quality evaluation results in Section 4.4.1.

Table 4.7. Subjective score for sequence “CT”

Method	QP	Score (10 observers: A to J)										Average score
		A	B	C	D	E	F	G	H	I	J	
H014	28	5	5	4	4	3	5	5	5	5	5	4.60
	32	5	5	4	3	1	4	4	4.2	4.7	4	3.89
	36	5	5	4	3	1	3	3	3.5	4.2	3.5	3.52
	40	5	5	4	3	1	2	2	3.3	3.5	2	3.08
Proposed	28	5	5	5	4	2	5	5	5	5	5	4.60
	32	5	5	4	3	2	4	4	4.2	4.6	4	3.98
	36	5	5	4	4	2	3	3	3.6	4.3	3	3.69
	40	5	5	4	4	1	3	2	3.3	3.6	2.5	3.34

Table 4.8. Subjective score for sequence “Echocardiography”

Method	QP	Score (10 observers: A to J)										Average score
		A	B	C	D	E	F	G	H	I	J	
H014	28	5	4	4	3	5	4	5	5	5	4.8	4.48
	32	5	4	4	3	3	3	3	4.5	4.6	4.2	3.83
	36	5	4	4	3	4	3	2	4	4	3.3	3.63
	40	5	4	4	3	2	1	1	3.8	3.6	2.5	2.99
Proposed	28	5	4	4	3	5	5	5	5	5	5	4.60
	32	5	4	4	3	5	4	4	4.5	4.7	4	4.22
	36	5	4	4	4	4	3	3	4	4	3	3.80
	40	5	4	4	3	3	2	2	3.7	3.8	2	3.25

Table 4.9. Subjective score of JVT-H014 and the proposed bit allocation scheme

Sequence	QP	Target Bit Rate (kbps)	Score		Score Gain
			H014	Proposed	
CT	28	438.19	4.60	4.60	0.00
	32	285.03	3.89	3.98	0.11
	36	197.55	3.52	3.69	0.17
	40	145.10	3.08	3.34	0.26
Echocardiography	28	2266.14	4.48	4.60	0.12
	32	1317.29	3.83	4.22	0.39
	36	667.08	3.63	3.80	0.17
	40	308.26	2.99	3.25	0.26

4.5 Region of Interest Bit Allocation Applications for Medical Sequences

In addition to the perceptual bit allocation approach, we will also exploit the approach of ROI bit allocation in this section. The ROI-based approach is particularly relevant to the compression of medical images and video sequences since in a typical medical video sequence, there are usually some regions of special interests to the doctors, and other regions of little diagnostic significance. It is often impossible to maintain the video quality of the non-interesting region at a given channel bit rate throughout. Instead, one or more ROIs are defined and the video quality of the ROI is ensured. Figure 4.6 is an example that shows what a ROI part is in our test video sequence. The region inside the

white square box is the ROI of the image. It is obvious that in Figure 4.6 the ROI includes the whole left ventricle (LV) that is of diagnostic interest.

(a) Original image

(b) ROI part

Figure 4.6. The 3rd frame of sequence “LV”

In [80], only the ROI part is compressed and transmitted. However in the telediagnosis situation, the non-ROI part may also have clues for diagnosis. Thus we not only compress the ROI, but also the non-ROI, although their interests are different. In what follows we propose two simple ROI-based video compression methods, which we call Lagrangian multiplier update (LMU) and QP update (QPU). For the former, the value of QP is kept unchanged for all macroblock's in the same image while for the latter, the value of QP will change according to whether a macroblock is within the ROI or not. These two methods are now described briefly as follows.

LMU method utilizes the video content information in bit allocation. First, the frames are segmented into different kinds of macroblocks, which are edge, texture and background. Second, for each macroblock, the Lagrangian multiplier is updated in mode selection so that more bits are assigned to the edge pattern and fewer bits are assigned to the texture pattern. Moreover, LMU method uses ROI segmentation. That is, the

macroblocks inside the ROI are segmented into different macroblock patterns, while the macroblocks outside the ROI are not segmented.

QPU method exploits a different approach by changing the value of QP according to the macroblock's position. Note that in the RDO mode decision process, the value of Lagrangian multiplier is decided empirically as Equation (2.7). It can be seen from the equation that a larger QP leads to a larger Lagrangian multiplier and thus a smaller amount of bits. If the macroblock is inside the ROI, the QP will be decreased and the macroblock will be assigned with more bits. If the macroblock is outside the ROI, the QP will be increased and the macroblock will be assigned with fewer bits. This method is motivated by our desire of maintaining high video quality inside the ROI without increasing the overall bit rates. However, in order to maintain the smoothness of video quality, the QP increment or decrement should be set to a small value (1 or 2).

Table 4.10. Target bit rate

Initial QP	Target bit rate (kbps)
28	81.38
32	45.55
36	26.49
40	15.22

We implement our ROI bit allocation scheme in the H.264 reference model JM12.0 [81]. The video sequence used in the experiment comes from a series of magnetic resonance imaging (MRI) images in a complete LV activity cycle. It consists of slices of a parallel stack of contiguous MRI LV sequences acquired in the short-axis plane of the LV, with each containing temporally resolved frames depicting LV motion in a single

heart cycle in the respective slices. The number of frames is 25. The frame size is 160×192 . The video format is 4:2:0 YUV. The experimental conditions are the same as shown in Table 4.1. Table 4.10 lists the target bit rate.

For QPU method, there are several ways of varying the QP values. One way is to decrease QP by 1 if the macroblock is inside the ROI, and to increase QP by 1 if the macroblock is outside the ROI. We call this method as QPU1. Another way is to decrease QP by 1 if the macroblock is inside the ROI, and to increase QP by 2 if the macroblock is outside the ROI. We call this method as QPU2. To evaluate the performance of our methods, we compare them with JM12.0.

Table 4.11. PSNR and bit rate

QP	JM12.0		LMU		QPU1		QPU2	
	Bit Rate (kbps)	PSNR (dB)	Bit Rate (kbps)	PSNR (dB)	Bit Rate (kbps)	PSNR (dB)	Bit Rate (kbps)	PSNR (dB)
28	80.74	36.35	81.24	36.38	83.26	36.23	82.02	36.17
32	45.12	33.85	45.66	33.85	47.10	33.93	46.84	33.77
36	25.61	31.36	26.15	31.42	26.78	31.56	26.30	31.42
40	14.98	29.09	15.09	29.24	15.66	29.34	15.46	29.28

Table 4.11 shows the results achieved by the JM12.0 and the proposed methods. As observed from Table 4.11, the proposed methods are able to improve the PSNR in most cases.

The average PSNR difference and average bit rate difference are shown in Table 4.12. As can be seen from the table, although the PSNR increases with the increasing of bit rate cost, the bit rate change is within an acceptable limitation.

Table 4.12. Average PSNR difference over complete range of bit rate and average bit rate difference over complete range of PSNR

LMU		QPU1		QPU2	
PSNR (dB)	Bit Rate (%)	PSNR (dB)	Bit Rate (%)	PSNR (dB)	Bit Rate (%)
-0.015	0.301	-0.060	1.425	-0.135	3.269

Table 4.13 is the corresponding average PSNR of the ROI part. Figure 4.7 shows the PSNR of encoded frames within the ROI part when QP = 32 and QP = 40. From Figure 4.7 and Table 4.13, it is clear that the proposed methods achieve better performance than JM12.0 does in ROI.

Table 4.13. Average PSNR inside ROI (dB)

QP	JM12.0	LMU	QPU1	QPU2
28	34.82	34.93	35.19	35.26
32	31.87	31.98	32.31	32.34
36	29.18	29.30	29.59	29.61
40	26.62	26.84	27.00	26.94

For different QPs, we also compare the reconstructed coding sequences of our proposed methods with those of JM12.0 in the video viewer. We observe that inside the ROI, the sequences of our proposed methods appears a little better vision quality than those of JM12.0. This is identical with the points achieved from Table 4.11, Table 4.12, and Figure 4.7. This means that our proposed method is effective for ROI video coding.

Figure 4.8 gives the result of PSNR vs. QP inside ROI part. It can be seen that when QP increases, the PSNR decreases. The PSNRs of all our proposed methods are higher than that of JM12.0. This shows our proposed methods perform better than JM12.0.

Moreover, our variable QP methods (QPU1 method and QPU2 method) perform better than the fixed QP method. This means the method of updating QP is better than updating Lagrangian multiplier. Notice that the video quality varies with the QP, our method can be used as a scalable scheme for different quality requirement.

(a) QP = 32

(b) QP = 40

Figure 4.7. PSNR vs. frame number inside ROI

Figure 4.8. PSNR vs. QP inside ROI

4.6 Summary

In this chapter, we have proposed a perceptual mode decision algorithm by updating Lagrangian multiplier according to the perceptual characteristics of video contents, so that more bits are allocated to edge patterns. Based on the perceptual mode decision algorithm, we have presented a perceptual bit allocation scheme for H.264. For medical video application, we have also proposed an efficient ROI bit allocation scheme. Experimental results have shown that our proposed perceptual bit allocation scheme outperforms H.264 rate control for medical video compression.

Chapter 5

Scalable Fast Mode Decision Scheme for H.264

5.1 Introduction

The basic idea of fast RDO based mode decision in H.264/AVC is to select the coding mode that achieves the best RD performance without searching all the modes, leading to the reduction of computational complexity. This is based on the observation that a large macroblock partition suits slow motion and simple texture video objects, while a small partition size suits fast motion or complex scenes. Moreover, the occurrence of having different partition sizes in motion compensation is not equal, and can be decided by using the information of the temporal and spatial contents.

A number of efforts have been made to reduce the computational complexity of H.264/AVC by using various fast mode decision algorithms, such as fast SKIP mode decision [51]-[54], fast intra mode decision [55]-[59], fast inter mode decision [60]-[64], and the combination of the above [65]-[69]. All the existing methods are based on the temporal correlation between current macroblock and its matching macroblock in the previous frame, and the spatial correlation between current macroblock and its neighboring macroblocks in the current frame. Therefore, these fast mode decision strategies are basically dependent on the characteristics of video content.

This chapter will present a new way for H.264/AVC fast mode decision. Unlike other methods which depend on temporal and spatial correlations, we study the probability distribution of coded modes. It is well known that, for most of the real life video sequences, macroblock coding modes such as SKIP and INTER16×16 have much higher occurrences than the other coding modes. Thus in RDO process, we prioritize the macroblock coding modes such that the highest probable mode will be tried first, followed by the second highest probable mode, and so on. The macroblock coding mode with the least occurring probability will be tried at last. In this process, the computed RD cost will be checked against a content adaptive RD cost threshold to decide if we should terminate the RDO process before trying all of the possible modes. By adjusting this threshold we can actually control the time when the early termination can be activated, and thus this threshold can be used to determine the tradeoff between time saving and PSNR loss. We could achieve a very significant time saving by increasing this threshold with compromise of PSNR, or we can achieve a very good PSNR performance by reducing the threshold, but the time saving will not be very significant.

The advantage of the above fast RDO algorithm is that the order of macroblock coding mode is in accordance with the actual occurring probability of the macroblock coding modes. This can make sure that the more popular coding modes are checked prior to the less popular modes. In this way many unpopular coding modes are skipped, and the computational time is significantly reduced. Therefore we can terminate the RDO process at any time when the RD cost is below a preset threshold. Another advantage of the proposed algorithm is that the threshold can be adjusted according to the user's preference of the tradeoff between time saving and PSNR loss, and thus provides a

scalable time saving mechanism. Note that in our scheme, the RD costs of the macroblock coding modes are calculated and compared against a preset threshold one after another, until the early termination requirement is met. This differs fundamentally from other existing fast mode decision strategies where a reduce number of coding modes are tested and the one that produce the minimum RD cost are selected.

As compared with the parallel mode decision strategy shown in Figure 2.7, our mode decision scheme is based on sequential decision strategy, such that we will compute the RD costs of macroblock coding modes one after another, in the order of descending occurring probability. This process will terminate immediately when the computed RD cost is below a threshold, which is adaptively decided by statistics of the neighbouring macroblocks. Figure 5.1 illustrates the sequential mode decision procedure.

Figure 5.1. Sequential mode decision scheme of RDO

5.2 Statistical Study on Macroblock Coding Modes in H.264

If the best coding mode can be determined at an early stage of RD cost computation, significant time saving can be achieved. The early termination strategy can be fulfilled making use of the local temporal and spatial contents, as well as a content adaptive threshold. In addition, motion estimation for any coding mode is performed only if there is a need to calculate the RD cost of this mode, and thus, the overall structure of the RDO process has to be modified to facilitate the early termination strategy.

5.2.1 Probability Distribution of Macroblock Coding Modes

It is observed that in encoding a natural video sequence, macroblocks in a slow-motion and low complexity frames are usually coded using SKIP or larger partition such as 16×16 , whereas macroblocks in a fast-motion or high-complexity frames are likely to be coded using smaller partitions such as 8×8 , 8×4 , 4×8 or 4×4 . Due to the strong temporal correlation between consecutive frames, the probability of encoding a macroblock using inter mode is much higher than using intra mode.

To verify the above observations, extensive experiments have been conducted on different sequences and at different QPs to find out the statistics of macroblock coding modes in test video sequences. Figure 5.2 shows an example of the macroblock coding mode statistics of twelve test sequences by using full search RDO.

(a) QCIF, QP = 28

(b) QCIF, QP = 40

(c) CIF, QP = 28

(d) CIF, QP = 40

Figure 5.2. Probability distribution of macroblock coding modes for different sequences

In Figure 5.2, mode SKIP, INTER16×16, INTER16×8, INTER8×16, INTER8×8, INTRA16×16 and INTRA4×4 are represented by number 1 to 7 respectively. It can be seen from Figure 5.2(a) and 5.2(c) that for slow-motion sequences such as “Akiyo”, “Claire”, “Container”, more than 85% of their macroblocks are encoded using the SKIP or INTER16×16 modes, and less than 5% of their macroblocks are encoded using

INTER8×8, INTRA16×16 and INTRA4×4 modes. On the other hand, for fast-motion and high-complexity sequences such as “Foreman”, “Mobile” and “Stefan”, more than 40% of their macroblocks are encoded using the coding modes with smaller partitions. For example, the probabilities of using smaller partitions for “Foreman”, “Mobile” and “Stefan” are 47%, 63% and 47% respectively. Furthermore, the probability of SKIP mode or the modes with larger partitions increases as QP increases, and the probability of the modes with smaller partitions decreases as QP decreases, which are shown in Figure 5.2(b) and 5.2(d). Therefore, significant time saving can be achieved if we design an intelligent early termination strategy during RDO by taking into account of the probability distribution of selected macroblock coding modes.

5.2.2 Mean Value and Standard Deviation of RD Cost

In RDO process, the RD cost of each macroblock coding mode must be computed in order to decide which mode would be eventually used. Thus an early termination strategy can be designed based on the RD cost of each coding mode. In order to activate the early termination correctly, we have conducted an experiment to explore the statistical properties of RD cost, such as their mean value and standard deviation, for different sequences under different QPs. In the statistical experiments, after finishing the mode decision of one macroblock, we record its mode number, RD cost, and the accumulated number of macroblocks of the mode. Then we calculate the mean value and standard derivation of RD cost of the mode. The procedure of obtaining the mean value and standard derivation is as follows. Let J_i be the RD cost of a macroblock which mode

belongs to a determined mode, i be the number of macroblocks of that mode, $E_{J,i}$ be the mean value of the mode's RD cost, and suppose the total number of macroblocks of the mode be n , then

$$\begin{aligned}
 E_{J,n} &= \frac{1}{n} \sum_{i=1}^n J_i \\
 &= \frac{1}{n} \left[(n-1) \frac{1}{n-1} \sum_{i=1}^{n-1} J_i + J_n \right] \\
 &= \frac{1}{n} \left[(n-1) E_{J,n-1} + J_n \right]
 \end{aligned} \tag{5.3}$$

Assume that the mean value of n macroblocks is equal to the mean value of $n-1$ macroblocks, that is, $E_{J,n} = E_{J,n-1}$, then:

$$\begin{aligned}
 \sigma_{J,n} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (J_i - E_{J,n})^2} \\
 &= \sqrt{\frac{1}{n} \sum_{i=1}^n (J_i - E_{J,n-1})^2} \\
 &= \sqrt{\frac{1}{n} \left[(n-1) \frac{1}{n-1} \sum_{i=1}^{n-1} (J_i - E_{J,n-1})^2 + (J_n - E_{J,n-1})^2 \right]} \\
 &= \sqrt{\frac{1}{n} \left[(n-1) \sigma_{J,n-1}^2 + (J_n - E_{J,n})^2 \right]}
 \end{aligned} \tag{5.4}$$

Table 5.1 is the statistical result for QCIF sequence “Foreman”. From this table, we can see that for different coding modes, the mean and standard deviation of RD cost are quite different. In most cases, SKIP mode has the lowest mean and standard deviation of RD cost, while INTRA4×4 mode has the highest mean value and standard deviation of RD cost. As QP increases, the mean value and standard deviation increase too. It can be seen that the standard deviation of RD cost has very large values and thus the RD cost of different macroblocks vary largely. Moreover, in most cases, the mean value of RD cost

for most coding modes is in accordance with its occurring probability. The macroblock coding mode which has higher occurring probability usually produces lower mean value of RD cost.

Table 5.1. Mean value and standard deviation of RD cost for sequence “Foreman”

Coding Mode	QP = 28		QP = 32		QP = 36		QP = 40	
	Mean Value	Standard Deviation	Mean Value	Standard Deviation	Mean Value	Standard Deviation	Mean Value	Standard Deviation
SKIP	3684.53	2315.38	6825.61	4637.07	12696.27	9058.78	23508.09	17039.48
INTER16×16	5572.83	2854.27	10685.00	5534.81	20693.56	10551.22	39327.44	19721.68
INTER16×8	7000.06	2993.13	13683.81	5878.98	26175.07	10734.74	48750.11	19887.81
INTER8×16	6528.57	2810.95	12802.26	5554.29	24964.86	11003.29	48819.00	21537.71
INTER8×8	9432.36	3162.65	18392.22	5930.96	33771.91	11576.23	73151.46	23049.07
INTRA16×16	1234.20	962.37	3474.17	3199.58	8930.49	7893.12	24144.37	17135.81
INTRA4×4	9756.36	3797.78	18599.62	6672.90	36940.29	10932.14	73401.14	19377.83

Figure 5.3 illustrates the mean value of RD cost vs. number of macroblocks for sequence “Foreman” at QP = 28. It is very clear that for different macroblock coding modes, their mean value of RD cost are different. Figure 5.4 illustrate the standard deviation of RD cost vs. number of macroblocks for sequence “Foreman” at QP = 28. From the figure, macroblock modes can also be classified according to their standard deviation of RD cost. This shows that mean value and standard deviation of RD cost are good measures to distinguish different coding modes. This is the reason that they can be used in early termination which details will be discussed later in this chapter.

Figure 5.3. Mean of RD cost vs. macroblock number for sequence “Foreman”

Figure 5.4. Standard deviation of RD cost vs. macroblock number for sequence
“Foreman”

5.2.3 Correlation Coefficient of RD Cost

Although RD cost varies largely for different modes, the RD cost of neighbouring macroblocks and their co-located macroblocks in the reference frame are highly correlated. This is evident from experiments. We use the correlation coefficient of RD cost between consecutive frames to represent the correlation of RD cost, which is defined as follows:

$$\rho = \frac{Cov_{i,i-1}}{\sqrt{Var_i \times Var_{i-1}}} \quad (5.3)$$

where $Cov_{i,i-1}$ is the covariance of RD cost between Frame i 's and Frame $i-1$'s macroblocks, Var_i is the variance of RD cost of Frame i 's macroblocks, Var_{i-1} is the variance of RD cost of Frame $i-1$'s macroblocks. Note that Frame i and Frame $i-1$ are two consecutive frames. Figure 5.5 shows the correlation coefficient of RD cost between consecutive frames at QP = 28 of four different sequences. In Figure 5.5, the average values of correlation coefficient for all the sequences are larger than 0.9. The average correlation coefficient for less complicated sequences such as “Akiyo” is 0.983, and that for fast sequences such as “Stefan” is 0.952. This implies that the RD costs are very similar between consecutive frames, and thus provides a good basis for predicating the RD cost of the current frame's macroblock, which can be used to activate the early termination. That is, the statistical properties such as the mean value and standard deviation of RD cost of previous macroblocks can be used during the early termination in mode selection.

Figure 5.5. Correlation coefficient of RD cost

5.3 Priority Mechanism and Early Termination Strategy

5.3.1 Priority Mechanism

Based on the observation as described in Section 5.2.1, we have designed an algorithm to sort the order of the macroblock coding modes for each macroblock to be coded according to their occurring probabilities. Occurring probability of a coding mode is the probability that the mode is selected as the best mode. Let n_i be the number of times that Mode i is selected as the best mode, n be the total number of previously processed macroblocks, then the occurring probability of Mode i is:

$$P_i = \frac{n_i}{n} \quad (5.4)$$

The macroblock coding mode that has the highest occurring probability will be placed at the beginning of the queue, and will have the highest priority to be checked in mode decision process, while the macroblock coding mode that has the lowest occurring probability will be placed at the bottom of the queue and will be the last to be checked. If one coding mode has met the early termination criterion such that its RD cost is below the given threshold, this mode will be chosen as the best mode, and the remaining coding modes in the queue will be skipped. After that the priority queue will be updated, which will be used for mode decision of the next macroblock. Figure 5.6 illustrates this mechanism of prioritizing the macroblock coding modes.

Figure 5.6. Priority mechanism

Since the order of the coding modes in the priority queue is in accordance with the occurring probability of that mode, the more popular modes are always checked prior to the less popular modes. This ensures that the more popular modes are kept while the less popular modes are skipped if necessary. Thus the computational time is reduced.

5.3.2 Early Termination Strategy

The objective of an early termination is to decide whether a macroblock coding mode has met the RD cost criterion so that the mode selection for the current macroblock can be terminated early without trying the rest of the coding modes in the queue. Based on the observations in Section 5.2.2 and Section 5.2.3, an early termination mechanism according to the mean value and standard deviation of RD cost is defined as follows,

$$J \leq E_J - \alpha \sigma_J \quad (5.5)$$

where J is the macroblock's RD cost as in Equation (2.3), E_J is the mean value of the mode's RD cost, σ_J is the standard derivation of the same mode's RD cost, and α is a positive constant coefficient.

As we have shown, each coding mode has its own mean value of RD cost that is different from the one of the others, and the best coding mode has the minimum RD cost. If the RD cost of the current mode satisfies Equation (5.5), this means the current mode is, or very closed to the optimal mode, and the PSNR loss will be negligible even if it is not the optimal mode. Therefore, the RDO process stops and the current mode is selected as the best coding mode. Here $E_J - \alpha \sigma_J$ means the best coding mode approaches in the direction of reducing the average RD cost so that video quality is maintained. In Equation (5.5), α is a parameter to control the video quality and computational time. If we want to save more time, α can be set to a lower value. On the other hand, if we want to maintain high video quality, α can be set to a higher value. Therefore the adjusting of α makes our fast RDO algorithm scalable in terms of time saving.

5.4 Fast Mode Decision Algorithm

Based on the proposed prioritization mechanism and the early termination measure, the fast RDO algorithm is proposed as follows.

- (1) Sort the coding modes according to their occurring probability, and place them into the priority queue.
- (2) Test the mode from the beginning of the priority queue. Compute its RD cost and check against early termination criterion.
- (3) If the RD cost satisfies Equation (5.5), select the current mode as the best mode (early termination) and go to Step (5).
- (4) If the current mode is not the last mode in the priority queue, go to Step (2); otherwise select the mode with the minimum RD cost as the best mode.
- (5) Update the priority queue of the coding modes according to the new probability distribution.

The proposed fast RDO algorithm is summarized in Figure 5.7.

Initially, for the first macroblock of the first P frame, all the modes are placed into the priority queue in the order of 1 to 7. Then a full search method is used to select the best coding mode for the first macroblock. After one coding mode has been selected as the best mode, the priority queue is updated according to the occurring probability of that mode such that the mode that has the highest occurring probability is placed at the beginning of the priority queue.

Figure 5.7. Flowchart of the proposed fast RDO algorithm

The mean value and standard deviation of RD cost are predicted dynamically according to that of the previous macroblocks. For the n th macroblock of the mode,

$$\hat{E}_{J,n} = \frac{1}{n} \left[(n-1) E_{J,n-1} + J_n \right] \quad (5.6)$$

$$\hat{\sigma}_{J,n} = \sqrt{\frac{1}{n} \left[(n-1) \sigma_{J,n-1}^2 + \left(J_n - \hat{E}_{J,n} \right)^2 \right]} \quad (5.7)$$

where $E_{J,n-1}$ is the true mean value of RD cost for the previous $n-1$ macroblocks, $\sigma_{J,n-1}$ is the true standard deviation of RD cost for the previous $n-1$ macroblocks. Initially, $E_{J,0} = 0$, $\sigma_{J,0} = 0$.

One special case is the INTER8×8 mode. In checking the INTER8×8 mode, the 8×8 block will be further partitioned into smaller blocks such as 8×8, 8×4, 4×8, and 4×4. The RD cost of the sub-blocks are computed separately and their summation is the RD cost of mode INTER8×8. Therefore, no matter what the size of the sub-block is, the coding mode is still considered as INTER8×8.

5.5 Experimental Results

To evaluate the performance of the proposed fast RDO algorithm, we compare it with JVT reference model software JM8.2 [82]. All the simulations are performed using a Pentium-4 3.00GHz processor with 512MB DDR RAM. The conditions of the experiment are listed in Table 5.2. The main difference of the experiment in this chapter and those in the previous chapters is that the rate control has not been performed in this chapter. Thus there is no requirement of the target bits.

In the experiments, the frame rate of the sequences is 30 frames per second. For QCIF sequences, the number of frames is 240. For CIF sequences, the number of frames is 120. For each sequence, four QP values of 28, 32, 36, and 40 are used. The comparison results are produced and tabulated based on the averaged difference of coding time (ΔTime), the averaged PSNR difference (ΔPSNR) and the averaged bit rate difference

(ΔBIT). In order to evaluate the time saving of the fast RDO algorithm, the time difference is defined as follows.

Table 5.2. Test conditions

Frame type	IPPP...
Frame rate	30 fps
Slice mode	OFF
RDO	ON
Hardmard	OFF
Search range	32
Restrict search range	No restriction
Symbol mode	CABAC
Partition mode	No data partition
Out File mode	Annex B

For QP_i ($i = 1 \dots 4$), let $T_{JM,i}$ denote the coding time used by JM8.2 encoder and $T_{FR,i}$ be the time taken by the fast RDO algorithm, the difference of coding time is defined as

$$\Delta\text{Time}_i = \frac{T_{FR,i} - T_{JM,i}}{T_{JM,i}} \times 100\% \quad (5.8)$$

Thus the average difference of coding time is as follows:

$$\Delta\text{Time} = \frac{1}{4} \sum_{i=1}^4 \Delta\text{Time}_i \quad (5.9)$$

PSNR and bit rate differences are calculated according to the numerical averages between the RD-curves derived from JM8.2 encoder and the fast RDO algorithm, respectively. The detailed procedures in calculating these differences can be found in [73], which is recommended by JVT Test Model Ad Hoc Group [74]. Note that PSNR and bit rate

difference should be regarded as equivalent, that is, there is either the increase in PSNR or the decrease in bit rate — not both at the same time.

The experimental results with $\alpha = 0.3$ are given in Table 5.3. As can be seen from Table 5.3, our algorithm has achieved a significant saving in the average encoding time compared to JM8.2, while at the same time the loss of video quality is negligible.

Table 5.3. Experimental results, $\alpha = 0.3$

Sequence	Format	Δ PSNR (dB)	Δ BIT (%)	Δ Time (%)
Akiyo	QCIF	0.022	-0.362	-47.475
Carphone	QCIF	-0.052	1.086	-27.243
Claire	QCIF	-0.116	1.890	-49.062
Coastguard	QCIF	-0.090	2.551	-24.268
Container	QCIF	-0.116	2.300	-45.137
Foreman	QCIF	-0.139	2.472	-22.850
Akiyo	CIF	-0.004	0.118	-50.335
Bus	CIF	-0.109	2.281	-22.086
Mobile	CIF	-0.095	2.209	-20.604
Paris	CIF	-0.093	1.884	-32.327
Stefan	CIF	-0.099	2.037	-28.064
Tempete	CIF	-0.117	2.921	-23.327

Table 5.4 shows the detailed performance results for QCIF sequence “Foreman” for different QPs with $\alpha = 0.3$. As QP increases, the amount of time saving increases. This is because in this case, the probability of SKIP mode increases. Since SKIP mode has the highest priority, it is checked prior to other coding modes. Thus the time saving will be more significant than in the case of lower QPs.

Table 5.4. Performance details for sequence “Foreman”

QP	JM8.2			Proposed			ΔTime_i (%)
	PSNR (dB)	Bit rate (kbps)	Time (s)	PSNR (dB)	Bit rate (kbps)	Time (s)	
28	36.35	141.51	680.084	36.23	142.20	562.421	-17.301
32	33.55	86.96	678.553	33.44	87.26	542.465	-20.056
36	30.92	54.38	679.362	30.82	54.53	511.468	-24.713
40	28.34	34.29	687.522	28.21	34.83	485.872	-29.330

1	1	1	1	4	5	5	3	2	4	2
1	1	1	2	1	1	5	1	2	2	3
2	1	1	1	1	1	3	5	2	2	4
1	1	5	2	2	4	2	5	2	2	2
4	2	5	5	2	1	3	4	2	2	4
1	1	3	5	1	1	1	4	1	2	2
1	1	2	5	5	3	3	2	2	2	4
1	1	1	3	5	5	4	3	2	2	2
4	4	2	2	4	4	5	5	5	3	3

1	1	1	1	3	5	5	2	2	3	4
3	3	2	2	1	1	5	2	2	2	2
2	1	1	2	1	1	2	5	3	2	1
1	2	5	3	2	5	2	5	3	2	1
1	2	3	5	3	1	2	4	2	1	1
1	1	2	4	3	1	2	1	2	2	3
1	1	2	4	3	3	5	1	2	2	2
1	1	1	5	5	5	4	4	3	2	2
4	4	2	2	5	4	4	5	5	3	5

(a) Macroblock modes by JM8.2

(b) Macroblock modes by proposed algorithm

Figure 5.8. The mode distribution for the 45th frame of sequence “Foreman”

Figure 5.8 shows the coding modes of different macroblocks in the 45th frame of QCIF sequence “Foreman” where $QP = 28$ and $\alpha = 0.3$. Figure 5.8(a) is the result without using fast mode decision scheme; Figure 5.8(b) is the result with the proposed fast RDO algorithm. It can be seen from these figures that there are nearly 60% macroblocks that have exactly the same coding modes, and the others are having the similar coding modes. We define the two coding modes are similar if their block sizes are

next to each other. For example, INTER8×8 is similar to INTER8×16 and INTER16×8. Although not all the modes are the same, the PSNR loss of the proposed algorithm is only 0.042dB. This high similarity in macroblock coding modes of these two schemes shows that the proposed fast RDO can quickly find the optimal or near optimal mode without high quality loss.

Figure 5.9 shows the original image and reconstructed images with JM8.2 and proposed method. Comparing Figure 5.9(b) and 5.9(c), it can be seen that the reconstructed image with full mode decision method and the reconstructed image with fast mode decision method are much similar. This shows that although our proposed fast mode decision method achieves 22.85% of average time saving as mentioned in Table 5.3, the quality loss of reconstructed image with our proposed method is very small when compared with the reconstructed image with full mode decision method. Thus our proposed fast mode decision method is effective.

Figure 5.9. The 45th frame of sequence “Foreman”

Table 5.5 gives the experimental results with $\alpha = 0.5$. Comparing Table 5.3 with Table 5.3, it can be seen that when α increases, the average time saving decreases.

Table 5.5. Experimental results, $\alpha = 0.5$

Sequence	Format	Δ PSNR (dB)	Δ BIT (%)	Δ Time (%)
Akiyo	QCIF	-0.010	0.198	-35.087
Carphone	QCIF	-0.057	1.164	-22.045
Claire	QCIF	-0.049	0.736	-36.328
Coastguard	QCIF	-0.057	1.596	-20.265
Container	QCIF	-0.082	1.592	-38.648
Foreman	QCIF	-0.099	1.763	-18.441
Akiyo	CIF	-0.022	0.429	-37.436
Bus	CIF	-0.067	1.383	-16.815
Mobile	CIF	-0.067	1.563	-17.156
Paris	CIF	-0.065	1.305	-26.405
Stefan	CIF	-0.055	1.116	-23.789
Tempete	CIF	-0.080	1.960	-18.431

Figure 5.10(a) and 5.10(b) give the results of average PSNR difference versus α and average time difference versus α , respectively. In Figure 5.10(a), for high motion sequences “Foreman” and “Stefan”, the average PSNR difference increases as α increases. In Figure 5.10(b), for all the sequences, the average time difference decreases as α increases. This shows that α can be used as a control parameter for the tradeoff between the reconstructed video quality and computational complexity. If we want to save more time, we can decrease α . Otherwise α can be increased to retain the coding quality.

(a) Average PSNR difference vs. α

(b) Average time difference vs. α

Figure 5.10. Average PSNR difference and average time difference

In Table 5.6 to Table 5.8, we compare our proposed algorithm with that of Grecos *et al* [65]. The reason why we choose Grecos's algorithm is that both performed the experiments with the same software (JM8.2) and both employed the same frame type (IPPP). Since the Δ PSNR used in Grecos's algorithm was in percentage (%) and only three QPs (QP = 28, 32, 40) were selected in Grecos's algorithm, we also obey the same

rule in comparison and give the results at QP = 28, 32 and 40 only. From these tables, it can be seen that except for sequence “Container”, our proposed algorithm achieves more time saving with acceptable loss of PSNR and bit rate compared to that of Grecos’s algorithm. Figure 5.11 to Figure 5.13 give three examples of the average PSNR difference and average time difference comparing between the Grecos’s algorithm and proposed algorithm. For example, from Figure 5.11, it can be seen that if we adjust α properly so that both algorithms produce the same PSNR loss of about 0.08%, our proposed algorithm will achieve about 20% average time saving as against only about 16% average time saving by Grecos’s algorithm.

Table 5.6. Comparison between Grecos’s algorithm and proposed algorithm, $\alpha = 0.3$

Sequence	QP	Δ PSNR (%)		Δ BIT (%)		Δ Time (%)	
		Grecos	Proposed	Grecos	Proposed	Grecos	Proposed
Carphone	28	-0.05	-0.08	-0.55	0.03	-10.27	-21.92
	32	-0.23	-0.29	-0.18	0.03	-14.33	-25.46
	40	0.03	-0.07	-0.15	0.13	-24.18	-33.30
Coastguard	28	-0.02	-0.21	-0.27	0.20	-7.91	-17.31
	32	-0.16	-0.29	-0.39	-0.11	-13.55	-22.08
	40	-0.19	-0.27	-0.79	0.60	-26.27	-30.54
Container	28	-0.08	-0.36	-0.07	0.55	-39.67	-36.37
	32	-0.12	-0.39	-1.92	0.21	-46.18	-44.68
	40	-0.03	-0.43	0.97	0.26	-50.77	-50.34
Foreman	28	-0.08	-0.33	-0.37	0.49	-9.22	-17.92
	32	-0.03	-0.33	0.13	0.34	-11.99	-20.17
	40	-0.03	-0.46	-0.05	0.16	-19.2	-28.82
Stefan	28	-0.05	-0.20	-0.2	0.34	-8.69	-23.99
	32	-0.09	-0.28	-2.47	0.16	-11.5	-26.36
	40	-0.26	-0.23	-0.23	0.19	-22.77	-32.37

Table 5.7. Comparison between Grecos's algorithm and proposed algorithm, $\alpha = 0.5$

Sequence	QP	Δ PSNR (%)		Δ BIT (%)		Δ Time (%)	
		Grecos	Proposed	Grecos	Proposed	Grecos	Proposed
Carphone	28	-0.05	-0.05	-0.55	0.27	-10.27	-17.43
	32	-0.23	-0.21	-0.18	-0.15	-14.33	-20.15
	40	0.03	-0.07	-0.15	0.00	-24.18	-27.59
Coastguard	28	-0.02	-0.15	-0.27	0.03	-7.91	-14.15
	32	-0.16	-0.16	-0.39	0.11	-13.55	-18.58
	40	-0.19	-0.08	-0.79	0.35	-26.27	-25.40
Container	28	-0.08	-0.22	-0.07	0.60	-39.67	-30.28
	32	-0.12	-0.27	-1.92	0.36	-46.18	-35.48
	40	-0.03	-0.11	0.97	-0.13	-50.77	-45.39
Foreman	28	-0.08	-0.17	-0.37	0.58	-9.22	-13.83
	32	-0.03	-0.24	0.13	0.29	-11.99	-16.35
	40	-0.03	-0.32	-0.05	0.99	-19.2	-23.48
Stefan	28	-0.05	-0.08	-0.2	0.36	-8.69	-18.33
	32	-0.09	-0.15	-2.47	0.09	-11.5	-22.79
	40	-0.26	-0.11	-0.23	0.25	-22.77	-28.34

Figure 5.11. Average PSNR difference and time difference for sequence “Carphone”

Table 5.8. Comparison between Grecos's algorithm and proposed algorithm, $\alpha = 0.7$

Sequence	QP	Δ PSNR (%)		Δ BIT (%)		Δ Time (%)	
		Grecos	Proposed	Grecos	Proposed	Grecos	Proposed
Carphone	28	-0.05	0.00	-0.55	0.26	-10.27	-11.09
	32	-0.23	-0.03	-0.18	0.31	-14.33	-14.64
	40	0.03	0.03	-0.15	-0.13	-24.18	-21.35
Coastguard	28	-0.02	-0.15	-0.27	-0.11	-7.91	-10.27
	32	-0.16	-0.10	-0.39	0.22	-13.55	-13.31
	40	-0.19	-0.15	-0.79	0.09	-26.27	-18.96
Container	28	-0.08	-0.08	-0.07	0.77	-39.67	-23.30
	32	-0.12	-0.12	-1.92	0.31	-46.18	-23.61
	40	-0.03	-0.11	0.97	0.00	-50.77	-30.70
Foreman	28	-0.08	-0.11	-0.37	0.59	-9.22	-9.93
	32	-0.03	-0.03	0.13	0.80	-11.99	-11.12
	40	-0.03	-0.35	-0.05	0.50	-19.2	-16.64
Stefan	28	-0.05	-0.06	-0.2	0.37	-8.69	-5.06
	32	-0.09	-0.03	-2.47	0.03	-11.5	-15.19
	40	-0.26	0.00	-0.23	0.63	-22.77	-23.60

Figure 5.12. Average PSNR difference and time difference for sequence "Container"

Figure 5.13. Average PSNR difference and time difference for sequence “Stefan”

5.6 Summary

In this chapter, we have proposed a fast RDO algorithm based on the occurring probability of different coding modes. The coding mode which has higher occurring probability will be tried first in the RDO process. Once the RD cost of the coding mode has met the early termination criterion, the RDO process will be stopped immediately without testing the rest of coding modes in the priority queue, thus significant time saving can be achieved. By adjusting a threshold which is proportional to the RD cost of the previously encoded frame, we can achieve good tradeoff between time saving and PSNR loss, and thus this approach is scalable. Simulation results have shown that our proposed algorithm achieves significant time saving with negligible PSNR loss when compared with JM8.2.

Chapter 6

Conclusion and Recommendations

6.1 Conclusion

As the most powerful video coding standard in history, H.264 employs RDO to achieve the best coding efficiency between rate and distortion. In this thesis, several efficient RDO techniques have been proposed for three major tasks that arise in the implementation of RDO for H.264. These tasks include rate control, perceptual bit allocation, and fast mode decision. By using these efficient RDO techniques, the RD performance of coded video sequences can be improved and the computational complexity can be reduced so that the improved codec can be used for real time application.

To solve the QP dilemma problem in H.264 rate control, we have proposed a minimum distortion based mode decision process that estimates QP after the full-search of motion estimation. We have also proposed a simple motion complexity measure to quantize the complexity of a frame's motion contents. Based on motion complexity and minimum distortion based mode decision, we have presented a new rate control scheme for H.264. In this scheme, we employ motion complexity to estimate target bits so that the bit allocation is in accordance with the complexity of frame's motion contents, and

we estimate QP according to the macroblock's own information. Our experimental results have shown that our rate control scheme is simple and effective. It outperforms the existing H.264 proposal for low motion sequences, and for high motion sequences at scene changes and high motions.

To achieve better perceptual video quality under bit rate constraint, we have proposed a perceptual mode decision algorithm by updating Lagrangian multiplier according to the perceptual characteristics of video contents, so that more bits are allocated to edge patterns. Based on the perceptual mode decision algorithm, we have presented a perceptual bit allocation scheme for H.264. For medical video application, we have also proposed ROI bit allocation scheme. Experimental results have shown that our proposed perceptual bit allocation scheme outperforms H.264 rate control for medical video compression.

To reduce computational complexity for real time application, we have proposed a fast RDO algorithm based on the occurring probability of different coding modes. The coding mode which has higher occurring probability will be tried first in the RDO process. Once the RD cost of the coding mode has met the early termination criterion, the RDO process will be stopped immediately without testing the rest of coding modes in the priority queue, thus significant time saving can be achieved. By adjusting a threshold which is proportional to the RD cost of the previously encoded frame, we can achieve good tradeoff between time saving and PSNR loss, and thus this approach is scalable. Simulation results have shown that our proposed algorithm achieves significant time saving with negligible PSNR loss when compared with JM8.2.

6.2 Recommendations

Scalable Video Coding (SVC) [84] is the extension of H.264. It enables the transmission and decoding of partial bit streams to provide video services with lower temporal or spatial resolutions or reduced fidelity while retaining a reconstruction quality that is high relative to the rate of the partial bit streams. It has achieved significant improvements in coding efficiency with an increased degree of supported scalability relative to the scalable profiles of prior video coding standards [85].

As in H.264, rate control and fast mode decision are also necessary requirements in SVC for real time applications. In order to improve coding efficiency in comparison with different spatial resolutions, additional inter layer prediction mechanism is incorporated in SVC. Under such a new mechanism, it is not a trivial task for simply applying the single layer rate control and fast mode decision methods for SVC. Therefore new methods that consider both the correlation between inter layers and within intra layers should be designed for SVC rate control and fast mode decision.

Anselmo and Alfonso proposed a rate control scheme for SVC that implements rate control by output buffer management [86]. The disadvantage of this scheme is that it only considers intra layer rate control and does not utilize the correlation between different layers. Liu *et al.* proposed a rate control scheme for spatial scalability and coarse gain scalability (CGS) of SVC [87]. In the scheme, MAD is either predicted from the previous frame of the current layer or from the current frame of the previous layer. Li *et al.* proposed a fast mode decision scheme for spatial scalability, temporal scalability and CGS of SVC [88]. It makes use of the mode distribution correlation between the base

layer and enhancement layers. Lin *et al.* proposed a fast mode decision scheme for CGS of SVC [89]. Mode decision is performed depending on a premade look-up table that reduces the number of possible modes. Lee *et al.* proposed a fast mode decision scheme for temporal scalability of SVC [90]. Correlation of corresponding motion compensated blocks is considered and hypothesis test is performed so as to reduce computation time. Although these methods have achieved better performance for SVC, the shortcoming still exists at that no method has fully utilized the spatial correlation and temporal correlation. Therefore how to extend our RDO technique to SVC is a valuable topic of future research.

The extension of our RDO techniques to SVC may be a combination of RDO techniques for inter layer and intra layer. Depending on different layers, different RDO strategies may be proposed. For intra layer rate control and fast mode decision, we may extend our efficient minimum distortion based rate control algorithm and priority based fast mode decision algorithm for H.264 to SVC directly. For inter layer, we may investigate RDO algorithms that utilize the spatial correlation between different inter layers. Such work may be done based on the statistical analysis of QP and mode distribution between spatially correlated macroblocks in different layers. In order to solve QP dilemma problem, QP estimation may be implemented by not only using QP model but also considering the relationship between current macroblock and its corresponding macroblock in the lower layer. In order to reduce computational complexity, fast mode decision may be implemented by looking up a priority table and performing early termination operation. Since the number of candidate coding modes in the priority table is reduced as compared with that of the complete modes, time saving may therefore be achieved.

Author's Publications

Journal Papers

1. Hongtao Yu, Feng Pan, and Zhiping Lin, "Content adaptive rate control for H.264," *International Journal of Innovative Computing, Information and Control*, vol. 1, no. 4, pp. 685-700, Dec. 2005.
2. Hongtao Yu, Zhiping Lin, and Feng Pan, "Applications and improvement of H.264 in medical video compression," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 12, pp. 2707-2716, Dec. 2005.
3. Feng Pan, Hongtao Yu, and Zhiping Lin, "Scalable fast rate-distortion optimization for H.264/AVC," *EURASIP Journal on Applied Signal Processing, special issue on Advanced Video Technologies and Applications for H.264/AVC and Beyond*, vol. 2006, Article ID 37175, pp. 1-10, Oct. 2006.
4. Hongtao Yu and Zhiping Lin, "A DCT-based method for automatic change detection in multi-temporal SAR images," submitted to *IEEE Transactions on Geoscience and Remote Sensing*.

Conference Papers

1. Hongtao Yu, Feng Pan, and Zhiping Lin, "A new bit estimation scheme for H.264 rate control," *Proceedings of 2004 IEEE International Symposium on Consumer Electronics*, pp. 396-399, Reading, UK, Sept. 2004.
2. Hongtao Yu, Zhiping Lin, and Feng Pan, "An improved rate control algorithm for H.264," *Proceedings of 2005 IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 312-315, Kobe, Japan, May 2005.
3. Hongtao Yu, Feng Pan, Zhiping Lin, and Yin Sun, "A perceptual bit allocation scheme for H.264," *Proceedings of 2005 IEEE International Conference on Multimedia and Expo*, pp. 313-316, Amsterdam, Netherlands, July 2005.
4. Zhiping Lin, Hongtao Yu, and Feng Pan, "A scalable fast mode decision algorithm for H.264," *Proceedings of 2006 IEEE International Symposium on Circuits and Systems*, pp. 3494-3497, Island of Kos, Greece, May 2006.
5. Hongtao Yu, Zhiping Lin, Ru San Tan, Thu Thao Le, Dhanjoo N. Ghista, "H.264 standard for left ventricle video compression and telediagnosis," *Proceedings of the Third IASTED International Conference on Telehealth*, pp. 193-197, Montreal, Canada, May 2007.
6. Hongtao Yu, Zhiping Lin, Geok Ling Chai, Chye Hwang Yan, and Nalliah Raman, "A DCT-based change detection method for multi-temporal SAR images," *Proceedings of 2008 IEEE Geoscience and Remote Sensing Symposium*, vol. III, pp. 166-169, Boston, USA, July 2008.

7. Hongtao Yu, Zhiping Lin, and Frederick Cheng Tat Teo, “An efficient coding scheme based on image alignment for H.264/AVC,” *Proceedings of 2009 IEEE International Symposium on Circuits and Systems*, pp. 629-632, Taipei, Taiwan, May 2009.

Bibliography

- [1] ITU-T and ISO/IEC JTC1, *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), Version 1: 2003, Version 6: June 2006.
- [2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560-576, July 2003.
- [3] G. J. Sullivan and T. Wiegand, "Video compression — from concepts to the H.264/AVC standard," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 18-31, Jan. 2005.
- [4] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 23-50, Nov. 1998.
- [5] E. H. Yang and X. Yu, "Rate distortion optimization for H.264 interframe coding: a general framework and algorithms," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1774-1784, July 2007.
- [6] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia*, Chichester: John Wiley & Sons Ltd, 2003.

- [7] A. Puri, X. M. Chen, and A. Luthra, "Video coding using the H.264/MPEG-4 AVC compression standard," *Signal Processing: Image Communication*, vol. 19, no. 9, pp. 793-849, Oct. 2004.
- [8] A. Joch, F. Kossentini, and P. Nasiopoulos, "A performance analysis of the ITU-T draft H.26L video coding standard," *Proceedings of the 12th International Packetvideo Workshop*, Pittsburgh, USA, Apr. 2002.
- [9] A. Luthra and P. Topiwala, "Overview of the H.264/AVC video coding standard," *Motorola Internal Report*, 2003.
- [10] S. Saponara, C. Blanch, K. Denolf, and J. Bormans, "The JVT advanced video coding standard: complexity and performance analysis on a tool-by-tool basis," *Proceedings of the 13th International Packetvideo Workshop*, Nantes, France, Apr. 2003.
- [11] T. Wiegand, H. Schwarz, A. Joch, F. Lossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 688-703, July 2003.
- [12] T. Hallbach, "H.264 video compression standard", *Proceedings of the 6th Nordic Signal Processing Symposium*, Bergen, Norway, Oct. 2003.
- [13] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression", *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74-90, Nov. 1998.

- [14] J. H. Wang, Z. B. Chen, Y. He, and Y. D. Chen, "A MAD-based rate control strategy," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 4th Meeting*, Document JVT-D070, Klagenfurt, Austria, July 2002.
- [15] N. Wang and Y. He, "A new bit rate control strategy for H.264", *Proceedings of 2003 Joint Conference of the 4th International Conference on Information, Communications and Signal Processing and 4th Pacific-Rim Conference on Multimedia*, Session 3A2.6, Singapore, Dec. 2003.
- [16] S. K. Kim and Y. S. Ho, "Rate control algorithm for H.264/AVC video coding standard based on rate-quantization model," *Proceedings of 2004 IEEE International Conference on Multimedia and Expo*, Taipei, Taiwan, June 2004.
- [17] Y. C. Wang and J. J. Leou, "A rate control method for H.26L video transmission," *Proceedings of 2003 IEEE International Conference on Multimedia and Expo*, Baltimore, USA, July 2003.
- [18] T. H. Tsai and J. J. Leou, "A rate control method for H.264 video transmission," *Proceedings of 2004 IEEE International Conference on Multimedia and Expo*, Taipei, Taiwan, June 2004.
- [19] S. Milani, L. Celetto, and G. A. Mian, "A rate control algorithm for the H.264 encoder," *Proceedings of the Sixth Baiona Workshop on Signal Processing in Communications*, Baiona, Spain, Sept. 2003.
- [20] Z. H. 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, no. 12, pp. 1221-1236, Dec. 2001.

- [21] I. H. Shin, Y. L. Lee, and H. W. Park, "Rate control using linear rate- ρ model for H.264," *Signal Processing: Image Communications*, vol. 19, no. 4, pp. 341-352, Apr. 2003.
- [22] P. Navakitkanok and S. Aramvith, "Improved rate control for advanced video coding (AVC) standard under low delay constraint," *Proceedings of 2004 IEEE International Conference on Information Technology: Coding and Computing*, vol. 2, pp. 664-668, Las Vegas, USA, Apr. 2004.
- [23] M. Q. Jiang, X. Q. Yi, and N. Ling, "Improved frame-layer rate control for H.264 using MAD ratio," *Proceedings of 2004 IEEE International Symposium on Circuits and Systems*, vol. III, pp. 813-816, Vancouver, Canada, May 2004.
- [24] M.Q. Jiang, "Frame layer bit allocation scheme for constant quality video," *Proceedings of 2004 IEEE International Conference on Multimedia and Expo*, Taipei, Taiwan, June 2004.
- [25] M. Q. Jiang, X. Q. Yi, and N. Ling, "Quantizer control for H.264/AVC streaming over networks," *Journal of Internet Technology*, vol. 5, no. 3, pp. 301-304, July 2004.
- [26] M. Q. Jiang and N. Ling, "On enhancing H.264/AVC video rate control by PSNR-based frame complexity estimation," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 281-286, Jan. 2005.
- [27] M. Q. Jiang and N. Ling, "On Lagrange multiplier and quantizer adjustment for H.264 frame-layer video rate control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 5, pp. 663-669, May 2006.

- [28] M. Q. Jiang and N. Ling, "Low-delay rate control for real-time H.264/AVC video coding," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 467-477, June 2006.
- [29] Z. G. Li, F. Pan, K. P. Lim, G. N. Feng, X. Lin, and S. Rahardja, "Adaptive rate control," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 6th Meeting*, Document JVT-F019, Awaji, Japan, Dec. 2002.
- [30] Z. G. Li, F. Pan, K. P. Lim, G. N. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 7th Meeting*, Document JVT-G012, Pattaya, Thailand, Mar. 2003.
- [31] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu, and Y. Lu, "Adaptive rate control with HRD consideration," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 8th Meeting*, Document JVT-H014, Geneva, May 2003.
- [32] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu, and Y. Lu, "Adaptive rate control for H.264," *Journal of Visual Communication & Image Representation*, vol. 17, no. 2, pp. 376-406, Apr. 2006.
- [33] S. W. Ma, W. Gao and Y. Lu, "Rate control on JVT standard," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 4th Meeting*, Document JVT-D030, Klagenfurt, Austria, July 2002.

- [34] S. W. Ma, W. Gao, Y. Lu, and D. B. Zhao, "Improved rate control algorithm," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 5th Meeting*, Document JVT-E069, Geneva, Swiss, Oct. 2002.
- [35] S. W. Ma, W. Gao, Y. Lu, and H. Q. Lu, "Proposed draft description of rate control on JVT standard," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 6th Meeting*, Document JVT-F086, Awaji, Japan, Dec. 2002.
- [36] S. W. Ma, Z. G. Li, and F. Wu, "Proposed draft of adaptive rate control," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 8th Meeting*, Document JVT-H017, Geneva, Swiss, May 2003.
- [37] S. W. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1533-1544, Dec. 2005.
- [38] ISO/IEC JTC1/SC29/WG11, "Test model 5," Document N0400, Apr. 1993.
- [39] A. E. Mohr, "Bit allocation in sub-linear time and the multiple-choice Knapsack problem," *Proceedings of 2002 Data Compression Conference*, Snowbird, USA, Mar. 2002.
- [40] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, San Francisco: W. H. Freeman, 1979.
- [41] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, no. 10, pp. 1385-1422, Oct. 1993.
- [42] S. Winkler, "A perceptual distortion metric for digital color video," *Proceedings of SPIE*, Vol. 3644, pp. 175-184, San Jose, USA, Jan. 1999.

- [43] J. L. Mannos and J. D. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525-536, July 1974.
- [44] S. Sengupta, S. K. Gupta, and J. M. Hannah, "Perceptually motivated bit-allocation for H.264 encoded video sequences," *Proceedings of 2003 IEEE International Conference on Image Processing*, Barcelona, Spain, Sept. 2003.
- [45] C. J. Tsai, C. W. Tang, C. H. Chen, and Y. H. Yu, "Adaptive rate-distortion optimization using perceptual hints," *Proceedings of 2004 IEEE International Conference on Multimedia and Expo*, Taipei, Taiwan, June 2004.
- [46] C. W. Tang, C. H. Chen, Y. H. Yu, and C. J. Tsai, "A novel visual distortion sensitivity analysis for video encoder bit allocation," *Proceedings of 2004 IEEE International Conference on Image Processing*, pp.3225-3228, Singapore, Oct. 2004.
- [47] M. M. Ghandi and M. Ghanbari, "A Lagrangian optimization rate control algorithm for the H.264/AVC encoder", *Proceedings of 2004 IEEE International Conference on Image Processing*, pp. 123-126, Singapore, Oct. 2004.
- [48] A. Pietrowcew, A. Buchowicz, and W. Skarbek, "Bit-rate control algorithm for ROI enabled video coding," *Proceedings of the 11th International Conference on Computer Analysis of Images and Patterns*, pp.514-521, Versailles, France, Sept. 2005.
- [49] D. Agrafiotis, D. R. Bull, N. Canagarajah, and N. Kamnoonwatana, "Multiple priority region of interest coding with H.264," *Proceedings of 2006 IEEE*

- International Conference on Image Processing*, pp. 53-56, Atlanta, USA, Oct. 2006.
- [50] Y. Liu, Z. G. Li, and Y. C. Soh, "Region-of-interest based resource allocation for conversational video communication of H.264/AVC," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 18, no. 1, pp. 134-139, Jan. 2008.
- [51] B. W. Jeon and J. Y. Lee, "Fast mode decision for H.264," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG 8th Meeting*, Document JVT-J033, Waikoloa, USA, Dec. 2003.
- [52] I. C. Choi, J. Y. Lee, and B. W. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 Part-10 AVC/H.264," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 12, pp. 1557-1561, Dec. 2006.
- [53] Z. Y. Wei and K. N. Ngan, "A Fast Macroblock Mode Decision Algorithm for H.264," *Proceedings of 2006 IEEE Asia Pacific Conference on Circuits and Systems*, pp. 773-776, Singapore, Dec. 2006.
- [54] A. Saha, K. Mallick, J. Mukherjee, and S. Sural, "SKIP prediction for fast rate distortion optimization in H.264," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 3, pp. 1153-1160, Aug. 2007.
- [55] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, G. N. Feng, D. J. Wu, and S. Wu, "Fast mode decision for intra prediction," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 7th Meeting*, Document JVT-G013, Pattaya, Thailand, Mar. 2003.

- [56] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. J. Wu, and S. Wu, "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 7, pp. 813-822, July 2005.
- [57] C. H. Tseng, H. M. Wang, and J. F. Yang, "Improved and fast algorithms for intra 4×4 mode decision in H.264/AVC," *Proceedings of 2005 IEEE International Symposium on Circuits and Systems*, vol. 3, pp. 2128-2131, Kobe, Japan, May 2005.
- [58] J. C. Wang, J. F. Wang, J. F. Yang, and J. T. Chen, "A fast mode decision algorithm and its VLSI design for H.264/AVC intra-prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 10, pp. 1414-1422, Oct. 2007.
- [59] B. G. Kim, "Fast selective intra-mode search algorithm based on adaptive thresholding scheme for H.264/AVC encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 127-133, Jan. 2008.
- [60] K. P. Lim, S. Wu, D. J. Wu, S. Rahardja, X. Lin, F. Pan, and Z. G. Li, "Fast inter mode selection," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) 9th Meeting*, Document JVT-I020, San Diego, USA, Sept. 2003.
- [61] D. J. Wu, F. Pan, K. P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahardja, and C. C. Ko, "Fast intermode decision in H.264/AVC video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 7, pp. 953-958, July 2005.

- [62] C. H. Kuo, M. Y. Shen, and C. C. J. Kuo, "Fast motion search with efficient inter-prediction mode decision for H.264," *Journal of Visual Communication & Image Representation*, vol. 17, no. 2, pp. 217-242, Apr. 2006.
- [63] Z. Zhou, J. Xin, and M. T. Sun, "Fast Motion Estimation and Inter-Mode Decision for H.264/MPEG-4 AVC Encoding", *Journal of Visual Communication & Image Representation*, vol. 17, no. 2, pp. 243-263, Apr. 2006.
- [64] A. C. W. Yu, G. R. Martin, and H. C. Park, "Fast inter-mode selection in the H.264/AVC standard using a hierarchical decision process," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 2, pp. 186-195, Feb. 2008.
- [65] C. Grecos and M. Y. Yang, "Fast inter mode prediction for P slices in the H.264 video coding standard," *IEEE Transactions on Broadcasting*, Vol. 51, No. 2, pp. 256-263, June 2005.
- [66] C. Grecos and M. Y. Yang, "Fast mode prediction for the baseline and main profiles in the H.264 video coding standard," *IEEE Transactions on Multimedia*, vol. 8, no. 6, pp. 1125-1134, Dec. 2006.
- [67] A. C. Yu, N. K. Ngi, and G. R. Martin, "Efficient intra- and inter-mode selection algorithm for H.264/AVC," *Journal of Visual Communications & Image Representation*, vol. 17, no. 2, pp. 322-344, Apr. 2006.
- [68] C. S. Kim and C. C. J. Kuo, "Feature-based intra-/inter coding mode selection for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 4, pp. 441-453, Apr. 2007.

- [69] J. H. Kim and B. G. Kim, "Fast block mode decision algorithm in H.264/AVC video coding," *Journal of Visual Communication & Image Representation*, vol. 19, no. 3, pp. 175-183, Apr. 2008.
- [70] T. H. 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, no. 1, pp. 246-250, Feb. 1997.
- [71] H. L. Lee, T. H. Chiang, and Y. Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 6, pp. 878-894, Sept. 2000.
- [72] JVT, Reference Model JM6.1e, Available: <ftp://standards.polycom.com>.
- [73] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T VCEG 13th Meeting*, Document VCEG-M33, Austin, USA, Apr. 2001.
- [74] JVT Test Model Ad Hoc Group, "Evaluation sheet for motion estimation," Draft version 4, Feb. 2003.
- [75] http://read.pudn.com/downloads101/sourcecode/multimedia/412227/avsnr.c_.htm.
- [76] F. Pan, "Adaptive image compression using local pattern information," *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1837-1845, Dec. 2002.
- [77] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, second edition, Prentice Hall, 2002.
- [78] A. M. Bazen and S. H. Gerez, "Systematic methods for the computation of the directional fields and singular points of fingerprints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 905-919, July 2002.

- [79] ITU-R, "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU-R Recommendation BT.500-11, June 2002.
- [80] P. Lambert, D. De Schrijver, D. Van Deursen, W. De Neve, Y. Dhondt, R. Van De Walle, "A real-time content adaptation framework for exploiting ROI scalability in H.264/AVC," *Proceedings of the 8th International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 442-453, Antwerp, Belgium, Sept. 2006.
- [81] JVT, Reference Model JM12.0, Available: <http://iphone.hhi.de/suehring/tml/download/>.
- [82] JVT, Reference Model JM8.2, Available: <ftp://standards.polycom.com>.
- [83] X. A. Lu, A. M. Tourapis, P. Yin, and J. Boyce, "Fast mode decision and motion estimation for H.264 with a focus on MPEG-2/H.264 transcoding," *Proceedings of 2005 IEEE International Symposium on Circuits and Systems*, pp. 1246-1249, Kobe, Japan, May 2005.
- [84] ITU-T and ISO/IEC JTC1, *Advanced Video Coding for Generic Audiovisual Services*, ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), Version 8 (including SVC extension): July 2007.
- [85] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103-1120, Sept. 2007.
- [86] T. Anselmo and D. Alfonso, "Buffer-based constant bit-rate control for scalable video coding," *Proceedings of 2007 Picture Coding Symposium*, Lisbon, Portugal, Nov. 2007.

- [87] Y. Liu, Z. G. Li, and Y. C. Soh, "Rate control of H.264/AVC scalable extension," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 116-120, Jan. 2008.
- [88] H. Li, Z. G. Li, and C. Y. Wen, "Fast mode decision algorithm for inter-frame coding in fully scalable video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 889-895, July 2006.
- [89] H. C. Lin, W. H. Peng, H. M. Hang, and W. J. Ho, "Layer-adaptive mode decision and motion search for scalable video coding with combined coarse granular scalability (CGS) and temporal scalability," *Proceedings of 2007 IEEE International Conference on Image Processing*, vol. 2, pp. 289-292, San Antonio, USA, Sept. 2007.
- [90] B. S. Lee, M. C. Kim, S. J. Hahm, C. S. Park, and K. S. Park, "A fast mode selection scheme in inter-layer prediction of H.264 scalable extension coding," *Proceedings of 2008 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, Las Vegas, USA, Mar. 2008.