FAST AND ADAPTIVE H.264/AVC VIDEO CODING FOR NETWORK BASED APPLICATIONS

A Dissertation

Presented to

The Faculty of the Graduate School

University of Missouri

In Partial Fulfillment

of the Requirements for the Degree of

Ph. D

---

by

LI LIU

Dr. Xinhua Zhuang, Dissertation Supervisor

DECEMBER 2009

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

FAST AND ADAPTIVE H.264/AVC VIDEO CODING FOR NEWWORK BASED APPLICATIONS

presented by Li Liu, a candidate for the degree of doctor of philosophy,and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Xinhua Zhuang

Professor Ye Duan

Professor Yunxin Zhao

Professor Michael Jurczyk

Professor Zhihai He

ACKNOWLEDGMENTS

I would like to acknowledge my heartfelt gratitude to my advisor, Dr. Xinhua Zhuang. This dissertation was accomplished with his professional knowledge, strong support, and invaluable advices on my PHD study and research. His generous personality and sharp foresight shall continuously inspire me on my future career. My thanks also go to Dr. Michael Jurczyk, Dr. Yunxin Zhao, Dr. Ye Duan, and Dr. Zhihai He for their precious time and efforts to serve as my committee members.

I would like to devote my sincerest and deepest gratitude to my parents. I could never forget their selfless support and warm encouragement since my date of birth. They are the greatest heroes in my life.

Finally, I would like to express my appreciation to my dear colleagues and friends. My life becomes colorful and joyful with your sharing and help.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

FAST AND ADAPTIVE H.264/AVC VIDEO CODING FOR NETWORK BASED APPLICATIONS

Li Liu

Dr. Xinhua Zhuang, Dissertation Supervisor

ABSTRACT

As the state of the art video coding standard, H.264/AVC achieves significant coding performance gain comparing to its predecessors. Nevertheless, the advance comes at huge complexity increase of the encoder, which may hinder its applications to real world. In addition, network applications impose some unique requirements on existing video coding algorithms. For instance, a variable bit rate output of the encoder has to be tuned into a constant rate bit stream to fit transmission channel bandwidth.

In this dissertation, two issues related to H.264/AVC video coding are to be addressed: coding complexity and bandwidth adaption (rate control), and corresponding solutions are provided. To reduce the coding complexity, the original mode decision process in H.264/AVC reference software is optimized for fast implementation. Moreover, two rate control algorithms are given to address different requirements of rate control: quality fluctuation reduction and accurate basic unit quantization decision.

Experiments are performed to test and validate the proposed algorithms. The results show that the proposed algorithms provide efficient solutions to the above problems and facilitate H.264/AVC coding standard for practical deployment.

Chapter 1
Introduction

1.1 Background

In past decades, with the development of digital techniques and the wide spread of the Internet, digital video and its related applications have become popular in daily life. Due to their nature, uncompressed digital video data in the original format can take staggering amount of room to store or transmit. Data compression techniques offer a possible solution to represent the vast amount of video data in a more compact and robust way so that their storage and transmission can be realized in much reduced costs in terms of size, bandwidth and power.

Data compression is achieved by removing redundancy information from the original signal. The history of data compression can be traced back to the late 1940s and early 1950s, when Claude Shannon created the fields of information theory [1,2]. The fundamental papers he published provided theoretical background for these data compression techniques being developed later on. According to whether distortion is introduced in the reconstructed data, current data compression techniques can be classified into two categories: lossless and lossy compression [3].

Lossless compression techniques usually exploit the statistical redundancy to represent the original signal without distortion so that the original data can be reconstructed from the compressed data at no difference. Some most widely used lossless data compression tools include run length coding, dictionary coders, Huffman coding [4], arithmetic coding [5], etc. Huffman coding and arithmetic coding are two most widely used lossless coding techniques. Huffman coding algorithm is created by David Huffman in 1952 [4], and the principle of arithmetic coding is first proposed by Peter Elias in the early 1960s. The basic idea of these two coding methods can be concluded as converting the original signal into another representation that maps frequently used symbols to a representation of fewer bits and infrequently used symbols to a representation of more bits, with a goal of using fewer bits in total. While Huffman coding can only assign symbols to codes of integer bits, arithmetic coding is capable of mapping the entire input file to one code and thus can obtain higher coding efficiency.

Instead of aiming at perfect reconstruction of the original data, lossy compression techniques allow to retrieve an approximation of the original data from its compressed format, in exchange for a higher compression ratio. Transform coding is a typical lossy compression method. First, original data is transformed into frequency domain, with an expectation that energy is compacted at some certain areas. After transforming, a subsequent quantization step is used to introduce distortion to transform coefficients. The more distortion is introduced, the less number of coding bits or higher compression ratio can be obtained. Over the years, a variety of linear transform methods have been

developed, including Hadamard Transform, Discrete Cosine Transform (DCT) [6], Discrete Wavelet Transform (DWT) [7] and many more, each with its own advantages and disadvantages.

Video coding is one of the most important applications of data compression. In literature, video coding and video compression are usually used interchangeably as the primary goal of video coding is to represent video data in less number of bits with data compression techniques. Video coding can be attained by taking advantage of spatial redundancy within a picture and temporal redundancy between neighboring pictures in a video sequence. In general, current video coding schemes all follow the hybrid coding structure framework, where a temporal prediction is used to reduce inter frame redundancy and then a transform coding is applied to reduce spatial redundancy in frequency domain.

To meet the industry requirement of standardizing existing video techniques, video coding standards were developed by two international organizations, ITU-T and ISO/IEC. The standards approved by ISO/IEC are called MPEG family, whose applications range from consumer video on CD (MPEG-1 1990) [8] to broadcast/storage standard or high definition TV (MPEG-2 1994) [9] and coding rectangular and arbitrary shaped objects (MPEG-4 Visual or part 2 1998) [10]. On the other hand, H.26x series of video standards published by ITU-T focus on improving the coding efficiency for bandwidth restricted telecommunication applications as the increasing number of video services and applications over the networks are creating great needs for better coding efficiency.

ITU-T published its first video coding standard H.261 [11] in 1990, and in 1995, ITU-T evolved H.263 video coding standards [12] (and later enhancements of H.263 known as H.263+ and H.263++) with higher compression ratio.

The state of the art video coding standard H.264/AVC [13] is a joint development by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG). In the early 1998, VCEG ITU-T SG16 Q.6 issued a call for proposals on a project called H.26L. Later on December of 2001, VCEG and MPEG formed Joint Video Team (JVT) to finalize the draft of the video coding standard as H.264/AVC in May 2003. Similar to previous coding standards, H.264/AVC is built on block-based motion compensated prediction, followed by DCT transform, quantization and entropy coding of DCT coefficients. However, the adoption of new techniques, such as 1/4 pixel prediction, multiple reference frame, in-the-loop deblocking filtering and RD (Rate Distortion) optimization, has enormously improved the coding efficiency of H.264/AVC encoder. As reported, H.264/AVC is able to yield a bit rate savings of about 50% over MPEG-2 for the same video quality [14,15].

1.2 Motivation and Contribution

While H.264/AVC coding standard provides powerful tools to obtain much better video compression performance, implementation of the standard for real world applications is not quite straightforward as several practical issues have to be considered, such as low complexity while maintaining high performance, rate scalability, error

resilience in packet loss environment and multimedia security. In this dissertation, two important issues, coding complexity and bandwidth adaption (rate control), will be addressed by the proposed solutions.

It's known that the great performance gain of H.264/AVC comes at much higher computational cost. Some introduced high complexity H.264/AVC features may hinder its implementation for real time applications, such as software platforms with limited power and memory resources. Therefore, it's highly desirable to optimize the H.264/AVC with reduced complexity at negligible performance compromise.

Since the optimization of H.264/AVC codec at lower level may heavily depend on the characteristics of platforms (e.g. DSP, FPGA, ASIC) on which it is mapped, in this dissertation, the study of optimization will focus at algorithm level that is applicable for any platform. With a comparable, meaningful configuration without one of some specific tools as the benchmark, J. Ostermann et al. reported complexity analysis of some major H.264/AVC encoding tools [16]. As noticed, among these tools, RD (Rate Distortion) optimization is the one imposing the highest complexity increase which produces up to 9% bit savings with a data transfer increase in the order of 120%.

In order to reduce the computational complexity rising from RD optimization, in the first part of this work:

1. We propose a bit estimation method to reduce computational load of entropy coding in RD optimization process. To make mode decision under RD optimization framework, JM reference software counts the number of bits to

code residual signal through the entropy coding step. By estimating coding bit for each mode using a reasonable model, the proposed method accelerates the implementation of RD optimization by skipping some necessary entropy coding step.

2. We deliver a fast mode decision algorithm that is capable of early eliminating possible inter mode choices for error resilient intra refresh decision. In packet loss environment, intra refresh is a simple and powerful tool to fight against error propagation. To estimate unknown channel errors, at encoder side, the optimal intra refresh by JM reference software is realized by simulation of *k*-pairs of decoders and channels at the cost of great complexity burden. With our proposed method, the encoder speed is accelerated by up to 10 times while similar or better reconstructed video quality is obtained.

Besides higher compression ratio, network based video applications impose more requirements on existing coding algorithms. To reliably deliver video from the sender to the receiver, transmission channel characteristics have to be considered. For instance, variant bit rate output from the encoder has to be tuned into a constant rate bit stream to fit the transmission channel bandwidth by rate control algorithms. In the second part of this dissertation, rate control for H.264/AVC video communication is studied and new algorithms are proposed.

1. We designed a frame level bit rate allocation algorithm based on a two pass H.264/AVC encoder structure to reduce quality fluctuation resulted from

constant rate bit allocation rate control. The method obtains residual signal from the first coding round, and then assigns frame level bit budget according to target distortion. Experimental results show that the algorithm can reduce standard deviation of PSNR by up to 1/3 at minor PSNR impact.

2. We present an adaptive and efficient rate control framework for low delay H.264/AVC video communication. To realize quantization decision improvement, two novel R-Q models are proposed for I frame and P frame, respectively. The P frame R-Q model is established through the ρ domain rate control theory. The I frame R-Q model measures frame level coding activity for quantization decision without performing computationally intensive intra prediction. Moreover, the I frame R-Q model also helps to alleviate frame skipping problem by adjusting I frame coding bit according to the buffer status.

1.3 Outline of the Dissertation

The dissertation consists of 8 chapters.

Chapter 1 briefs the background and gives motivation and contribution of the dissertation work.

Chapter 2 and chapter 3 are study on H.264/AVC encoder optimization. Chapter 2 proposes a method to reduce RD optimization complexity though estimating entropy coding bits. Chapter 3 introduces a fast mode decision algorithm for error resilient intra decision.

Chapter 4 summarizes existing rate control work. In the next two chapters, research contributed to H.264/AVC rate control is presented. Chapter 5 delivers a frame level bit rate allocation algorithm to alleviate the quality fluctuation problem. Chapter 6 introduces a new rate control framework with the P frame square root R-Q model and the I frame source rate model.

Chapter 7 surveys the status of the next generation video coding standard. New coding techniques are summerized by their roles in the coding process.

Chapter 8 concludes the dissertation. A summary of completed research and a plan of future work are given.

Chapter 2
CABAC based Bit Estimation for fast RD Optimization

At macroblock level, H.264/AVC offers up to 7 inter coding modes, 4 intra 16x16 coding modes, and 9 intra 4x4 coding modes, where the best mode decision is based on an efficient RD optimization process. According to our test results, compared with an coding process without RD optimization, RD optimized mode decision can improve PSNR quality for around 0.2~0.3 dB while reducing bit rate up to 10%.

To obtain actual distortion and coding bit for RD cost computation, RD optimization process goes through DCT/Q, IQ/IDCT, and entropy coding step for all possible coding modes. Though RD optimization improves coding performance, it also incurs a considerable complexity increase of the encoder.

Recently, several fast coding methods by replacing the entropy coding step in RD optimization process with coding bit estimation were proposed in previous work [17,18]. H.264/AVC standard supports two different entropy coding methods: CAVLC and CABAC. As known, compared with CAVLC, CABAC provides higher coding efficiency at average bit rate saving of 9%-14% [19]. However, these previous study on fast bit estimation and their experimental results are all based on CAVLC entropy coding. When

applied to entropy coding process using CABAC, these methods cannot be very accurate.

In this chapter, we study H.264/AVC CABAC entropy coding and then propose an adaptive model for fast and accurate CABAC bit estimation so as to reduce RD optimization complexity. The rest part of this chapter is organized as follows. Section 2.1 and 2.2 reviews the RD optimization techniques and CABAC coding in H.264/AVC, respectively. Section 2.3 proposes the CABAC based fast bit estimation method. Section 2.4 is the experimental result. Section 2.5 concludes this chapter.

2.1 RD Optimization in H.264/AVC

Video data in their original format contain a large amount of correlation, both within the same picture and among pictures. This fact can be utilized to predict current signal from past so that only the difference between the prediction and current signal needs to be coded. Based on this premise, prediction is introduced in a block-based hybrid video encoder to exploit redundancy. The prediction coding modes for each block can be classified as [20]:

INTRA coding: A prediction mode in which the picture content of a macroblock region is represented without reference to a region in any previously decoded picture.

INTER coding: A prediction mode in which the picture content of a macroblock region is represented as the sum of a motion-compensated prediction using a motion vector, plus (optionally) a decoded residual difference signal representation.

It was demonstrated that by adding more coding efficient modes in the rate-distortion

sense, the overall performance increases [20]. Comparing to its predecessors, more prediction modes are supported by H.264/AVC. At macroblock level, the final coding mode is to be chosen from a mode set containing 7 inter coding modes, 4 intra 16x16 coding modes, and 9 intra 4x4 coding modes. For each possible coding mode, its coded residual signal has different R-D characteristic, and the goal of a mode selection with RD consideration is to optimize an overall fidelity defined by

$$\min\{D\}, \textit{subject to } R < R_e \tag{2.1}$$

That is to minimize distortion D, subject to a constraint R_e, where D denotes the source distortion by quantization, i.e. difference between the original signal and the reconstructed signal, and R represents the resulting coding bit, including coding bit of macroblock header, motion information, and transformed coefficients.

The constrained optimization task in (2.1) can be solved using a classical Lagrange multiplier method [20]. The method converts the original constrained problem into an unconstrained one with a formulation given as

$$\min\{J\}, \textit{where } J = D + \lambda R \tag{2.2}$$

where an optimal mode should be chosen to minimize the rate distortion cost function for a particular λ. In H.264/AVC, Lagrange multiplier λ is associated with quantization parameter QP by

$$\lambda_{\mathrm{mod}e} = m \times 2^{QP/3} \tag{2.3}$$

where m is a constant.

As shown in Fig. 2.1, for all possible coding modes, this RD optimization process

must go through DCT/Q, IQ/IDCT, and entropy coding step to get the corresponding distortion and coding bit. Nevertheless, coding results from this RD optimization are only intermediate and will not be used for the final output. Hence, computational load caused by the RD optimization process is expected to be reduced with a small impact on final coding performance.

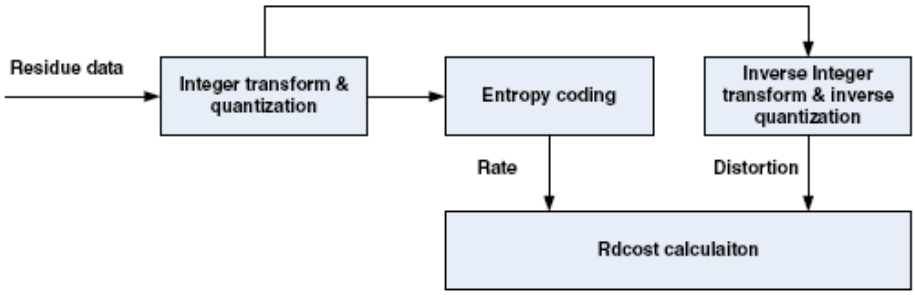


Fig. 2.1. Illustration of H.264/AVC RD optimization process [18]

2.2 CABAC in H.264/AVC

Two possible entropy coding methods are supported by H.264/AVC standard: CAVLC and CABAC. In this section, we brief H.264/AVC CABAC implementation as the coding bit estimation model in 2.3 is based on CABAC coding process.

In H.264/AVC, CABAC based entropy coding process consists of three elementary steps [19]: 1) Binarization; 2) Context modeling; 3) Binary arithmetic coding. In the first step, a given nonbinary valued syntax element is uniquely mapped to a binary sequence, a so-called bin string. Prior to the binary arithmetic coding process, the bin string enters the context modeling stage, where a probability model is selected such that corresponding

choice may depend on previously encoded syntax elements or bins. After assigning a context model, the bin value along with its associated model is passed to the regular coding engine, where the final stage of arithmetic encoding together with a subsequent model updating takes place.

In H.264/AVC, CABAC is used to code two categories of syntax elements:

1) The elements related to macroblock type, submacroblock type, or information of prediction modes.

2) All residual data elements.

To help understand estimating residual data coding bit, we give more details on the residual entropy coding process. After prediction to get residual data, a coded block flag (coded_block_flag) is transmitted for a given block of transform coefficients. If the coded block flag is insignificant, i.e. zero, no further information is sent for the block; otherwise, a significance map specifying the positions of significant, i.e. nonzero coefficients, is encoded. Afterwards, the coefficient level magnitude as well as the sign is coded for each significant transform coefficient and transmitted in reverse scanning order. In Fig. 2.2, a flow diagram is given to show the CABAC coding scheme for a transform coefficient block.

2.3 Fast Bit Estimation for Block Residual

From the description of residual entropy coding process in section 2.2, it can be seen that block-wise bit consumption results from three types of symbols: the coded block flag,

the significance map, and the level information. In this section, we analyze these coding parts to find a model for each individual part.

2.3.1 Coded Block Flag

The coded_block_flag is a one-bit symbol indicating if there are significant coefficients inside a single block of transform coefficients. We assign 1 bit to estimate the number of bits to code this symbol.

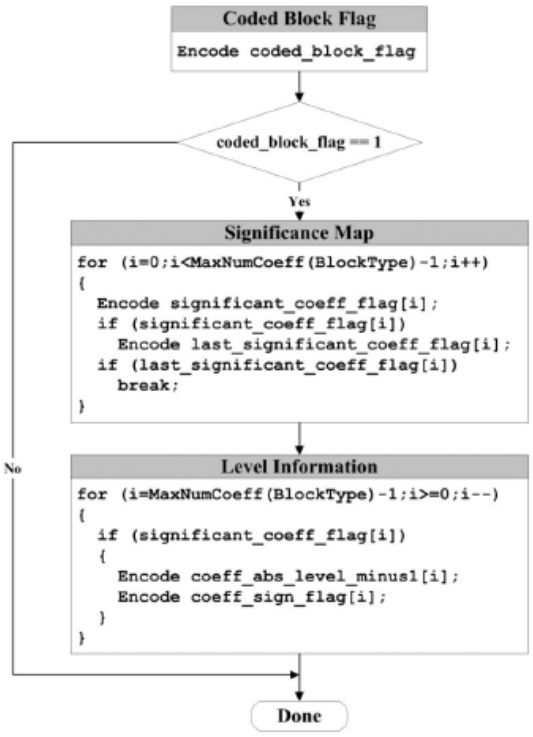


Fig. 2.2. Flow diagram of the CABAC coding scheme

for a block of transform coefficients [19]

2.3.2 Significance Map

If there are significant coefficients inside current transform coefficient block, a

binary-valued significance map will be coded. For each coefficient in a specified scanning order, a one-bit symbol significant_coeff_flag is transmitted. If symbol significant_coeff_flag is one, i.e., current scanning position contains a significant coefficient, a further one bit symbol last_significant_flag is sent, which indicates if current significant coefficient is the last one inside the block or if further significant coefficients fellow. After map scan, both significant_coeff_flag and last_significant_flag are sent to CABAC coding engine.

Above process shows that, bit to code significant map is associated with two factors: the number of nonzero coefficients and the number of zero coefficients before the last nonzero coefficient. According to our experimental results, bit to code significant map increases when these above two factors become larger. Hence, we propose that bit to code the significance map is estimated by

$$R_{Significance\ Map} = C_z \times N_z + C_{sm} \times N_s \tag{2.4}$$

where N_z is the number of zero coefficients before the last nonzero coefficient, N_s is the number of nonzero coefficients. C_z and C_{sm} are weighted coefficients for this linear estimation formula.

2.3.3 Magnitude and Sign Information

In H.264/AVC, the value of a quantized coefficient is referred as coefficient level. After coding the significant map, coefficient levels in current block also need to be coded to represent the residual data. A coefficient level is represented by two coding symbols:

coeff_abs_level_minus1, i.e. the coefficient level magnitude minus one, and coefff_sign_flag, the sign of the significant coefficient level. As zero coefficients have been coded in significance map, using coeff_abs_level_minus1 instead of original magnitude value for nonzero coefficient levels can produce a code word book beginning from 0 instead of 1 in the binarization step. However, for the sake of simplicity, we still use the original magnitude to refer the coding of coeff_abs_level_minus1 for the rest part of this chapter.

Since original coefficient level magnitude is in decimal format, it has to be binarized before entering the binary coding engine. After the binarization process, a decimal magnitude value is converted to a bin string, the length of which tends to increase as the value of its corresponding magnitude becomes larger. Hence, it is reasonable to assume that relationship exists between coefficient level magnitude and corresponding number of coding bit: the larger the magnitude, the more bit it costs to encode. This assumption can be further explained as follows.

Previous studies have shown that measured PDF (Probability Density Function) of prediction residual or transformed prediction errors can be modeled by a highly peaked Laplacian distribution [21,22]

$$p(k) = \frac{\lambda}{2} e^{-\lambda|k|} \tag{2.5}$$

where k is the coefficient level, λ is the model parameter, and $p(k)$ is the probability of an coefficient level k. According to this distribution model, the entropy of magnitude $|k|$

is $c\lambda k$, and the entropy of magnitude $|k+1|$ is $c\lambda(k+1)$, where c is a factor. That is, ideally, the number of bits coding a coefficient level magnitude $k+1$ is expected to be $c\lambda$ larger than coding a coefficient level magnitude k. Hence, estimated number of bits coding the coefficient level magnitude should have a linear relationship with the value of the magnitude.

To verify above relationship, we count average number of bits coding each coefficient magnitude from different encoded video sequences and plot them in Fig. 2.3. As seen, for a level magnitude smaller than 5, average number of coding bits is close to the magnitude value itself. That is, the number of bits linearly increases by 1 as the magnitude increases from $k-1$ to k, when k is smaller than 5. However, when a magnitude is larger than 5, relationship between the magnitude and corresponding average number of coding bits tends to be highly nonlinear. This is because as the magnitude increases, its distribution cannot be well described by (2.5).

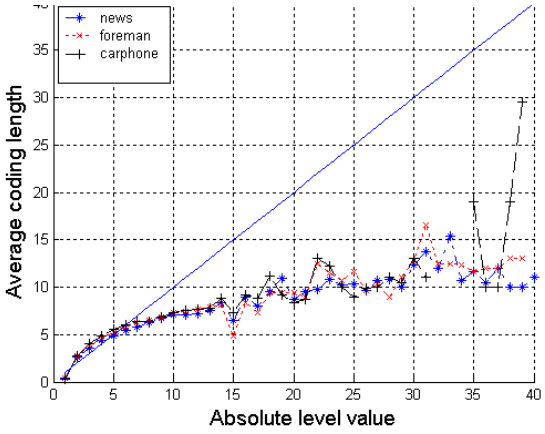


Fig. 2.3. Average number of bits to code the level magnitude (QP=20)

Based on above observations, we propose an adaptive method to estimate the number

of bits to code the magnitude information. For a magnitude smaller than T (see (2.8) for the definition of T), we adopt the value of the magnitude to estimate the coding bit. For a magnitude larger than T, the coding bit is adaptively calculated by the average number of bits to code this magnitude from previous k frames. This method is under two considerations. For a magnitude larger than T, adaptive estimation provides good accuracy. On the other hand, though adaptive estimation can also be applied for a magnitude smaller than T, according to our experiments, it does not provide significant performance difference compared with using the simple magnitude value as estimation. Moreover, since the accurate H.264/AVC prediction generates a large portion of coefficient levels whose magnitudes are smaller than T, using the simple estimation for these small value magnitudes also reduces estimation complexity.

Therefore, we propose to estimate total number of bits to code coefficient level magnitudes (coeff_abs_level_minus1) in a block as

$$R_{coeff\_abs\_level} = \sum_{i=1}^{total\_coeff} Est(L_i) \tag{2.6}$$

where L_i is the magnitude of the ith nonzero coefficient level, $Est(L_i)$ is the estimated number of bits to code L_i as defined in (2.7). In (2.7), $avg(k)$ stands for the average bit coding magnitude k computed from previous frames.

$$Est(k) = \begin{cases} k & if\ k \le T \\ avg(k) & otherwise \end{cases} \tag{2.7}$$

T is defined as

$$T = \begin{cases} 5 & QP > 15 \\ 8 & otherwise \end{cases} \tag{2.8}$$

The next part to be considered is coding coefficient level sign. Since sign is only available for these significant coefficients, as in [17], we estimate sign coding bit according to the number of significant coefficients.

$$R_{sign} = C_{si} \times N_s \tag{2.9}$$

Therefore, bit estimated to code coefficient levels in a block is given by

$$R_{Level} = C_{si} \times N_s + R_{coeff\_abs\_level} \tag{2.10}$$

2.3.4 Fast Bit Estimation Model

Summarizing discussions in 2.3.1, 2.3.2, and 2.3.3, the total number of bits to code a coefficient block based on CABAC entropy coding is estimated as

$$R_{estimated} = C_z \times N_z + C_s \times N_s + R_{coeff\_abs\_level} + 1 \tag{2.11}$$

where $C_s = C_{si} + C_{sm}$.

To get the value of coefficient C_z and C_s in (2.11), we have done several experiments using different video sequences at different QP values. The results show that when C_z is set to be 1 and C_s to be 2.5, the best coding performance is achieved. Hence, with coefficient values plugged in, (2.11) is rewritten as

$$R_{estimated} = N_z + 2.5 \times N_s + R_{coeff\_abs\_level} + 1 \tag{2.12}$$

Fig.2.4 gives the performance of block wise bits estimation based on (2.12). In Fig.2.4, x-axis denotes the exact number of bits after CABAC entropy coding and y-axis denotes the estimated number of coding bits by using (2.12) without entropy coding. It

can be seen that (2.12) gives a good estimation of the actual coding bits. Though errors still exist, they evenly distribute along the actual bits.

2.4 Experimental Results

By replacing original residual CABAC coding step in RD optimization process with fast bit estimation, the proposed method is implemented based on H.264/AVC reference software JM10.1. To evaluate its performance, the proposed method is compared with three other methods: original H.264/AVC with RD optimization enabled (RDO_{on}), original H.264/AVC with RD optimization off (RDO_{off}) and a fast bit estimation method based on CAVLC entropy coding process (Est_{CAVLC}) from [17].

In Est_{CAVLC}, the estimated number of bits to code a residual block is given by:

$$\mathrm{R}_{est} = T_c + T_z + SAT_l + 0.3\sum_{k=1}^{T_c} f_k \tag{2.13}$$

where T_c is the total number of significant coefficients, T_z is the total number of insignificant coefficients before the last significant coefficient, SAT is the sum of significant coefficient level magnitudes, and f_k is the frequency of the kth nonzero coefficient of a recorded block.

In (2.13), the frequency of the kth nonzero coefficient f_k is defined as: given a string of coefficients [0,3,0,1,-1,-1,9,1,0,0…] after zig-zag scan, the frequency of the first nonzero coefficient (3) is 1 and frequency of the last nonzero coefficient (1) is 7.

The chosen test sequences, which cover high, medium, and low motion contents, are

all in QCIF format and 100 frames long. Various QP values are tested for each video sequence to evaluate the performance at different coding bit rates.

Table 2.1 shows performance comparisons in terms of bit rate, PSNR, and coding time, where ΔBitrate and ΔPSNR is bit rate and PSNR difference compared with RDO_{on} method, respectively. ΔT is coding time saving ratio, which is defined as follows:

$$\Delta T = \frac{T_{proposed} - T_{RDOon}}{T_{RDOoff} - T_{RDOon}} \times 100\% \tag{2.14}$$

As seen, compared with RDO_{on} method, the proposed method can attain average RD optimization time saving ratio for around 30%, while maintaining similar PSNR (average 0.02 dB less) and reducing 0.22% bit rate averagely at the same time.

Please note in Table 2.1, as the principle of both Est_{CAVLC} and the proposed algorithm is to replace the residual entropy coding step with estimation, these two methods are capable of achieving same encoding time saving ratio. Hence, only one ΔT is exhibited for these two methods. In terms of coding performance, though both methods can achieve similar PSNR, the proposed method yields less bit output increase than Est_{CAVLC} (-0.22% vs. 1.68%), which means higher coding efficiency. This proves that as CAVLC and CABAC share different coding features, when entropy coding adopts CABAC, higher estimation accuracy can be obtained by the proposed method than the CAVLC based method Est_{CAVLC}. As a result, better coding performance can be achieved as well.

RD cures of three methods (RDO_{on}, RDO_{off} and the proposed) are also illustrated in

Fig.2.5. As shown, RD curves of the proposed method almost overlap with these of RDOon, which means the performance of our approach is very close to the performance of RDOon method. That is, no performance degradation is introduced by adopting the fast estimation method.

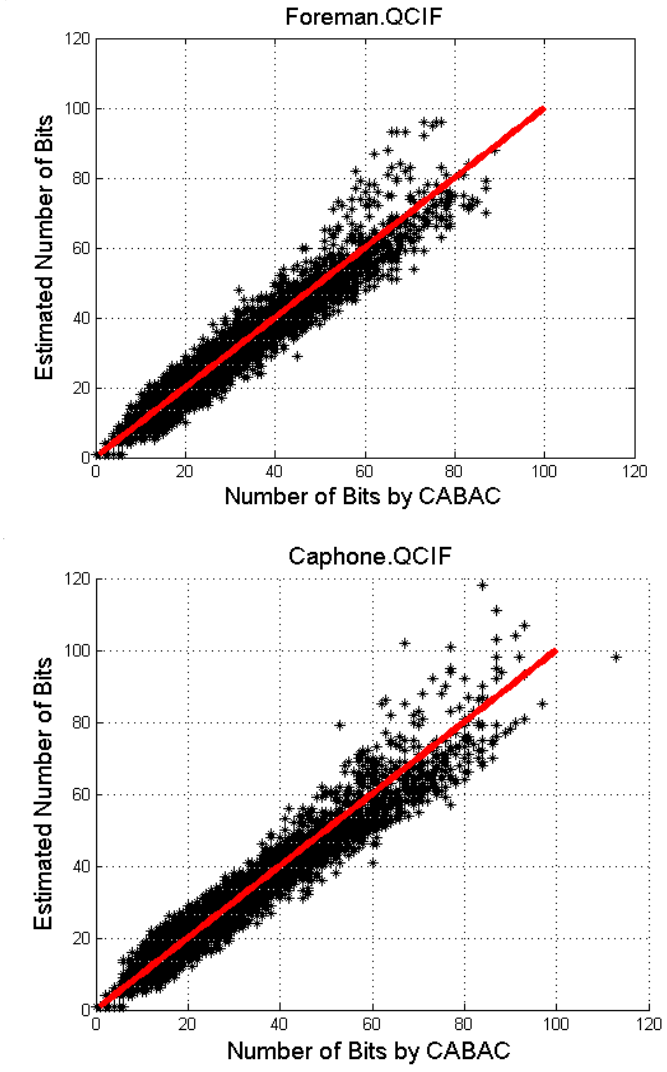


Fig. 2.4. Estimated number of bits VS. actual number of bits by CABAC

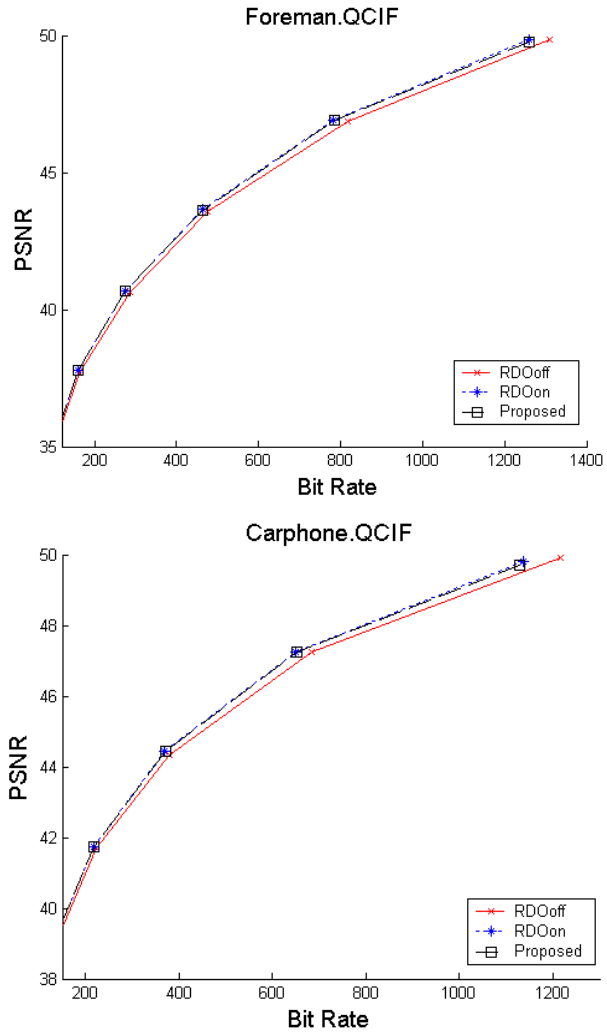


Fig. 2.5. RD curve comparisons of three Methods

2.5 Conclusions

In this chapter, we propose a bit estimation method based CABAC for fast implementation of RD optimized mode decision in H.264/AVC. Experimental results show that the proposed method greatly reduces the complexity of RD optimization for mode decision, while still achieving comparable performance as H.264/AVC encoding process with RD optimization enabled.

Table 2.1 Performance comparisons of four coding methods

Sequence	Q	RDO_{on}		RDO_{off}		Est_{CAVLC}		Proposed		
		Bit Rate	PSNR	ΔBitRate x100%	ΔPSNR	ΔBitRate x100%	ΔPSNR	ΔBitRate x100%	ΔPSNR	ΔT
Akiyo	18	117.62	45.55	3.70	-0.06	0.95	0.02	-0.03	-0.05	36.38
	26	37.07	39.69	5.13	-0.02	3.29	0.07	0.732	-0.01	25.47
	34	12.93	33.98	5.34	0.04	4.87	0.09	-0.08	-0.03	16.63
News	18	237.6	44.18	4.80	-0.05	0.91	0	0.16	0	42.92
	26	92.94	38.23	3.18	-0.05	1.34	0.01	-0.76	-0.02	29.31
	34	33.76	32.35	4.59	-0.11	2.19	0.02	1.01	-0.01	20.76
Carphone	18	369.9	44.43	2.35	-0.09	0.85	0.03	0.08	0.01	37.55
	26	120.26	38.75	2.35	-0.05	1.28	0.13	-0.77	-0.02	27.48
	34	38.31	33.13	5.90	0.15	2.06	0.17	-0.26	0.06	18.45
Foreman	18	462.18	43.65	2.54	-0.09	0.63	0.02	-0.01	-0.03	41.50
	26	159.34	37.77	2.71	-0.06	1.91	0.07	0.4	0	28.92
	34	60.34	32.42	3.46	0.07	2.86	0.17	-0.53	0	17.34
Coastguard	18	908.81	42.41	0.66	-0.18	0	-0.04	0.01	-0.03	46.78
	26	309.42	35.64	-0.86	-0.17	0.26	-0.05	-0.57	-0.08	34.96
	34	67.95	29.84	1.38	-0.12	1.81	0	-2.72	-0.1	21.99
				3.15	-0.05	1.68	0.04	-0.22	-0.02	29.76

Chapter 3
Fast Mode Decision and Motion Estimation for H.264/AVC Coding in Packet Loss Environment

In packet switch networks, packets may be discarded due to network congestion, or be treated as lost due to long end-to-end delay. Packet loss can cause severe problems for reliable video communication as the predictive loop propagates errors.

H.264/AVC coding standard offers both intra and inter coding mode at various block sizes. Since intra coding does not refer from previous frames, in packet loss environment, it is well recognized that intra coding can be a good choice to stop error propagation and enhance decoded video quality. To achieve high coding efficiency, such best coding mode decision is often made upon RD (rate distortion) optimization constraints, where the overall distortion in packet loss environment consists of not only quantization errors but also distortion resulted from channel error propagation. In order to estimate channel error propagation to help intra refresh, in JM reference software, the encoder simulates *k*-pairs of decoders and channels at the cost of great complexity burden.

In this chapter, we present a fast mode decision and motion estimation method for "intra refresh" based error resilient H.264/AVC video coding. Unlike the original

algorithm in H.264/AVC JM reference software, whose decision process must go through all possible coding modes, our method can decide the early termination of inadequate coding mode choices based on the estimated channel distortion map.

This chapter is organized as follows. Section 3.1 is the background introduction. Section 3.2 presents a RD optimized "intra refresh" algorithm using model based channel distortion estimation. Section 3.3 discusses the possibility of using channel distortion to help fast mode decision and motion estimation. Section 3.4 presents the proposed algorithm in details. Section 3.5 shows the experimental results. Section 3.6 is the conclusion.

3.1 Preliminaries

A challenge in practical video communications, where video contents are compressed and stored for delivery over networks, is that the original data are usually coded without considering varying channel conditions that may be encountered later. Such kind of compressed video is vulnerable to transmission errors which can cause severe quality degradation at receiver side. To enhance the robustness of video communication systems against packet loss, a variety of techniques have been proposed [23-25]. One of the most popular schemes is the "intra refresh" algorithm as it requires no syntax modification and hence is inherently compatible with existing coding standards.

Intra coding prevents error propagation by making no reference from previous decoded frames which may be damaged at receiver side. However, error resilient features

provided by intra refresh could be costly, as in general intra coding is not as efficient as inter predicted coding and can result in increasing number of coding bits. For the sake of coding efficiency, the problem of intra mode selection should be addressed by achieving balance between coding bit increase and error robustness.

Early intra refresh algorithms either periodically intra code the whole frame or contiguous blocks, or randomly select intra mode for blocks [26]. These algorithms introduce intra refresh rate by heuristic methods without considering content heterogeneity within a picture. Adaptive methods, such as [27,28], were proposed to frequently intra update regions that tend to have significant changes across pictures. A more efficient solution is to consider intra update under the rate distortion optimization framework [29-31]. Unlike the distortion in error free situation, the new distortion in packet loss environment is redefined by including both potential channel distortion and original quantization distortion. The channel distortion is either due to error concealment applied after packet loss, i.e., error concealment distortion, or error propagation from previous erroneous reference frames, i.e., error propagation distortion.

Since actual end-to-end channel distortion is unknown at encoder side, it must be estimated possibly based on coding information and package loss rate. Current methods of estimating channel distortion fall into two categories: simulation based and model based. In simulation based methods such as [31], the encoder simulates k copies of independent channels, whose statistical behaviors are decided by identically and independently distributed random variables. In addition, k copies of decoders are each

operated at encoder side based on each channel simulation individually. Hence, the expected distortion at decoder side can be estimated if k is chosen large enough with known packet loss rate p at encoder side. The major issue of this method is its high implementation complexity as it operates k pairs of channels and decoders at the same time, where k is a relative large number. By contrast, model based distortion estimation (MBDE) algorithms [29,30] measure the distortion based on predefined models at better computational efficiency.

In H.264/AVC, as the number of possible coding modes increases, RD optimization process becomes more complex. In this chapter, we present a fast mode decision and motion estimation method for "intra refresh" based error resilient H.264/AVC video coding. Unlike the original MBDE algorithm, whose decision process must go through all possible coding modes, the proposed method can decide the early termination of inadequate coding mode choices based on the estimated channel distortion map to reduce coding complexity.

3.2 RD Optimized Intra Refresh for Error Resilience

In this part, we first brief the Lagrange cost function in packet loss environment, and then introduce our channel distortion estimation method.

Here, we denote channel packet loss rate as p and assume p is already known at encoder side. This is reasonable as p can be adaptively calculated from information provided by the transmission protocol, such as real time control protocol (RTCP).

3.2.1. Lagrange Cost Function in Packet Loss Environment

In packet loss environment, assume that a packet gets lost at receiver side and error concealment is performed by the decoder. The concealed picture is different from the reconstructed one at encoder side. If concealed macroblocks are used for motion compensation, errors will propagate to future frames via reference. Therefore, even for an error free packet, only intra coded macroblocks may be perfectly reconstructed as inter coded macroblocks can refer from a reconstructed frame containing errors. Here, we use D_{ref} to denote the difference between signals reconstructed at encoder side and at decoder side.

Obviously, in packet loss environment, the end-to-end distortion should count the contribution not only from quantization errors but also from potential channel distortion caused by imperfectly transmission. Hence, overall distortion at decoder side should be rewritten as

$$D = (1-p)(D_{ef} + D_{ep}) + pD_{ec} \tag{3.1}$$

where D_{ef} is quantization distortion, D_{ec} denotes error concealment distortion when a macroblock is lost, and D_{ep} is the error propagation distortion if its reference frame is erroneous. Here we assume quantization distortion and error propagation distortion are linearly addible.

In original H.264/AVC, λ is designed best for RD decision in error free transmission and can be derived as follows. It's well known that the source distortion $D(R)$ depends on rate in high resolution quantization as [32]

$$D = \beta 2^{-\alpha R} \tag{3.2}$$

D can be also expressed in terms of quantization Q as

$$D = \frac{Q^2}{12} \tag{3.3}$$

The constrained RD optimization is given by

$$\min\{J\},\ J = D + \lambda R \tag{3.4}$$

By inserting (3.2) and (3.3) into (3.4), for error free transmission, (3.4) can be rewritten as

$$\min\{J\},\ J(Q) = \frac{Q^2}{12} + \lambda(-\frac{1}{\alpha}\log_2\frac{Q^2}{12\beta}) \tag{3.5}$$

The first order derivation of (3.5) is given as

$$\frac{\partial J(Q)}{\partial Q} = \frac{Q}{6} - \lambda(\frac{2}{\alpha \log 2}\frac{1}{Q}) \tag{3.6}$$

By setting (3.6) to 0, λ can be solved as

$$\lambda = \alpha \log 2 \frac{Q^2}{12} \tag{3.7}$$

In packet loss environment, original D in (3.2) should be replaced by the overall distortion (3.1). At encoder side, to consider potential channel distortion error at the decoder, the constrained RD optimization is changed from (3.4) to

$$\min\{J\},\ J=(1-p)(D_{ef}+D_{ep})+pD_{ec}+\lambda_{loss}R \tag{3.8}$$

Therefore, the first order derivation of (3.8) is given as

$$\frac{\partial J}{\partial Q}=\frac{\partial((1-p)D_{ef}+\lambda_{loss}R)}{\partial Q} \tag{3.9}$$

Please note, since error concealment distortion D_{ec} and error propagation distortion D_{ep} are independent of quantization, they are omitted in (3.9).

λ_{loss} is solved by setting (3.9) to 0 [31].

$$\begin{aligned}\lambda_{loss}&=(1-p)\alpha\log 2\frac{Q^2}{12}\\&=(1-p)\lambda\end{aligned} \tag{3.10}$$

By using λ_{loss}, the modified Lagrange cost function for RD optimization mode decision (3.8) in packet loss environment becomes

$$\begin{aligned}\min\{J\},\ J&=(1-p)(D_{ef}+D_{ep})+pD_{ec}+(1-p)\lambda R\\&=(1-p)(D_{ef}+D_{ep})+(1-p)\lambda R\end{aligned} \tag{3.11}$$

As $(1-p)$ is a common factor shared at the right side of (3.11), without loss of generality, (3.11)can be changed to

$$\min\{J\},\ J=(D_{ef}+D_{ep})+\lambda R \tag{3.12}$$

Meanwhile, when a macroblock is coded with intra mode, it does not refer to previous frames, resulting in a zero error propagation distortion in (3.12). Accordingly, the intra coding rate distortion cost J is simplified as

$$J_{intra}=D_{ef}+\lambda R \tag{3.13}$$

3.2.2 Model Based Channel Distortion Estimation

In order to use formula (3.11) and (3.13) for mode decision under RD optimization constraint, possible error propagation distortion D_{ep} from reference frames must be estimated at encoder side. It can be found that D_{ep} is caused by the imperfect reconstruction distortion D_{ref} from its reference frame. Similar to [30], we define a channel distortion map on block (4x4) basis to track potential reconstruction difference D_{ref} leading to error propagation. That is, each block is associated with a value D_{ref} indicating potential error propagation could be caused by referencing that block at the decoder.

Assume channel distortion maps of previous coded frames are available. To estimate a block based D_{ep}, we use a weighted average of reconstruction difference D_{ref} of four blocks overlapping with current block's motion compensated block in the reference frame, where the weights are proportional to the overlapped areas between the motion compensated block and each of the four blocks (Fig. 3.1). When a particular inter mode is applied to current macroblock, its potential error propagation distortion D_{ep} is estimated by summing D_{ep} of its 16 constituent blocks.

After a macroblock in current frame selects the optimal coding mode, its block based potential reconstruction difference D_{ref} is computed by

$$D_{ref} = (1-p)D_{ep} + p(D_{ec} + D_{ec\_ep}) \tag{3.14}$$

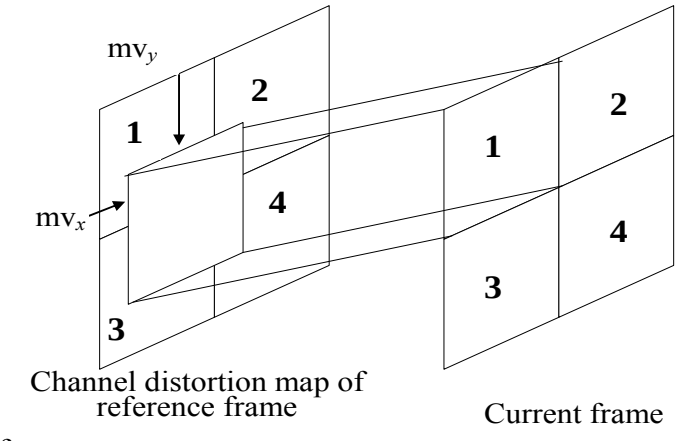


Fig. 3.1. Error propagation distortion estimation

In (3.14), D_{ep} is block's error propagation distortion related to the optimal mode decision o^*. D_{ec} refers to error concealment distortion if current MB is lost, and D_{ec_ep} is the estimated error propagation distortion from frames referenced for error concealment. The calculation of both D_{ec} and D_{ec_ep} is dependent on deployed error concealment method at the decoder, but can be estimated at the encoder.

3.3. Fast Mode Decision and Motion Estimation Using Channel Distortion Map

By adopting model based channel distortion estimation, computational cost on error prediction part is reduced. However, the overall coding complexity for RD optimized intra refresh remains too high for real-time applications.

In packet loss environment, the distortion relevant to RD optimization decision includes additional error propagation distortion (see (3.11)). Fig. 3.2 illustrates frame by frame comparisons between quantization distortion and error propagation distortion of

the coded foreman sequence using the model based distortion estimation algorithm MBDE. As seen, error propagation distortion tends to surpass quantization distortion as frames contain more motions (e.g., from frame 150 to 250) or loss rate increases (e.g., from 0.05 to 0.10). This observation implies that error propagation distortion can become a major factor in Lagrange cost function (3.11) when its value is relatively larger than quantization distortion. This assumption can be verified using results from Table 3.1.

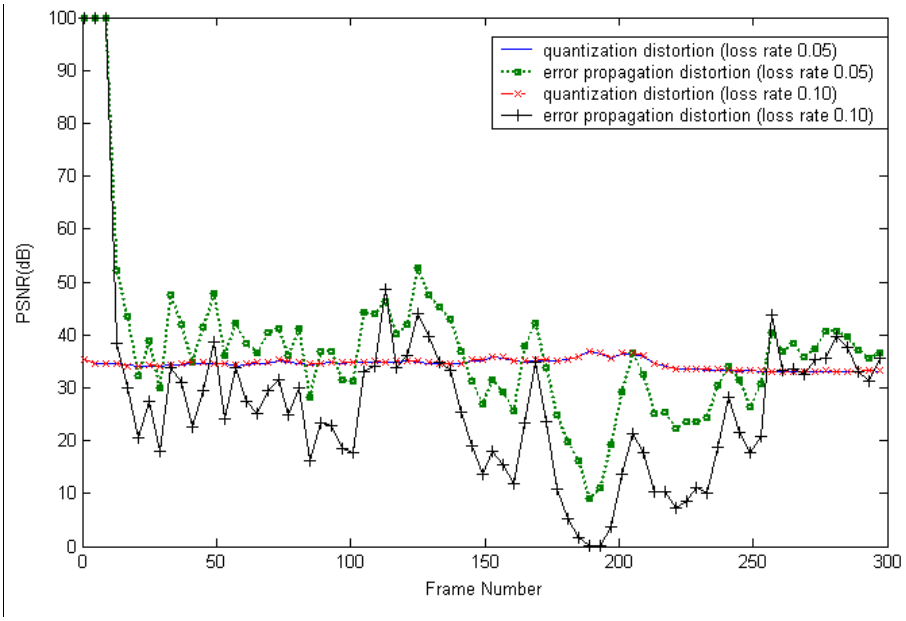


Fig. 3.2. Error propagation distortion and quantization distortion comparison of coded foreman video sequence (QP = 31)

For each macroblock, let $D_{ep}(o)$ denote error propagation distortion related to a possible inter coding option o, so $D_{ep}(o)$ is dependent on motion vector and reference frame choice of o. Let $minD_{ep}$ be the minimum value in set $\{D_{ep}(o)\}$ and $minJ_{intra}$ denote the minimum intra coding cost J given by (3.13). Table 3.1 shows statistical information from coded foreman sequence using the original MBDE algorithm: column 1 is the loss

rate, column 2 and 4 denote the corresponding ratio of intra coding macroblocks, and column 3 and 5 give percentage of intra coding macroblocks whose $minD_{ep}$ surpasses $minJ_{intra}$. As seen, for a relatively large number of macroblocks, without considering additional quantization distortion and coding rate, their $minD_{ep}$ surpasses $minJ_{intra}$. As loss rate increases, the number of such kind of macroblocks also increases. Hence, Table 3.1 confirms that in packet loss environment, error propagation distortion can become a dominant factor in RD optimized mode decision process considering the larger value of error propagation distortion. For these macroblocks whose $minD_{ep}$ is larger than $minJ_{intra}$, early termination of inter mode choices should be decided as their overall inter coding cost cannot be smaller than intra coding cost. Based on this observation, we propose a fast mode decision and motion estimation algorithm for H.264/AVC video coding in packet loss environment to reduce coding complexity.

Table 3.1 Statistical information of coded sequences

	Foreman		Coastguard	
Loss Rate	Intra Coding MB	$minD_{ep} > minJ_{intra}$	Intra Coding MB	$minD_{ep} > minJ_{intra}$
2.5%	32.65%	13.*14*%	28.9%	12.3%
5%	43.28%	25.38%	42.1%	29.8%
10%	55.40%	51.71%	60.8%	46.39%

3.4. Channel Distortion Map based fast Mode Decision and Motion Estimation Algorithm

The proposed two stage fast mode decision and motion estimation algorithm takes advantage of estimated error propagation distortion $D_{ep}(o)$ from channel distortion map for decision-making. Since channel distortion map is defined on block basis, computational cost of estimating $D_{ep}(o)$ is quite small. By using $D_{ep}(o)$ as assistance to decide early termination of possible inter coding choices, encoding speed can be accelerated significantly.

In the first stage, the algorithm quickly determines if possible inter modes should be terminated by comparing the estimated error propagation distortion to the minimum intra coding cost. If inter modes are not excluded from the first step, fast motion estimation algorithm is adopted to search for corresponding candidate motion vector.

3.4.1. Fast Mode Decision

Observations in Section 3.3 suggest the existence of macroblocks whose $minD_{ep}$ surpasses $minJ_{intra}$. Intuitively, for such kind of macroblocks, intra mode should be selected according to RD optimization criterion, as inter coding cost cannot be smaller than intra coding cost, where both costs are RD based. Nevertheless, the calculation of $minD_{ep}$ in Section 3.3 depends on the optimal motion vector decision from the motion estimation process, which means the corresponding computational cost of motion estimation cannot be reduced.

To solve this dilemma, considering the fact that the channel distortions between adjacent blocks are highly correlated, we choose to compute $D_{ep}(o)$ based on the predicted motion vector instead of the optimal motion vector decision from motion estimation process. Hence, we have the following fast mode decision algorithm.

Step 1: Decide the best intra mode by minimizing formula (3.13). Mark the corresponding RD cost as $minJ_{intra}$.

Step 2: For each inter coding option o, generate its predicted motion vector.

Step 3: Using the predicted motion vector, calculate estimated channel distortion $D_{ep}(o)$ for each possible inter coding option o. Mark the minimum $D_{ep}(o)$ as $minD_{ep}$.

Step 4: Compare $minD_{ep}$ with $minJ_{intra}$. If $minD_{ep}$ is larger than $minJ_{intra}$, best intra mode from Step 1 is selected as the final coding mode; otherwise, fast motion estimation is performed for these inter coding options whose $D_{ep}(o)$ is smaller than $minJ_{intra}$ so that RD cost between intra modes and inter modes can be compared to make the optimal mode decision.

3.4.2 Fast Motion Estimation

If possible inter modes are not terminated at the first stage, fast motion estimation algorithm can be applied to speed up motion vector search process. In JM12.1, the problem of choosing the best motion vector is also formulated by minimizing a Lagrange cost function:

$$mv^* = argmin(D(.) + \lambda_{motion} R(.)) \tag{3.15}$$

where $D(.)$ is the sum of absolute difference (SAD), and $R(.)$ represents the number of bits to code motion information.

The proposed fast motion estimation algorithm selects the best motion vector from a set of candidates by terminating these candidates that may have large error propagation distortion associated. The algorithm comprises following major steps.

Step 1: Consider up to 9 candidate motion vectors including 4 motion vectors from neighboring locations, 3 motion vectors in previous frame, zero motion vector and median motion vector [33].

Step 2: Let $D_{ep}(o_j)$ denote error propagation distortion associated with the jth sub-macroblock of an inter coding option o. Eliminate these candidates whose $D_{ep}(o_j)$ is larger than $minC_{intra}/k$, where k is the number of sub-macroblocks in mode o.

Step 3: Use (3.15) to select two best motion vector candidates. The reason of picking two instead of just one candidate is to avoid falling into local minimal spots.

Step 4: Refine motion vector candidates by using diamond search of radius 2.

Step 5: According to (3.15), choose the optimal motion vector with the minimum cost from candidates generated in Step 4.

3.5 Performance Evaluation

The proposed fast mode decision and motion estimation algorithm is implemented using H.264/AVC reference software JM12.1. The experiment adopted default configurations unless modifications specified below. Even though extensive experiments

have been carried out, we select to present results for three sequences, i.e., Carphone, Foreman, and Coastguard. In all experiments, QCIF test sequences are encoded at 7.5 frames per second with a fixed quantization parameter QP. The channel packet loss is simulated according to the packet loss testing conditions specified in [34].

The performance of simulation based error estimation method in JM12.1 (k = 30), original MBDE method, and our proposed method is compared in terms of PSNR, bit-rate and speed. For each compressed video, the packet loss situation is simulated 300 times; and the reconstructed video PSNR is calculated by using $PSNR = \frac{1}{300}\sum_{i=1}^{300} PSNR_i$, where $PSNR_i$ is the average $PSNR$ of the ith decoded sequence over all frames.

Table 3.2 shows the experimental results from three algorithms at different packet loss rates. As seen, compressed video using the proposed method is capable of achieving similar decoding quality as the simulation based method. That is because the channel distortion map is capable of capturing possible error propagation features according to the known loss rate. Meanwhile, the corresponding encoding process by the proposed method performs at much higher speed. Early termination of inadequate inter coding modes provides computational complexity saving of corresponding motion prediction time. Moreover, as the loss rate increases, compared to the simulation based error estimation method in JM12.1 (k=30) and the original MBDE, speed up gain of the proposed method becomes larger as well, since more macroblocks fit the early termination criterion stated in Section 3.4.

3.6 Conclusions

In this chapter, we propose a new fast mode decision and motion estimation method for “intra refresh” based error resilient H.264/AVC video coding. The method accelerates coding speed by deciding the early termination of inadequate coding mode choices based on the estimated channel distortion map. Compared to simulation based error estimation method in JM12.1 (k=30) and the original MBDE method, speed up gain of the proposed method can increase up to 13 and 10 times, respectively.

Table 3.2 Performance evaluation of the proposed scheme

Loss Rate	Method	Comparison Metric	Sequence		
			Carphone	Foreman	Coastguard
0.05	Simulation Based	PSNR	31.1dB	30.53dB	29.98dB
		Bit rate	69.5kbits/s	98.40kbits/s	118.5kbit/s
	MBDE	ΔPNSR	-0.04dB	0.58dB	0.08dB
		ΔBit%	-3.52	4.7	-3.80
		Speed up	1.25	1.29	1.25
	Proposed	ΔPNSR	0.18	0.66dB	0.24dB
		ΔBit%	1.31	5.17	-2.11
		Speed up	9.06	9.19	6.83
0.10	Simulation Based	PSNR	29.78dB	28.84dB	29.2dB
		Bit rate	77.5kbits/s	112kbits/s	136.5kbit/s
	MBDE	ΔPNSR	0.84dB	0.57dB	0.71dB
		ΔBit%	3.29	2.23	6.22
		Speed up	1.24	1.30	1.29
	Proposed	ΔPNSR	0.62dB	0.64dB	0.57dB
		ΔBit%	-3.16	1.78	6.90
		Speed up	10.88	11.66	10.47
0.20	Simulation Based	PSNR	28.26dB	26.53dB	27.83dB
		Bit rate	88.3kbits/s	129kbits/s	157kbit/s
	MBDE	ΔPNSR	1.08dB	0.7dB	0.47dB
		ΔBit%	1.96	5.81	4.77
		Speed up	1.24	1.30	1.27
	Proposed	ΔPNSR	1.17	0.71dB	0.4dB
		ΔBit%	4.23	4.65	4.77
		Speed up	12.4	13.13	11.87

Chapter 4
Rate Control: The Review

Rate control aims to achieve perceivable video quality for video communication applications under real-world constraints, such as bandwidth, delay, buffer sizes, etc. A practical rate control algorithm normally consists of two parts [3]: bit allocation and quantization control. The bit allocation part determines the number of bits to be allocated among different coding units, such as GOP (Group of Picture), frame, slice, or macroblock. After a bit budget R is allocated, the quantization control part determines quantization Q for each coding unit so that the number of output bits is close to the allocated bit budget.

In this chapter, we first review previous work related to bit allocation and quantization control, and then survey a H.264/AVC rate control framework JVT-G012 currently adopted by H.264/AVC JM reference software.

4.1 Rate Control Review

Rate control has been widely studied for digital video coding standards and applications, such as TM5 for MPEG-2 [35], VM8 for MPEG-4 [36,37], TMN8 for H.263+ [38], and so on.

A rate control algorithm regulates the encoder output bitstream to meet channel bandwidth and buffer constraint while pursuing high video quality through two steps: bit allocation and quantization control.

4.1.1 Bit Allocation

Due to the complicated video coding process and content variation across a video sequence, it is not feasible to only adjust encoding parameters to achieve constant bit rate output of the video encoder. In real world, rate control is realized with the introduction of buffers at both encoder and decoder side. With assistance of buffers, a variable output rate bit stream from the encoder can be stored and smoothed to transmit at a constant channel transmission bit rate. At any moment the encoder generates more bit than a buffer can hold, buffer overflow occurs. When buffer overflow is anticipated, the encoder usually chooses to skip some frames to avoid buffer overflow, which can result in motion discontinuity and quality degradation of the reconstructed video. In contrast, if the buffer level is too low, there can be a period of time that no output can be extracted from the buffer and the channel bandwidth is wasted.

Bit allocation considers available bandwidth and buffer status to assign budgets to different coding levels so that the targeted output bit rate from the buffer can fit the bandwidth without buffer overflow and underflow.

At frame level, constant rate bit allocation is commonly chosen. This bit allocation strategy tends to assign same bit budget for each frame as it assumes video contents vary

gradually from frame to frame. However, since video content can be highly dynamic and nonstationary across frames, constant rate bit allocation may result in quality fluctuation and negative effects on subjective video quality. To deal with these problems, in practical, semi-constant rate bit allocation algorithms are often used as well. Semi-constant schemes adaptively adjust allocation so that frames of higher complexity will be assignment with more bits and vice versa.

In previous works, several constant or semi-constant rate bit allocation schemes are proposed. Vetro and Sun proposed a heuristic method for MPEG-4 rate control to achieve bit allocation by considering buffer constraint [39]. In [40], Sun and Ahmad introduced PID (Proportional–Integral– Differential) controller to reduce the deviation between current buffer occupancy and target buffer level and minimize the buffer overflow and underflow. The algorithm defines a error signal, which reflects the difference between the target buffer fullness $\frac{B_s}{2}$ and the actual buffer level $B_{f,t}$ at time t as

$$E_t = \frac{(\frac{Bs}{2} - B_{f,t})}{\frac{Bs}{2}} \tag{4.1}$$

The error signal is then sent to the PID controller

$$PID_t = K_p\left(E_t + Ki \cdot \int_0^t E_\tau \cdot d\tau + K_d \cdot \frac{dE_t}{dt}\right) \tag{4.2}$$

where K_p, K_i, and K_d are the proportional, integral, and derivative control parameters, respectively. The algorithm effectively reduces the deviation between the current buffer fullness and the target buffer fullness, and minimizes the buffer overflow or underflow [40].

Li and Lin proposed a model-based method, in which the buffer dynamics are represented by a simple fluid-flow traffic model and the linear tracking theory is adopted to design the frame level control law [41]. This fluid-flow traffic model based method emphasizes on the buffer control without considering the complexity difference across the video sequence.

Unlike the constant rate bit allocation, constant distortion bit allocation aims to reduce the temporal artifacts, such as quality flicker and motion jerkiness, while regulating the output bit rate to the bandwidth and buffer constraint. A good constant distortion rate control should reduce the distortion variation across a coded sequence at no or trivial cost of average quality degradation.

Constant distortion bit allocation has been well studied in previous work. In [42], Xie and Zeng proposed to use MAD as video content complexity measurement and allocate bit budget based on MAD value of current frame. The relationship between coding bit and MAD at frame level is modeled by

$$R = k\sqrt{MAD} \tag{4.3}$$

By imposing a condition that each coded frame should have a close $C = \frac{bitrate}{Framerate}$, bits T_n assigned to current frame is given as

$$T_n = C\sqrt{\frac{MAD_n}{MAD_{n-1}}} \tag{4.4}$$

After frame level bit budget is allocated according to (4.4), a multiple pass coding process is to allow the final coding results to fully take advantage of the bit assignment.

Though the method gave a possible solution for the constant distortion allocation problem, its drawbacks is obvious. The multiple pass coding process imposes abundant complexity on the encoder Shen et al. [43] followed the principle of using MAD as complexity measurement and incremental PID for bit allocation so that more bits will be allocated to those bit demanding frames. In [3], we proposed a constant distortion bit allocation method for H.263+ encoder which directly targets on minimizing PSNR variation to alleviate fluctuation. Based on prediction residual signal at frame level, the algorithm derives a close-form approximation of rate distortion relationship, and assigns bit budget for each frame according to this approximation to enable the distortion of current frame to be close to the average distortion of previous coded frames.

4.1.2 Quantization Control

To assign a quantization, coding unit R-Q relationship has to be known. In general, the R-Q relationship is found to be highly nonlinear. It depends on the prediction residual and the entropy coding method used by the encoder. Either empirical or analytical methods can be adopted to estimate the R-Q relationship. While the empirical method needs to sample and interpolate the actual (R,Q) coding results of a given encoder, the analytical approach is capable of deriving R-Q curves by only using a close-form mathematical function based on the statistical property of real R-Q curves. For different coding sources, the close-form function adjusts its parameters to better approximate the

actual R-Q curve. Because of its less complexity, the analytical approach is more widely used.

Derived from the entropy of a Laplacian distributed coefficient source, a rate model was proposed in [44] and successfully applied for H.263 rate control

$$R(Q)=\begin{cases}\frac{1}{2}\log_2(2e^2\ \frac{\delta^2}{Q^2})\frac{\delta^2}{Q^2}>\frac{1}{2e}\\ \frac{e}{\ln 2}\frac{\delta^2}{Q^2}\qquad \frac{\delta^2}{Q^2}\le\frac{1}{2e}\end{cases} \tag{4.5}$$

The distortion model is simply derived from the quantization errors of a uniform random variable by a quantization Q to help solve the optimal bit allocation at MB level.

$$D=\frac{Q^2}{12} \tag{4.6}$$

Based on the Laplacian distribution assumption of DCT coefficients as well, Chiang et al. proposed a quadratic rate-distortion model [21,22]. Using a Cauchy density function to model DCT coefficient distribution by

$$p(x)=\frac{1}{\pi}\frac{\mu}{\mu^2+x^2}\quad x\in R \tag{4.7}$$

Kamaci et al. built a slightly different R-Q model [45].

$$\begin{cases}D=cR^{-\gamma}\\ D(Q)=bQ^{\beta}\end{cases} \tag{4.8}$$

In [46-48], He and Matria proposed the ρ domain rate control algorithm, where ρ is the zero ratio of transformed coefficients after quantization. Different from other methods that model the relationship between R and Q directly, ρ domain method establishes a

one-to-one correspondence between ρ and Q, and rate R is found to have a linear relationship with ρ.

4.2 H.264/AVC Rate Control

Quantization control for H.264/AVC is more difficult as quantization parameter QP affects both rate control and RD optimization process. The coding process of a MB with rate control involved is illustrated by Fig 4.1.

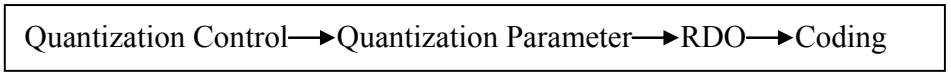


Figure 4.1 Coding Process of a MB with rate control

Because λ in the RDO Lagrange function is affected by QP, to perform RD optimization for a macroblock, QP has to be known. Therefore, QP should be decided by quantization control using residual signal before RD optimization. However, residual signal from motion estimation and mode decision is only available after RD optimization. As seen, this is a typical chick-egg dilemma.

To get around this dilemma, in JVT-G012 [49], Li et al. proposed to estimate MAD (Mean Absolute Difference) based upon a linear prediction model using MAD from prior. The predicted MAD is then treated as a measurement of coding complexity and used for quantization control. With predicted MAD, quantization is decided without using residual signal from RDO, and then RDO is performed with the quantization parameter QP decided from the quantization control part. JVT-G012 has been adopted by the

H.264/AVC JVT (Joint Video Team) reference software. In the rest part of this section, we brief the major ideas of JVT-012 rate control algorithm.

The algorithm consists of three consecutive components: GOP level bit allocation, frame level bit allocation and basic unit level quantization decision. The basic unit is defined as a group of successive macroblocks in the same frame. When the basic units arise, each shall contain at least one macroblock. After bit budget is assigned for each basic unit, the quadratic R-Q model [50,51] is adopted to decide quantization under the consideration of unit characteristics.

4.2.1 Bit Allocation at GOP Level

GOP is a group of successive pictures within a coded video sequence. The bit allocation of JVT-G012 follows a fluid flow traffic model proposed in [41]. Generally speaking, the number of bits to be allocated at different levels in a GOP depends on several factors: available bandwidth (C), frame rate (f), virtual buffer occupancy at the ith frame (V_i), actual output of the ith frame (B_i), and the total number of frames in a GOP (N_{GOP}). Here, the virtual buffer is a bridge between variant bit rate of encoder output and constant bit rate of channel input, where the buffer level will be increased a bit when extra number of bits is produced by the encoder than channel input, and vice versa. By considering these factors, the encoder output bit rate resulted from rate control is expected to be close to channel bit rate.

GOP level rate control decides the total number of bits to be allocated for all non-coded frames in current GOP. At the beginning of a GOP, the total number of bits T_0 assigned for current GOP is computed as follows

$$T_0 = \frac{C}{f} \times N_{GOP} - V_0 \tag{4.9}$$

where V_0 is the virtual buffer occupancy at the beginning of current GOP.

For rest pictures, the remaining bit budget T_i at the ith frame in current GOP will be updated by

$$T_i = T_{i-1} - B_{i-1} \tag{4.10}$$

where B_{i-1} is the number of coding bits from the i-1th frame.

4.2.2 I Frame Quantization Decision

The first picture is always coded as I frame in a GOP. In JVT-G012 [49], there is no bit allocation for I frame, and I frame quantization is determined in a relatively simple way. For I frame in the first GOP, quantization parameter QP, an index of the actual quantization value, is decided according to available bandwidth and frame rate by

$$QP_1(0) = \begin{cases} 40 & bpp \le l1 \\ 30 & l1 < bpp \le l2 \\ 20 & l2 < bpp \le l3 \\ 10 & bpp > l3 \end{cases} \tag{4.11}$$

In (4.11), *bpp* (Bits Per Pixel) is computed by $\frac{R}{f \times N_{pixel}}$, where R is the target bit rate, N_{pixel} is the number of pixels in a picture, and f is frame rate. l_1, l_2, and l_3 are predefined values being adjusted to input frame size.

For I frame in the rest GOPs, QP is decided by average P frame QP from previous GOP.

$$QP_n(0) = \max\left\{ \begin{array}{l} QP_{n-1}(0) - 2, \\ \min\left\{ QP_{n-1}(0) + 2, \dfrac{SumPQP(n-1)}{N_p(n-1)} - \min\left\{ 2, \dfrac{N_{n-1}}{15} \right\} \right\} \end{array} \right\} \qquad (4.12)$$

where $N_p(n\text{-}1)$ is the total number of stored pictures in the *n-1*th GOP, $SumPQP(n\text{-}1)$ is the sum of average picture level quantization parameters for all stored pictures. QP*n-1*(*0*) is the I frame QP in the *n-1*th GOP. $N_{n\text{-}1}$ is total number of frames in the *m-1*th GOP.

As seen, the I frame quantization in JVT-G012 [49] is very rough as it did not take the I frame and the consequent inter coded P frame complexity into account,.

4.2.3 P Frame Bit Allocation at Frame Level

For rest inter predicted pictures, before coding the *i*th picture in current GOP, the algorithm assigns target bit rate R_i for the *i*th frame as a convex linear combination of $\hat{R}_i$ and $\tilde{R}_i$.

$$R_i = \beta \times \hat{R}_i + (1-\beta) \times \tilde{R}_i \qquad (4.13)$$

where $\hat{R}_i$ is the target bit assignment considering remaining bits T_i in current GOP and $\tilde{R}_i$ is the target bit rate assignment at sequence level. β is a weighted factor set to be 0.5 here. $\hat{R}_i$ is calculated by

$$\hat{R}_i = T_i / (N_{GOP} - i) \qquad (4.14)$$

and $\tilde{R}_i$ is computed by

$$\tilde{R}_i = \frac{C}{f} + \gamma \times (S_i - V_i) \tag{4.15}$$

where S_i is target buffer level and V_i is current buffer level. γ is a weight factor and its typical value is 0.5. S_i is decided for each GOP after coding the I frame so as to reduce buffer occupancy. After coding the first inter predicted frame in a GOP, the initial value of target buffer level is set to:

$$S_2 = V_2 \tag{4.16}$$

The target buffer level for subsequent stored picture is determined by

$$S_i = S_{i-1} - \frac{S_2}{N_{GOP} - 2} \tag{4.17}$$

4.2.4 P Frame Basic Unit Quantization Control

At basic unit level, a quadratic rate distortion model is adopted to model R-Q relation [50,51]:

$$R_{residual} = c_1 \times \frac{MAD}{Q} + c_2 \times \frac{MAD}{Q^2} \tag{4.18}$$

where $R_{residual}$ is the number of bits for residual signal coding, MAD is mean absolute difference of current coding unit, C_1 and C_2 are two model coefficients. In H.264/AVC, quantization step QP is used as an index for actual quantization Q with the relationship

$$Q = 2^{(\mathrm{QP}-4)/6} \tag{4.19}$$

As stated, to dissolve the chick-egg dilemma, JVT-G012 introduced a linear model to predict the MAD of current basic unit by the MAD of the basic unit in the co-located

position from previous frame. Let MAD_p denote the predicted MAD and MAD_c denote the actual MAD of basic unit in the co-located position from previous frame, the predicted MAD_p of current basic unit is computed by:

$$MAD_p = a_1 \times MAD_c + a_2 \tag{4.20}$$

where a_1 and a_2 are two coefficients. The initial value of a_1 and a_2 are set to 1 and 0, respectively. They are updated after coding each basic unit.

At basic unit level, target bit assigned to a unit is proportional to estimated MAD of that unit. In current frame, denote MAD estimation of the jth unit in a frame by $MAD_{p,j}$, and remaining bit budget in this frame after coding the kth unit by L, bits B assigned to the next $k+1$th unit is decided by

$$B = \frac{L}{\sum_{j=k+1}^{n} \mathrm{MAD}_{p,j}} \times \mathrm{MAD}_{p,k+1} \tag{4.21}$$

With the concept of basic unit, models (4.18) and (4.20), the steps to perform quantization control at basic unit are given as follows [49]:

1. Compute a target bit assignment for current frame by using the fluid traffic model and linear tracking theory [41].

2. Predict the MAD of current basic unit and remaining basic units by the linear model (4.20) using the actual MAD of basic unit in the co-located position from previous frame.

3. Allocate bit to the next non-coded basic unit in current frame using (4.21).

4. Compute quantization by using the quadratic R-D model [50,51].

5. Perform RDO for each MB in current basic unit using quantization parameter derived from step 4.

Chapter 5
Frame-Level Constant-Distortion Bit Allocation for Smooth H.264/AVC Video Quality

Rate control for H.264/AVC is challenging as the well-known chick-egg dilemma rising from the RD optimization decision process. Current H.264/AVC reference software adopts a rate control framework proposed in JVT-G012 [49]. Though JVT-G012 provides a rate control solution for H.264/AVC, there are several issues not addressed enough in this dissertation:

1. At frame level, based on the assumption that video complexity varies gradually from frame to frame, the method tends to allocate similar budget for each frame. However, when high motion or scene change occurs, this semi-constant rate bit allocation can cause negative visual artifacts, such as quality flicker and motion jerkiness.
2. The quantization control R-Q model is based on a simple rate distortion close form $D = \beta e^{-\alpha R}$. We hope to find a more accurate R-Q model so that better coding performance can be obtained.
3. In JVT-G012, R-Q model is only available for P frame. In practice, the simple I frame quantization decision in JVT-G012 is found to have some drawbacks.

For instance, it tends to assign a relative small quantization for I frame, which can lead to possible buffer overflow and overall quality degradation.

To provide a solution for the first issue, in chapter 5, a frame level bit allocation method is proposed. By adaptively adjusting frame level budget according to coding complexity and target distortion, the proposed method can effectively alleviate quality fluctuation across the reconstructed video sequence.

In chapter 6, problem 2 and 3 is addressed by a novel P frame R-Q model for basic unit quantization control and an I frame source model with picture activity measurement.

The remainder of chapter 5 is organized as follows. Section 5.1 briefs the constant distortion rate control problem. Section 5.2 describes the two-stage encoder structure used for constant distortion rate control. Section 5.3 discusses enhanced frame level bit allocation. Section 5.4 presents the improved basic unit quantization decision. Section 5.5 gives the experimental results and the last section is the conclusion.

5.1 Constant Distortion Rate Control

While producing high quality video under bandwidth and buffer constraint, constant distortion rate control targets at smoothing the video quality across the coded frames as well. In [3], we proposed a constant distortion bit allocation method for H.263+ encoder that directly targets on minimizing PSNR variation to alleviate distortion. Based on prediction residual signal at frame level, the algorithm derives a close-form approximation of rate distortion relationship, and assigns bit budget for each frame

according to this approximation to enable the distortion of current frame to be close to the average distortion of previous coded frames.

However, method in [3] cannot be applied to H.264/AVC encoder directly as the well-known chick-egg dilemma resulted from RD optimization decision process: the frame-level prediction residual is not available until the whole frame is encoded. To decouple this dilemma, we adopt a two-stage encoder structure [52] so that frame-level coding complexity can be estimated for bit allocation purpose. Moreover, an enhanced frame-level and basic unit level rate control is proposed in this work to achieve better coding quality.

5.2 Two-Stage Encoder Structure

To allocate frame level bit budget according to $D(R)$ relationship, residual complexity has to be measured first. In order to dissolve the inter-dependency problem between RD optimization and rate control in H.264/AVC so that frame level bit budget can be allocated based on residual complexity, we employ a two-stage encoder structure proposed in [52]. In this structure, the RD optimization procedure will be performed at stage-1 followed by the rate control process at stage-2. Before stage-2 coding starts, frame-level bit budget is allocated using $D(R)$ relationship according to a target distortion so as to reduce quality fluctuation across the frames (to be discussed in details in Section 5.3).

The major ideas of the two stage encoder structure can be summarized as follows, , where QP_1 and QP_2 are quantization parameter used in stage-1 and stage-2, respectively.

Stage 1: By averaging previous frame's QP_2, a constant QP_1 is used to perform macroblock RD optimization in current frame. The macroblock residual then goes through DCT/Q and IQ/IDCT to obtain a reconstructed macroblock, which is required for intra prediction of subsequent macroblocks.

Stage 2: With the bit assignment at frame-level, quantization parameter QP_2 is decided for each basic unit to fulfill the target bit. Instead of going through RD optimization again, the optimal macroblock type and prediction mode are adopted directly from stage-1 here. That is: if a current macroblock is inter coded in stage-1, inter predicted residual from stage-1 is simply re-quantized by QP_2; otherwise, if a current macroblock is intra coded, its residual signal is re-computed with the best intra mode from stage-1 before being quantizated by QP_2, since the neighboring pixels can be different from these in the stage-1.

This two stage encoder structure is built based on the experimental results in [52]. It is shown that the decrease of coding gain is minor even though QP_1 and QP_2 are different as long as their difference Δ is restricted to a small range, i.e.,

$$\left|QP_1 - QP_2\right| \le \Delta \quad where\ \Delta \le 3 \tag{5.1}$$

Few observations can be made from this two-stage encoder structure. First, as the coding gain difference is small, the residual in stage-1 can be considered as a close-form

prediction of final RD optimization results in stage-2. By analyzing the characteristics of stage-1 residual, it shall help us to decide bit allocation and basic unit rate control in stage-2. Secondly, since stage-2 encoding demands only one additional forward and inverse DCT and one more quantization process, the computational complexity increase of this two-stage encoder structure is relatively small comparing to the overall H.264/AVC encoding cost. Fig. 5.1 illustrates the structure of adopted two stage encoder.

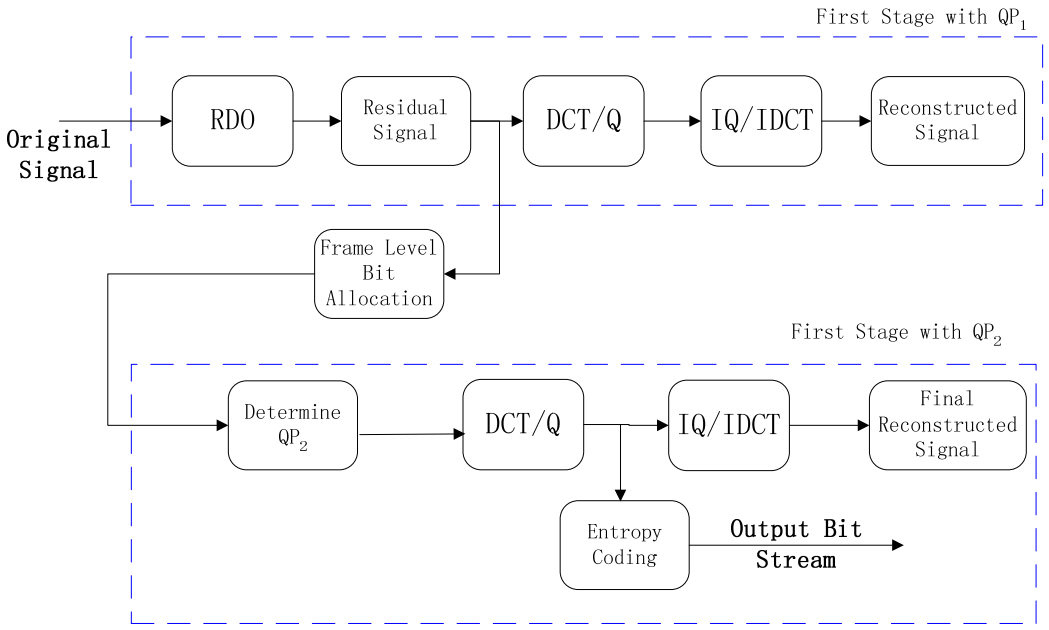


Fig.5.1 Two stage encoder structure

5.3 Frame-Level Bit Allocation

The purpose of constant distortion bit allocation is to minimize quality fluctuation across the entire video sequence. The bridge between distortion and coding bit is built upon the ρ-domain rate control theory [46].

According to the ρ-domain rate control algorithm, where ρ is the zero ratio of quantized coefficients, coding bit rate is observed to follow a linear relationship with ρ and the distortion has shown an exponential relationship with ρ as in (5.2), where σ^2 is the picture variance, θ and α are constants for a frame.

$$\begin{aligned} R &= \theta \cdot (1-\rho) \\ D &= \sigma^2 e^{-\alpha(1-\rho)} \end{aligned} \tag{5.2}$$

If denote the number of zero coefficients as $\hat{N}$, (5.2) becomes

$$\begin{aligned} R &= \theta \cdot (N-\hat{N}) \\ D &= \sigma^2 e^{-\beta(N-\hat{N})} \\ \hat{N} &= N \times \rho \end{aligned} \tag{5.3}$$

where *N* is the total number of coefficients. This ρ domain rate control algorithm has been successfully applied in the optimal bit allocation inside a video frame, including bit allocation among video objects or different groups of macroblocks within a frame [46-48].

To use ρ domain rate-distortion model for frame-level bit allocation, prior to stage-2, the average distortion of (*K-1*) previous coded frames is calculated as the target distortion for current frame. Then at frame level, the number of zero coefficients $\hat{N}$ corresponding to the target distortion is estimated. Based on the estimated $\hat{N}$, a bit budget is allocated by the linear relationship between $\hat{N}$ and *R*. Finally, the allocated bit budget is shaped using buffer constraints.

5.3.1 Target Frame-Level Distortion

The primary goal of constant distortion bit allocation is to alleviate the quality variation. It's easy to see that the closer the distortion of a coded frame is to average distortion, the smaller the quality variation is. Therefore, to ensure current frame distortion to be close to the distortion of previously coded frames, the target distortion $\hat{D}_K$ is decided by averaging the distortion of (*K-1*) previous encoded frames as

$$\hat{D}_K = \frac{1}{K-1}\sum_{i=1}^{K-1} D_i \tag{5.4}$$

5.3.2 Estimate of Number of Zeros

To assign bit budget according to the target distortion, the relationship between distortion and the number of zero coefficients $\hat{N}$ has to be known. That is, σ^2 and β in (5.2) must be computed according to the features of current frame. In this part, we first derive σ^2 and β based on stage-1 residual information, and then estimate ρ using the known $D(\rho)$ relationship.

After RD optimization, the residual from stage-1 can be transformed by 4x4 DCT. Denote the histogram of the absolute DCT coefficients by *H*, the distortion caused by quantization *Q* can be calculated as

$$D(Q) = \sum_{x=0}^{M} \int_{x-0.5}^{x+0.5} (y - r_Q(x))^2 \, p(y) dy \tag{5.5}$$

where x is the rounded absolute DCT coefficient value, M is the entry number of the histogram, $r_Q(x)$ is the reconstruction value of x defined by the quantization scheme, $p(y)$ is the underlying probability density function of the absolute DCT coefficients and $\int_{x-0.5}^{x+0.5} p(y)dy = \frac{H(x)}{N}$.

Moreover, to model the absolute DCT coefficient distribution, we approximate the $p(y)$ as uniformly distributed in each relatively small interval [x-0.5, x+ 0.5]. Thus, (5.5) can be further simplified as [3]

$$D(Q) = \frac{1}{N}\sum_{x=0}^{M}(x - r_{Q_{step}}(x))^2 H(x) \tag{5.6}$$

As seen, the entire D-Q curve is determined by (5.6) once the histogram H is calculated and the reconstruction function $r_Q(x)$ is defined. Since $D(Q)$ is a monotonically increasing function, without loss of generality we may assume that for some Q^*, $D(Q^*\text{-}1) < \hat{D}_k < D(Q^*)$. Denote the number of zero DCT coefficients for a specific Q as $N(Q)$, then we have

$$\begin{aligned} D(Q^*) &= \sigma^2 e^{-\beta(N - N(Q^*))} \\ D(Q^* - 1) &= \sigma^2 e^{-\beta(N - N(Q^* - 1))} \\ \hat{D} &= \sigma^2 e^{-\beta(N - \hat{N})} \end{aligned} \tag{5.7}$$

From (5.7), the number of zero DCT coefficients $\hat{N}$ corresponding to the target distortion $\hat{N}$ can be solved in close-form without explicitly knowing σ^2 and β as [3]

$$\hat{N} = \frac{\ln(\frac{D(Q^*)}{\hat{D}_K}) \times N(Q^* - 1) + \ln(\frac{\hat{D}_K}{D(Q^* - 1)}) \times N(Q^*)}{\ln(\frac{D(Q^*)}{D(Q^* - 1)})} \tag{5.8}$$

It can be seen that $\hat{N}$ is a linear combination of $N(Q^*)$ and $N(Q^*\text{-}1)$. The weights are determined by three log-ratio functions involving two bounding values of the target distortion $\hat{D}_k$.

Based on the features of the two stage encoding structure, residual from stage-1 should be highly similar to the residual from stage-2. Hence, $\hat{N}$ here will be used as estimation for the number of zero coefficients in stage-2.

5.3.3 Estimate of θ

As shown by (5.3), the relationship between the number of zero DCT coefficients $\hat{N}$ and a target bit rate R in stage-2 is linear: $R = \theta \cdot (N - \hat{N})$. To estimate the slope θ, we use the information from stage-1 by

$$\theta = \frac{R_o}{N - N_o} \tag{5.9}$$

where N_o and R_o represent the number of zero DCT coefficients after quantization with QP_1 and the resulting coding bit from stage-1, respectively. As N_o and R_o are known after stage-1 coding, there are no extra cost for the estimation of θ.

5.3.4 Enhanced Frame Level Bit Allocation

To meet the desirable distortion at the frame-level, after $\hat{N}$ calculated by (5.8) and θ estimated by (5.9), the bits allocated for coding the picture at stage-2 can be computed by

$$\hat{R} = \theta(N - \hat{N}) \tag{5.10}$$

Since above budget $\hat{R}$ accounts only for coding DCT coefficients, headers and motion vectors will add more bits

$$\hat{R} = \theta(N - \hat{N}) + R_{MV} + R_H \tag{5.11}$$

where R_{MV} and R_H are the bit budgets allocated for motion vectors and headers individually. They are estimated by bit coding motion vectors and headers from stage-1 coding process.

Meanwhile, the bit budget assigned above to maintain a constant distortion must not cause buffer overflow or underflow. Thus, the bit budget $\hat{R}$ by (5.11) needs to be further shaped with buffer constraints in two steps. In the first step, we followed the fluid-flow model in JVT-G012. The actual bit budget allocation *T* for current frame uses a weighted combination of $\tilde{R}$ and $\hat{R}$

$$T = \beta \times \hat{R} + (1 - \beta) \times \tilde{R} \tag{5.12}$$

where $\hat{R}$ is computed by (5.11), $\tilde{R}$ is a bit budget considering current buffer status, and β is a weight factor set to be 0.9 here. $\tilde{R}$ is defined the same as (4.15) by

$$\tilde{R} = \frac{C}{f} + \gamma \times (S - V) \tag{5.13}$$

In the second step, an even stronger buffer constraint will be applied if potential buffer overflow or underflow is predicted

$$T = \begin{cases} \frac{C}{f} + T_o \cdot B_{\max} - V & if\ (T + V - \frac{C}{f}) > T_o \cdot B_{\max} \\ \frac{C}{f} + T_u \cdot B_{\max} - V & if\ (T + V - \frac{C}{f}) < T_u \cdot B_{\max} \\ T & otherwise \end{cases} \tag{5.14}$$

where T_o and T_u are the overflow ratio and underflow ratio, respectively, and B_{max} is the buffer size, which is set to be *0.3*C* in simulation part.

5.4 Enhanced Basic Unit Rate Control

At basic unit level, the residual is assumed to be Laplacian distributed so that the quadratic R-Q model is adopted to make quantization decision as [50,51]

$$R_{residual} = c_1 \times \frac{MAD}{Q} + c_2 \times \frac{MAD}{Q} \tag{5.15}$$

To use (5.15) for quantization decision, basic unit MAD must be known. In order to dissolve the chick-egg dilemma, in JVT-G012, MAD is predicted through a linear prediction function by (4.20). Since video contents are dynamic, linear prediction can result in large errors when scene change or large motion occurs. Recall the two stage encoder structure in Section 5.2, only the basic unit rate control is performed in stage-2, where the MB type and prediction mode are kept unchanged from stage-1. This indicates macroblock's MAD in stage-2 should be very close to that in stage-1. Therefore, instead of performing the linear prediction using (4.20), we propose to estimate basic unit MAD

according to coding results from stage-1. For each macroblock, denote the MAD from stage-1 as MAD_1, the predicted MAD_2 in stage-2 is set to be the same as MAD_1 and used in (5.15) to decide the final quantization parameter QP_2.

5.5 Performance Evaluation

The proposed constant distortion bit allocation (CDBA) algorithm is implemented on H.264/AVC reference software JM12.1 and compared with JVT-G012[49], a semi-constant rate bit allocation algorithm.

Extensive simulations were performed on many standard sequences. Here we present a subset of representative results. The selected test sequences are News, Carphone, Foreman and a combined sequence of News and Foreman, all in QCIF format. The experiment settings are as fellows. Baseline profile is adopted and 200 frames are encoded for each sequence at 20fps. The first frame is coded as I frame and rest frames are all P frames. In the proposed algorithm, the bit allocation of the first 20 frames uses the constant rate bit allocation method to have a stable average distortion.

Table 5.1 illustrates performance comparisons of proposed CDBA algorithm and JVT-G012. As seen, the proposed algorithm is capable of generating comparable average PSNR while keeping much smaller standard deviations. This means that less quality fluctuation throughout the coded sequences is observed. Please note, since video sequences used in our test is only 10 seconds long, both rate control algorithms may have control errors depending on how many bits are left in the buffer at the end of encoding.

However, the error becomes negligible when the sequence has much longer duration than the buffer size (in second).

Subjective video quality comparison is also given in Fig. 5.2. The frames compared are the scene change frame in the combined sequence coded by two different methods. As seen, as the CDBA algorithm is able to detect scene change and assign more bits accordingly to keep its distortion close to previous coded frames, the frame coded by the CDBA algorithm has better subjective quality. By contrast, the frame coded by JVT-G012 suffered from quality degradation as inadequate bit assignment in this bit consuming scenario. From above comparisons, it can be concluded that the advantage of the proposed CDBA scheme is obvious in terms of delivering smoother H.264/AVC video quality.

(a) (b)

Fig. 5.2. Subjective quality comparison of the combined sequence coded at 40Kbps. 101st coded frame of the sequence: (a) CDBA vs. (b) Constant Rate Bit Allocation

5.6 Conclusions

In this chapter, we propose a new two stage constant distortion bit allocation algorithm with enhanced rate control for H.264/AVC encoder. The experimental results show that compared with JVT-G012, the proposed CDBA algorithm can deliver similar average PSNR for all testing video sequences, while reducing quality fluctuation across the sequences at the same time.

Table 5.1 Performance comparisons of proposed CDBA and JVT-G012

Sequence	Target bitrate (Kbps)	Algorithm	Actual bitrate (Kbps)	PSNR (dB)	σ^2 of PSNR
News	40	CDBA	40.23	32.67	0.72
		Constant	40.12	32.72	1.05
	60	CDBA	60.11	35.87	0.85
		Constant	60.06	35.99	1.06
Carphone	40	CDBA	41.21	31.89	2.24
		Constant	41.56	31.83	3.34
	60	CDBA	60.16	34.41	2.12
		Constant	60.17	34.40	2.27
Foreman	40	CDBA	41.19	28.67	2.05
		Constant	41.68	28.77	2.76
	60	CDBA	60.22	31.85	1.53
		Constant	60.22	31.87	1.71
Combined	40	CDBA	40.68	30.83	2.51
		Constant	40.67	30.86	3.09
	60	CDBA	60.09	33.96	2.19
		Constant	60.10	34.03	2.61

Chapter 6
H.264/AVC Rate Control with Enhanced R-Q Model and Bit Allocation

This chapter presents an adaptive and efficient rate control framework for low delay H.264/AVC video communication. At basic unit level, to realize quantization decision improvement, two novel R-Q models are proposed for I frame and P frame, respectively. The P frame R-Q model is established through the ρ domain rate control theory to compute quantization according to target bit assignment. The I frame R-Q model measures frame level coding activity for quantization decision without performing computationally intensive intra prediction, which alleviates the frame skipping problem.

The remainder of this chapter is organized as follows. Section 6.1 describes a quadratic ρ domain rate model and then derives the P frame square root R-Q model. Section 6.2 presents the I frame source model. Section 6.3 gives detailed description of the bit allocation approach and the basic unit level quantization control. Section 6.4 exhibits experimental results, and Section 6.5 concludes the chapter.

6.1. P Frame Square Root R-Q Model

The ρ domain source rate model is reported to be more accurate than the Q domain source rate model as it results in more precise estimation of the residual signal

complexity [46-48]. In this section, we first propose an enhanced quadratic ρ domain rate model to better measure the relationship between zero ratio ρ and coding bit rate R. Next, we show that SATD/Q can be expressed as a second order polynomial of nonzero ratio 1-ρ, where SATD refers to sum of absolute transformed differences. By combining these two results, we obtain a final usable square root R-Q model.

Traditional coding technology always exploits video sequence redundancy by prediction, either temporally or spatially. For P frame, after motion compensated prediction, residual is processed using DCT transform. When the distribution of DCT coefficients is known, there is a one-to-one mapping between quantization Q and zero ratio ρ

$$\rho = \frac{1}{N} \sum_{|x|<Q} D(x) \tag{6.1}$$

In (6.1), x is a coefficient, and $D(x)$ is the corresponding number of x in this distribution. N is the total number of coefficients, or the sum of all $D(x)$. Assume that the linearity between R and ρ within the same frame holds, according to ρ domain rate control algorithm [48], bit rate R and distortion D can be modeled by

$$\begin{aligned} R &= \theta \times (1 - \rho) \\ D &= \delta^2 e^{-\alpha(1-\rho)} \end{aligned} \tag{6.2}$$

where δ^2 is the variance of a frame, θ and α are constant model parameters for a frame.

It is obvious that the original ρ domain rate control method heavily depends on the linear relationship assumption between R and ρ. However, according to previous studies

[53] and our observations, the linear $R(\rho)$ relationship does not always hold strong enough. Once the linearity becomes weak, the original ρ domain rate control may lose its accuracy. In Fig.1, some $R(\rho)$ examples are illustrated by using different test sequences. From the plot, it is easy to observe that the linearity assumption doesn't always hold tightly. That is, $R(\rho)$ curve is not always a straight line and tends to be nonlinear for some testing sequences.

6.1.1. Quadratic ρ Domain Rate Model

To better measure $R(\rho)$ relationship, we propose to use a quadratic model instead of the original linear model so that further accuracy can be attained

$$R = \theta_1(1-\rho) + \theta_2(1-\rho)^2 \tag{6.3}$$

By using the new model defined in (6.3), $R(\rho)$ curves are plotted again in Fig. 1 as the solid lines, where θ_1 and θ_2 are obtained through curve fitting and constant for each frame. As seen, the new $R(\rho)$curves by the quadratic model exhibit a better match with the actual $R(\rho)$ curves within wide range of ρ.

To further compare the estimation accuracy of the linear and the quadratic $R(\rho)$ model, in Table 1, we also show the rate estimation errors in terms of RMSE and E^2 of coded frames by using (2) and (3), respectively. E^2 is a quantity to measure the degree of data variation from a given model [52,54]. It is defined as

$$E^2 = 1 - \frac{\sum_i (X_i - \hat{X}_i)^2}{\sum_i (X_i - \bar{X}_i)^2} \tag{6.4}$$

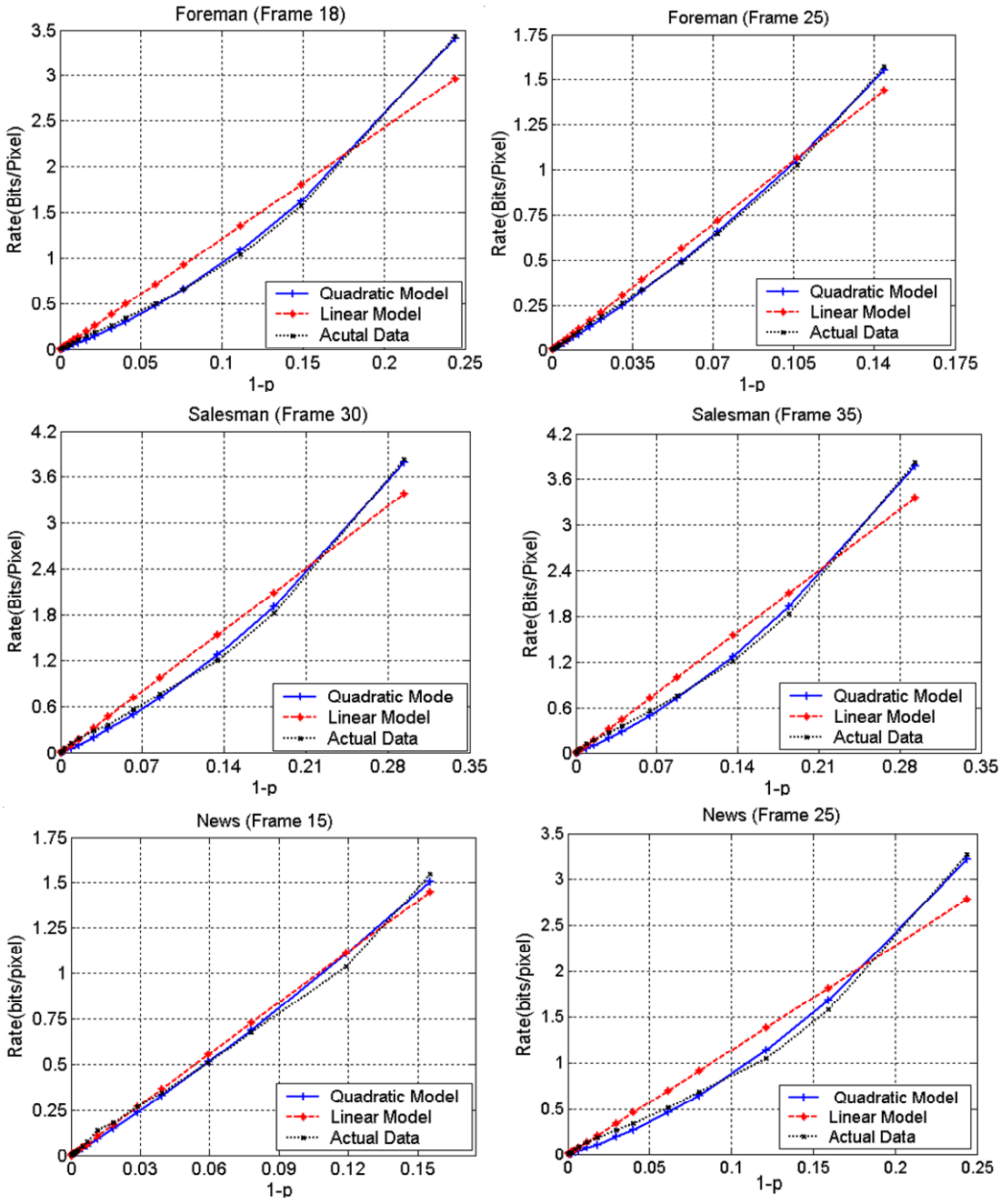


Fig. 6.1. The relationship between coding bit rate R and zero coefficient ratio ρ

where X_i and $\hat{X}_i$ are the actual and estimated values of data points, respectively, and $\bar{X}_i$ is the mean of all data points. According to the definition of E^2 in (6.4), for any reasonable model, E^2 should have a value between 0 and 1 as the second term of the

right-hand-side of (6.4) is less than one [52]. The more accurate the model, the closer the value of E^2 is to 1. As shown in Table 1, based on the quadratic model in (6.3), RMSE is reduced and E^2 is closer to 1. Therefore, we claim (3) can lead to higher estimation accuracy of $R(\rho)$ relationship.

6.1.2. Square Root R-Q Model

Similar to all other non-scalable video coding standards, H264/AVC encoder controls output bit rate by adjusting its quantization. As stated in 2.1, there is a one-to-one mapping between zero ratio ρ and quantization Q. To get $Q(\rho)$ relationship for rate control, we utilized an analytical method. In previous works [52], SATD is confirmed to be a good estimation of residual signal complexity. Here, we study the relationship between SATD/Q and zero ratio ρ and show that SATD/Q can be expressed as a second order polynomial of 1-ρ.

Again, let the statistical attributes of DCT coefficients be approximated by a Laplacian distribution $p_l(x)$ given by [21,22]

$$p_l(x) = \frac{\lambda}{2} e^{-\lambda|x|} \tag{6.5}$$

where λ is coefficient varying to different sources.

In a source given distribution by (6.5), the corresponding zero ratio ρ from a quantization Q is expressed by

$$\begin{aligned} \rho &= 2 \times \int_0^Q \frac{\lambda}{2} e^{-\lambda x} dx \\ &= 1 - e^{-\lambda Q} \end{aligned} \tag{6.6}$$

Table 6.1 Comparisons of the quadratic model and the linear model

Sequence	Frame Number	Linear Model		Quadratic Model	
		RMSE	E^2	RMSE	E^2
Foreman	18	0.134	0.961	0.022	0.999
	25	0.036	0.989	0.010	0.999
Salesman	30	0.128	0.973	0.037	0.998
	35	0.133	0.971	0.040	0.998
News	15	0.027	0.994	0.020	0.999
	25	0.136	0.961	0.042	0.999
Mother-Daughter	5	0.110	0.986	0.064	0.998
	10	0.045	0.996	0.036	0.999
Carphone	15	0.029	0.990	0.014	0.999
	25	0.030	0.990	0.015	0.999

Also, according to (6.5), SATD/Q can be written as

$$\begin{aligned} \frac{SATD}{Q} &= \sum_{x=1}^{N}\sum_{y=1}^{N}\frac{\left|e_{x,y}\right|}{Q} \\ &\approx \sqrt{\sum_{x=1}^{16}\sum_{y=1}^{16}\frac{\left|e_{x,y}\right|^2}{Q^2}} \\ &\approx \sqrt{2\int_0^{\infty}\frac{|x|^2}{Q^2}\cdot p_l(x)\,dx} \end{aligned} \tag{6.7}$$

where $e_{x,y}$ is the value of a transform coefficient at a MB position (x,y). With (6.7),

SATD/Q can be expressed as

$$\frac{SATD}{Q} \approx \sqrt{2\sum_{i=0}^{\infty} i^2 \int_{Qi}^{Q(i+1)} p_l(x)\,dx} = \sqrt{2\sum_{i=1}^{\infty} i^2 (e^{-\lambda Qi} - e^{-\lambda Q(i+1)})} \tag{6.8}$$

(6.8) converges to

$$\sqrt{2\frac{(1-e^{-\lambda Q}-1)(1-e^{-\lambda Q}-2)}{\left(1-e^{-\lambda Q}\right)^2}} = \frac{1}{1-(1-\rho)}\sqrt{2(1-\rho-1)(1-\rho-2)} \tag{6.9}$$

By taking its Taylor expansion, (6.9) can be expressed as a second order polynomial

$$\frac{SATD}{Q} = \alpha_1(1-\rho) + \alpha_2(1-\rho)^2 + \alpha_0 \tag{6.10}$$

To verify the validity of the second order polynomial SATD/Q-ρ model, we plot the actual SATD/Q-ρ relationship and the curve produced by (6.10) in the same figure (see Fig. 6.2). As shown, the estimation curve by using (6.10) matches the actual SATD/Q-ρ curve quite well at different ρ values.

With the SATD/Q-ρ formula, in the next step in (6.10), we combine it with our proposed ρ domain rate model (6.3) to obtain a usable R-Q relationship. After inverting (6.3) into (6.10) and dropping insignificant terms, we have the final square root R-Q model as

$$\frac{SATD}{Q} = \beta_1 R + \beta_2 \sqrt{R} + \beta_0 \tag{6.11}$$

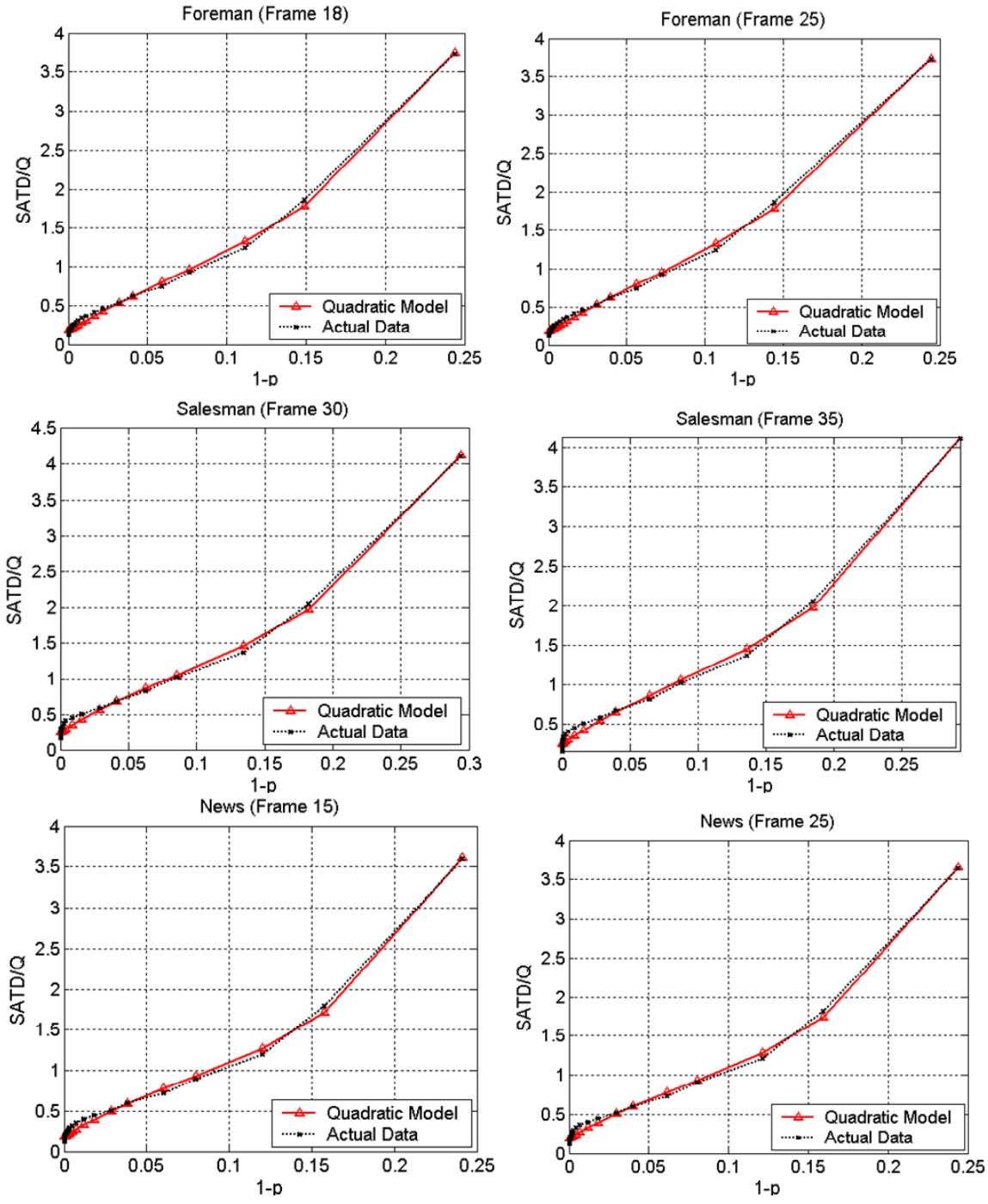


Fig. 6.2 The relationship between SATD/Q and zero coefficient ratio ρ

6.2. I Frame R-Q Model with Activity Measurement

In order to attain high coding efficiency, H.264/AVC I frame coding adopts intra prediction to explore spatial correlation within a frame. A proper I frame quantization decision is crucial as it can not only alleviate buffer management problems such as buffer overflow and underflow, but also generate a bit allocation between I frame and consequent inter coded frames that improves overall encoded frame quality. Unlike P frame, I frame quantization is decided only at frame level without adjustable basic unit quantization in H.264/AVC reference software JM12.3.

As introduced, in JVT-G012 [49], I frame quantization is determined in a relatively simple way. To assign a proper I frame quantization based on video characteristics, two issues have to be addressed: bit budget dispatch according to the frame level complexity in a GOP and the I frame R-Q relationship so that quantization can be decided with the bits assigned. The first issue is related to frame level bit allocation, which will be explained in more details in 6.4.2. Here, only the solution for the second issue will be presented.

Since the introduction of intra prediction, similar to P frame, modeling the relationship between bit rate R and quantization Q according to actual residual requires I frame to be coded, which does not comply with RDO process. To get around this problem, one solution is computing the image activity as the degree of the image complexity and a characteristic measurement of rate distortion relation [55]. Some typical

activity measuring criterion [55] include edge intensity, gray level variance, and entropy variances. In [56], Pan et al. designed an image activity measurement according to the block based DCT coefficients. However, as intra predicted residual can have different energy features from the original image, such measurement cannot be very accurate. In [57], Czuni et al. proposed to calculate quantization from spatial and temporal features including average edge content by training these features through a neural network. A drawback of this method is that how these features affect the model parameters was not clarified. Moreover, the method is not practical for some real time applications as training through the neural network needs a large number of data and certain time. Jing and Chau introduced an intra R-Q model by using a quadratic formula [58]. However, adaptive update of the model parameters is neglected. The scheme in [59] estimates intra picture activity by computing the variance of spatial image values, but it has to deal with the window size problem, which may result in overestimating the residual complexity of a textural macroblock if a too large statistical window size is used.

Based on previous works and experimental results, we propose to measure H.264/AVC I frame coding activity C by the gradient information that is defined as

$$C = \left|Grad_x\right| + \left|Grad_y\right|$$
$$\left|Grad_x\right| = \sum_{j=o}^{M-1}\sum_{i=0}^{N-2}\left(\left|Y_{j,i} - Y_{j,i+1}\right| + \left|U_{j,i} - U_{j,i+1}\right| + \left|V_{j,i} - V_{j,i+1}\right|\right) \tag{6.12}$$
$$\left|Grad_y\right| = \sum_{j=o}^{M-2}\sum_{i=0}^{N-1}\left(\left|Y_{j,i} - Y_{j+1,i}\right| + \left|U_{j,i} - U_{j+1,i}\right| + \left|V_{j,i} - V_{j+1,i}\right|\right)$$

In (6.12), *M, N* is the number of rows and columns, respectively. Y is luminance value. U and V represent chrominance values. The primary reason choosing gradient information to measure image activity is that gradient magnitude reflects image change intensity. This property is inherently compatible with the purpose of intra coding: applying intra prediction so that only spatial differences are to be coded. Another advantage is that gradient information can be obtained before actually coding the frame at costs of low additional computational complexity.

After selecting the activity measurement, curve fitting was performed on extensive testing data to derive the source R-Q model as

$$QP = \upsilon_1 * \ln(R / C) + \upsilon_0 \tag{6.13}$$

where υ_1 and υ_0 are two model parameters.

To show effectiveness of the proposed model, the relationship between QP and resulting I frame coding bits *R* from various test sequences is plotted in Fig. 6.2. For each plot, the frame is encoded as I frame with QP values ranging from 15 to 50. As seen in Fig. 6.3, R-Q curves resulting from (6.13) are very close to the data retrieved from actual coding results.

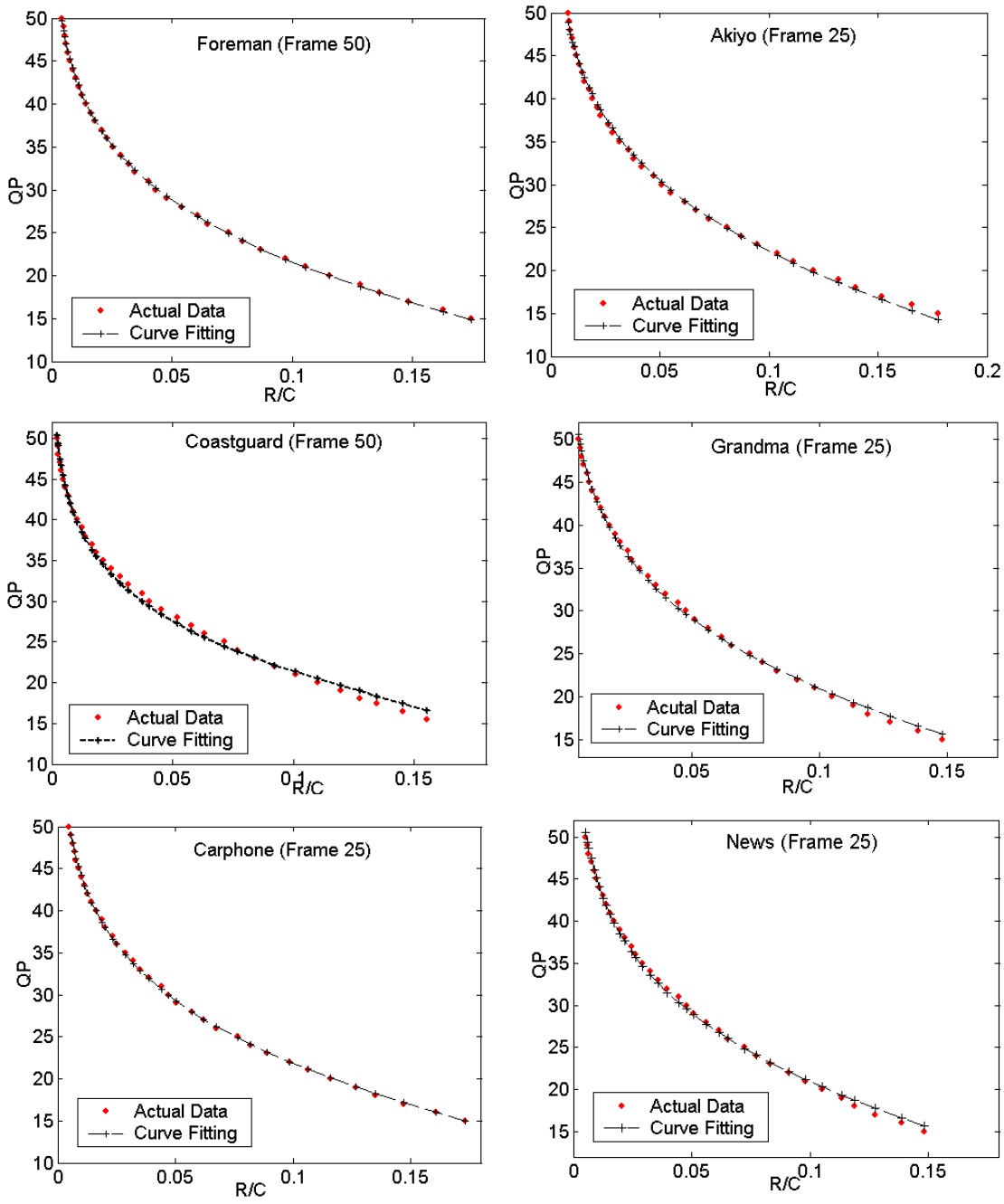


Fig. 6.3. I Frame R-Q relationship modeled by the Logarithmic model (6.13)

6.3. Rate Control Framework with enhanced Bit Allocation

Bit allocation is an important component of a rate control algorithm. In JVT-G012 [49], bit allocation is achieved with the fluid-flow traffic model proposed in [41]. Compared with previous bit allocation algorithms, the fluid-flow traffic algorithm is simple but effective. However, in practice, this algorithm still has some drawbacks. For instance, for consequent GOPs after the first one, I frame bit allocation in [49] did not consider buffer status at the beginning of the current GOP. Coding bits of different intra frames can change significantly, as the image complexity varies. The simple decision of I frame QP as average P frame QP from previous GOP can cause buffer overflow at the beginning of each GOP because excessive number of bits may be generated due to a small quantization. As a result, consequent frames have to be skipped to maintain the buffer status. This is undesirable as skipped frames can result in motion discontinuity and degrade both objective and subjective quality of the reconstructed video. To avoid this problem, I frame R-Q relationship and current buffer status should be considered to decide the I frame quantization adaptively.

In this section, framework of the proposed rate control mechanism will be presented in details. Since the bit allocation part is an improvement on the original H.264/AVC bit allocation, only the different parts, which are P frame basic unit control, I frame bit

allocation and quantization decision, will be focused. In addition, how to update model parameters will also be addressed as well.

6.3.1 Bit Allocation at GOP Level

GOP level rate control allocates bits for all non-coded frames in current GOP. At the beginning of a GOP, the total number of bits T_0 to be assigned for current GOP is computed as follows [49]

$$T_0 = \frac{C}{f} \times N_{GOP} - V_0 \tag{6.14}$$

where C is the target output bit rate, N_{GOP} is the number of pictures in a GOP, and V_0 is virtual buffer occupancy at the beginning of current GOP. For rest pictures in current GOP, the remaining bit budget T_i at the ith frame will be calculated as [49]

$$T_i = T_{i-1} - B_{i-1} \tag{6.15}$$

where B_{i-1} is the number of coding bits from the $i-1$th frame.

6.3.2 I Frame Rate Control

To decide I frame quantization, frame level budget is first assigned according to complexity estimation. With the restriction of reducing buffer overflow and frame skipping, quantization is finalized according to the I frame R-Q model.

6.3.2.1. I Frame Bit allocation

Under the assumption that adjacent GOPs are of similarity features, frame level coding complexity in current GOP is predicted by known frame coding complexity from previous GOP.

The budget dispatched to the I frame in current GOP is decided by

$$B_I = T_0 \frac{W_I}{W_I + kW_p \times N_p} \tag{6.16}$$

where W_I and W_P is the coding complexity prediction of I frame and P frame from previous GOP, respectively, and N_p is the number of P frames in a GOP. k is a weighted factor set to be 1.35 here. Both W_I and W_p are calculated as a product of its corresponding average number of coding bits and average value of quantization parameters. For the first GOP, W_I / W_P is set to be 15 to allow the initial I frame bit allocation.

6.3.2.3. I Frame QP Decision

To use the source model for the purpose of reducing buffer overflow and minimizing undesirable frame skipping, I frame quantization is decided in an iterative way. With the assigned bit budget and activity measurement, an initial quantization is computed by (6.13). To judge if buffer overflow will occur or not, the buffer status V_I after coding the I frame will be predicted by

$$V_1 = V_0 + \exp(QP - \upsilon_0/\upsilon_1) \times C - C / f \tag{6.17}$$

where $R = \exp(QP - \upsilon_0/\upsilon_1) \times C$ is the I frame bit estimation with current QP, and C/f is buffer output rate.

If buffer overflow is anticipated, QP will be increased by 1 to allow coarser quantization. Then the number of bits generated from this new QP will be calculated again to decide whether the buffer constraint is met or not. This process continues until a proper QP value is found. Though it's always possible to increase QP until no buffer overflow occurs, an over-large quantization can harm the overall coding performance as the I frame is coded at low quality. Therefore, the QP difference between current I frame and previous I frame will be further restricted to less or equal to 3 to ensure that no severe quality degradation and discontinuity are caused.

In (6.13), the initial value of υ_1 and υ_1 are set to be -9.0 and -1.0, respectively, and they will be updated by using the method in 6.4.4.

6.3.3 P Frame Rate Control

6.3.3.1. P Frame Bit Allocation

Before coding the *i*th P frame in current GOP, the algorithm allocates target bit assignment R_i as a convex linear combination of $\hat{R}_i$ and $\tilde{R}_i$

$$R_i = \beta \times \hat{R}_i + (1-\beta) \times \tilde{R}_i \tag{6.18}$$

where $\hat{R}_i$ and $\tilde{R}_i$ is defined the same as in [49]. $\hat{R}_i$ is the target bit assignment considering remaining bits T_i in current GOP and $\tilde{R}_i$ is the target bit assignment at sequence level. β is a weighted factor set to be 0.5 here.

6.3.3.2. SATD Prediction

In H.264/AVC, a well known problem is the chicken and egg dilemma between rate control and RD optimization. To decouple the dilemma, a linear model is used to predict current basic unit SATD by using SATD of the co-located basic unit from previous frame. Let $SATD_p$ and $SATD_a$ denote predicted SATD of current basic unit and SATD of the co-located basic unit from previous frame, respectively. The linear prediction model is then given by

$$SATD_p = \gamma_1 \times SATD_a + \gamma_2 \tag{6.19}$$

where γ_1 and γ_2 are two coefficients of the prediction model, and their initial values are set to 1 and 0, respectively. γ_1 and γ_2 are updated after coding each basic unit.

6.3.3.3 P Frame Basic Unit Quantization Decision

After the basic unit bit budget is allocated [49], quantization will be decided by using the P frame R-Q model with the assigned bit budget and the predicted $SATD_p$. For P frame, the overall bit budget accounts not only for DCT coefficient bit, but also for header bit R_H and motion vector bit R_{MV}. Therefore, the number of bits assigned to DCT coefficients should be further shaped by

$$R_{cf} = R_{overall} - R_{MV} - R_H \tag{6.20}$$

The corresponding quantization Q can be computed as

$$Q = \frac{SATD_p}{\beta_1 R_{cf} + \beta_2 \sqrt{R_{cf}} + \beta_0} \tag{6.21}$$

6.3.4 Post Processing

Though the R-Q models proposed in section 6.2 and 6.3 are capable of matching actual R-Q curves precisely, it's not enough to have fixed model parameters for different video frames. The parameters in these models need to be adaptively adjusted to fit the current video contents. Therefore, after coding stage, the encoder has to update parameters in R-Q models (6.11), (6.13) and SATD prediction formula (6.19).

First, data points from past coding units are selected using the sliding window mechanism. That is, after a new unit is encoded, its corresponding data will be stored to update the data points in the sliding window. Next, for these selected data points within the sliding window, the encoder collects corresponding components, such as R and Q. Using the linear regression technique, the updated model parameters can be obtained by minimizing the overall difference between model outputs and actual data values. Finally, after new model parameters are derived, the encoder performs further a refinement step. By applying the refinement process, some outlier data points will be excluded. The final model parameters are derived using the same formula in previous step based on these new data points.

The updated model parameters β_1, β_2 and β_0 can be obtained as

$$\beta_1 = \frac{A(DN-F^2)-E(CN-DF)+G(CF-D^2)}{B(DN-F^2)-C(CN-DF)+D(CF-D^2)}$$
$$\beta_2 = \frac{B(EN-FG)-C(AN-DG)+D(AF-ED)}{B(DN-F^2)-C(CN-DF)+D(CF-D^2)}$$
$$\beta_0 = \frac{B(DG-EF)-C(CG-AF)+D(CE-DA)}{B(DN-F^2)-C(CN-DF)+D(CF-D^2)} \quad (6.22)$$

where

$$A = \sum_{i=1}^{N} \frac{SATD_i}{\Delta_i} R_i, \quad B = \sum_{i=1}^{N} R_i^2, \quad C = \sum_{i=1}^{N} \sqrt[2]{R_i} R_i$$
$$D = \sum_{i=1}^{N} R_i, \quad E = \sum_{i=1}^{N} \frac{SATD_i}{\Delta_i} \sqrt[2]{R_i}, \quad F = \sum_{i=1}^{N} \sqrt[2]{R_i} \quad (6.23)$$
$$G = \sum_{i=1}^{N} \frac{SATD_i}{\Delta_i}$$

For I frame R-Q model update, the goal is minimize the following equation

$$\min\left(\sum_{i=1}^{n} (QP_i - (\upsilon_1 * \ln(R_i / C_i) + \upsilon_0))^2\right) \quad (6.24)$$

where n is the total number of data points.

By taking the first order derivation of (6.24)

$$\frac{\partial}{\partial \upsilon_1} = r_1\left(\sum_{i=1}^{n} x_i^2\right) + r_0\left(\sum_{i=1}^{n} x_i\right) - \sum_{i=1}^{n} QP_i x_i = 0$$
$$\frac{\partial}{\partial \upsilon_2} = r_1\left(\sum_{i=1}^{n} x_i\right) + r_0(n) - \sum_{i=1}^{n} QP_i = 0 \quad (6.25)$$
$$x_i = \ln(R_i / C_i)$$

The updated I frame R-Q model parameters γ_1 and γ_2 can be obtained as

$$\upsilon_1 = \frac{\left(\sum_{i=1}^{n} QP_i x_i\right)(n) - \left(\sum_{i=1}^{n} QP_i\right)\left(\sum_{i=1}^{n} x_i\right)}{\left(\sum_{i=1}^{n} x_i^2\right)(n) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)}$$
$$\upsilon_2 = \frac{\left(\sum_{i=1}^{n} QP_i\right)\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} QP_i x_i\right)\left(\sum_{i=1}^{n} x_i\right)}{\left(\sum_{i=1}^{n} x_i^2\right)(n) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)} \quad (6.26)$$

6.3.5 Frame Skipping Strategy

Coded frames are lost when buffer overflow happens, which can cause severe error propagation as the lost frames can be used as reference frames. If buffer overflow is anticipated, some of future frames are usually skipped to prevent overflow from happening.

While there are different frame skipping strategies, a well recognized method from MPEG VM8 [36,37] rate control algorithm will be adopted here. Current buffer status is checked before coding the next frame. If buffer occupancy exceeds 80% of the buffer capacity, the next frame will be skipped and buffer fullness is updated by subtracting. This process continues until the buffer level is below 80% of the buffer capacity.

6.4. Simulation Results

In this section, the proposed rate control framework is evaluated by comparing its performance with JVT-G012 rate control algorithm [49]. The implementation of both rate control algorithms is based on H.264/AVC reference software JM12.3.

Two sets of experiments are conducted to compare two algorithms in different aspects. First, I frame bit allocation and quantization decision are turned off for the proposed algorithm so that the difference between two algorithms exists in the P frame basic unit quantization decision part. In this experiment, we focus on the performance improvement by adopting the new P frame R-Q model without changing the bit allocation part. In the

second experiment, with constrained buffer capacity, proposed I frame rate control part is added to see how it can provide further improvement on buffer management by preventing overflow.

6.4.1 P Frame R-Q Model Comparison

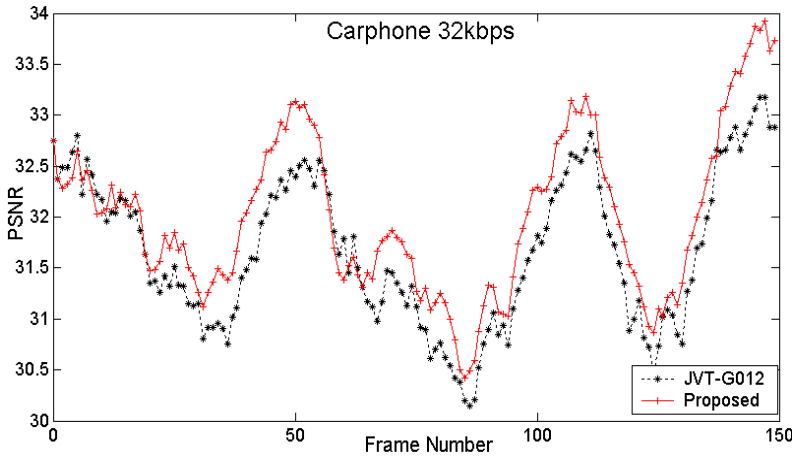
In this experiment, only the first frame is coded as I frame and all remaining frames are inter-coded P frames (IPPPPP….). Because JVT-G012 is used as the benchmark, the I frame quantization is determined by the same rule in JVT-G012 for the sake of fair comparison. Moreover, there is no buffer size specified, which indicates that potential buffer overflow and underflow are implicitly omitted. As a result, no frame skipping is performed as well. Video sequences are coded at 30 fps. All possible macroblock partition modes are enabled with the best mode decision by RDO. Search range is 32 pixels and the number of reference frame is 1. Basic unit is set to be one macroblock.

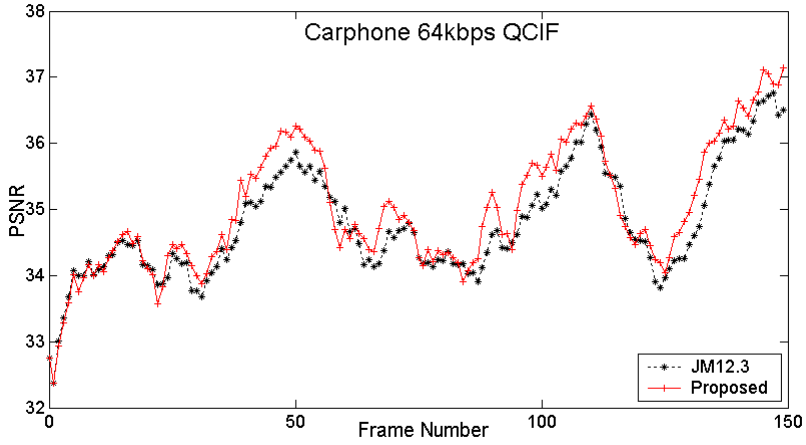
Video sequences of different motion intensities and formats are tested for both algorithms at various target bit rates. Table 2 presents performance comparison results from this experiment in terms of average luminance PNSR and actual bit rates. As seen, when compared with JVT-G012 [49], the proposed basic unit control with square root R-D model gives a smaller bit rate mismatch for major video sequences. Furthermore, the proposed rate control framework generates higher average luminance PSNR for all reported test sequences at different target bit rates. An increase of luminance PSNR up to

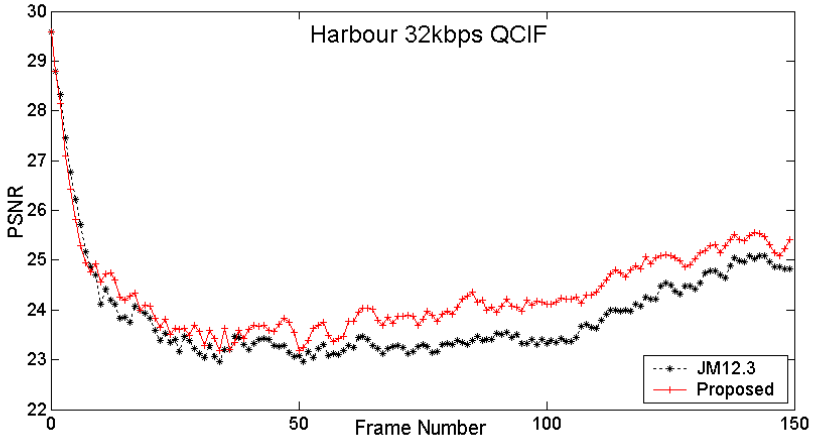
0.47 dB for coded QCIF videos and up to 0.34 dB for coded CIF videos can be found from Table 6.2.

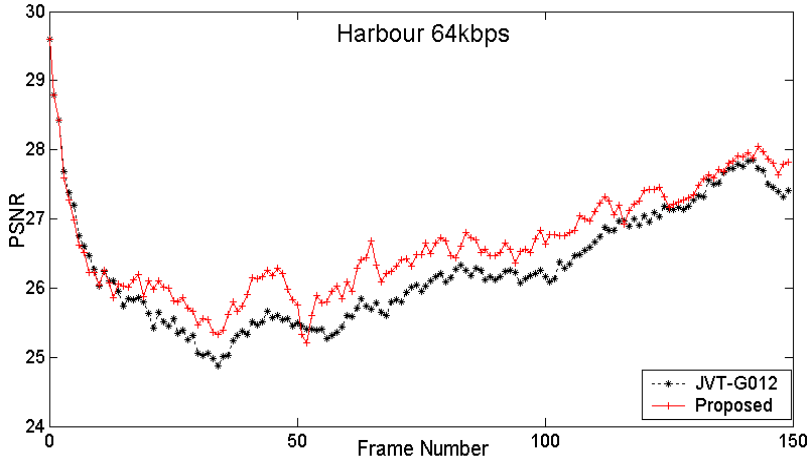
Frame by frame PSNR comparison of reconstructed video sequences is also plotted in Fig. 6.4. From the plot, the PSNR improvement by the proposed rate control scheme over JVT-G012 [49] is evident.

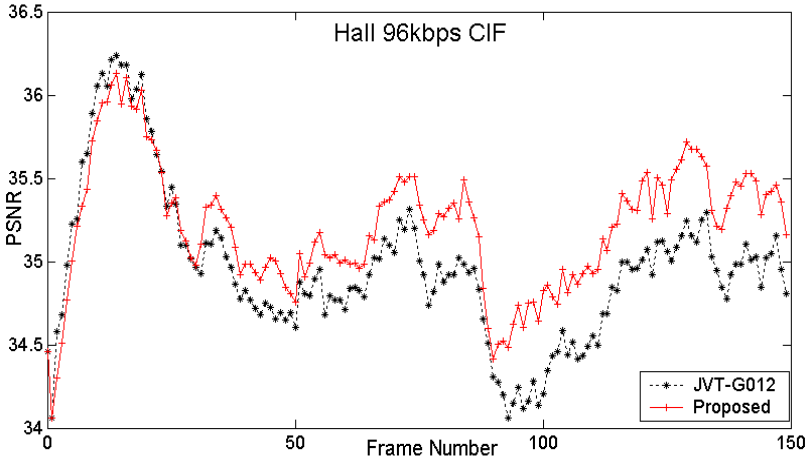
Please note the performance of a rate control algorithm can be influenced by many factors. In this experiment, we exclude other factors so that the performance comparisons can be focused on basic unit quantization decision. For this purpose, the bit allocation part of two rate control algorithms is kept identical and so as other factors. From the comparison results, it is fair to conclude that major contribution to video quality improvement in terms of higher PSNR is from the proposed basic unit quantization part because the new model can provide more accurate R-Q relationship estimation.

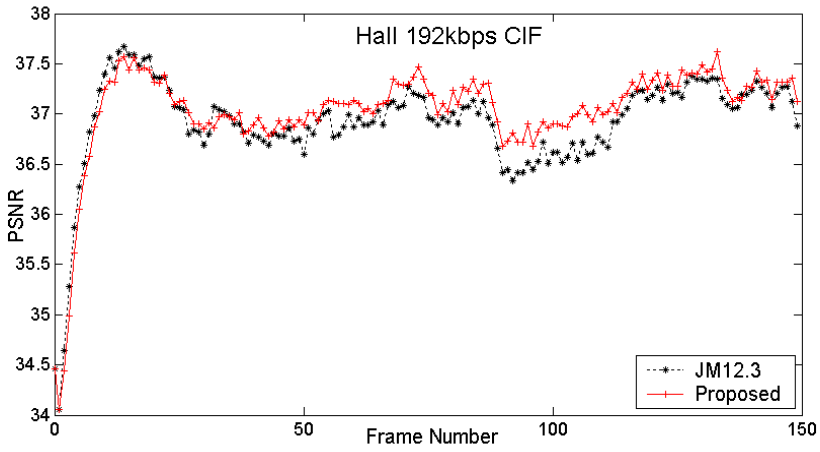


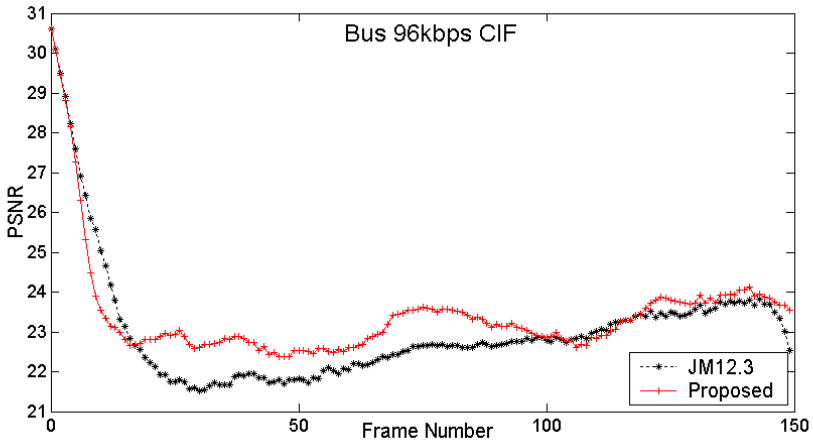

Carphone 64kbps QCIF
PSNR
Frame Number
JM12.3
Proposed


Harbour 32kbps QCIF
PSNR
Frame Number
JM12.3
Proposed


Harbour 64kbps
PSNR
Frame Number
JVT-G012
Proposed


Hall 96kbps CIF
PSNR
Frame Number
JVT-G012
Proposed
36.5
36
35.5
35
34.5
34
0
50
100
150


Hall 192kbps CIF
PSNR
Frame Number
JM12.3
Proposed
38
37.5
37
36.5
36
35.5
35
34.5
34
0
50
100
150


Bus 96kbps CIF
PSNR
Frame Number
JM12.3
Proposed
31
30
29
28
27
26
25
24
23
22
21
0
50
100
150

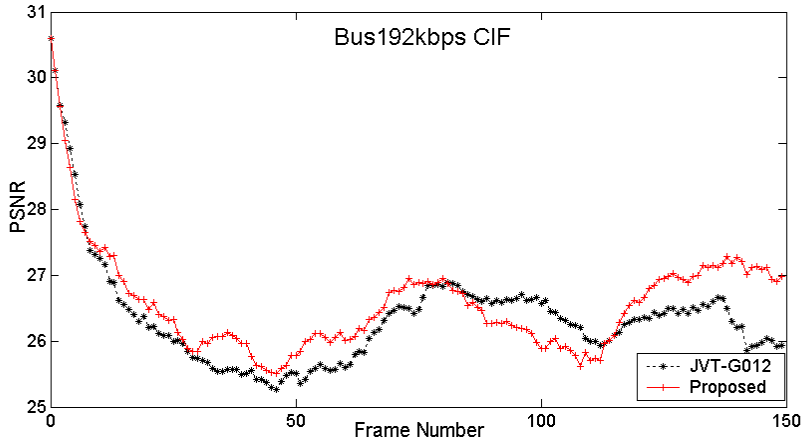


Fig. 6.4. Frame by frame PSNR comparison at different bit rates (Experiment 1)

6.4.2 I Frame Quantization Decision Evaluation

In this experiment, I frame quantization decision strategy is enabled for the proposed algorithm. Buffer constraint is applied for both algorithms with a size of 1/3 bit rate, i.e., the maximal buffer delay is limited to 333ms. The same frame skipping strategy described in 6.3.5 is adopted by both algorithms. I frame is inserted every 30 frames and the rest configuration parameters are kept the same as that in experiment 1. To maintain decoder side output rate, a skipped frame will be concealed by inserting its previous coded frame until the next coded frame appears [36,37].

Table 6.2 Rate control performance comparison (Experiment 1)

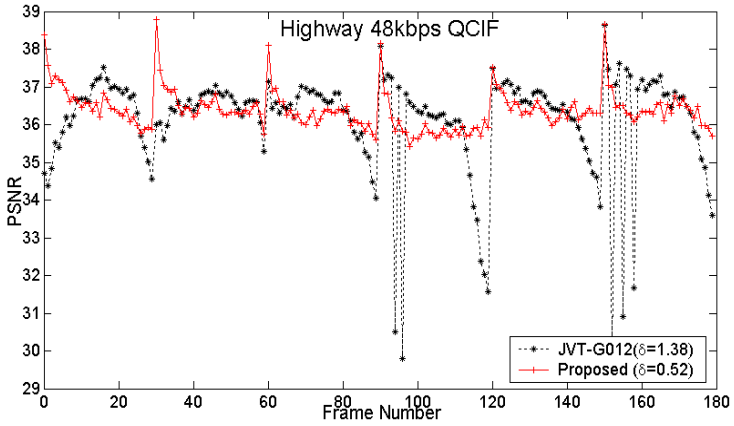
Test Sequence	Format	Target Bitrate (kb/s)	JVT-G012		Porposed		
			Actual Bitrate (kb/s)	PSNR (dB)	Actual Bitrate (kb/s)	PSNR (dB)	PSNR Gain (dB)
Salesman	QCIF	32	32.11	31.96	31.99	32.24	0.28
		64	64.11	35.75	64.07	35.96	0.21
Foreman	QCIF	32	32.21	29.09	32.01	29.23	0.14
		64	64.25	32.69	63.87	32.83	0.14
News	QCIF	32	32.18	31.49	32.1	31.81	0.32
		64	64.34	35.42	64.18	35.83	0.41
Carphone	QCIF	32	32.13	31.67	32.07	32.01	0.34
		64	64.2	34.79	64.1	35.01	0.22
Highway	QCIF	32	32.19	35.04	32.29	35.26	0.22
		64	64.19	37.23	64.22	37.3	0.07
Harbour	QCIF	32	32.16	23.91	32.03	24.38	0.47
		64	64.25	26.3	64	26.64	0.34
Mobile	QCIF	32	32.12	21.61	32.1	21.68	0.07
		64	64.22	24.36	64.03	24.42	0.06
bus	CIF	96	96.43	23.04	96.12	23.42	0.38
		192	192.16	26.57	192.61	26.61	0.04
Hall	CIF	96	96.35	34.96	96.2	35.21	0.25
		192	192.29	36.92	192.18	37.03	0.11
Akiyo	CIF	96	96.21	39.13	96.11	39.19	0.06
		192	192.24	41.88	192.3	41.88	0
Foreman	CIF	96	96.43	31.06	96.04	31.13	0.07
		192	192.85	34.02	192.15	34.11	0.09
Mother-Daughter	CIF	96	97.06	37.23	97.56	37.27	0.04
		192	194.06	40.05	193.88	40.09	0.04
News	CIF	96	96.44	33.85	96.33	33.94	0.09
		192	193.05	37.29	192.81	37.6	0.31

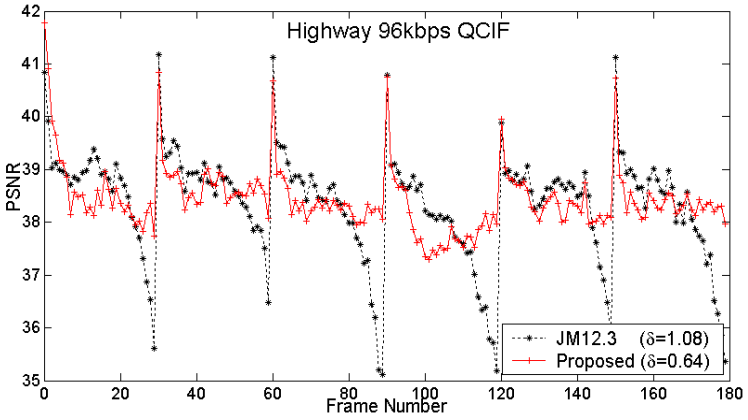
From the experimental results in Table 6.3, several important observations can be found. With the consideration of current and future buffer status, I frame quantization is decided more reasonably by the proposed framework so that the number of skipped frames are greatly reduced. Please note as the QP difference is restricted to 3, for sequences such as carphone (QCIF) and mother-daughter (QCIF), some frames always

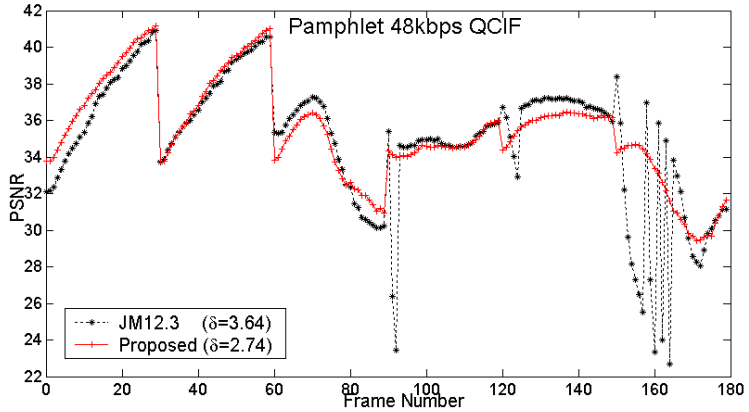
have to be skipped. In addition, as the I frame QP value decided by the proposed method takes the frame level complexity into consideration, the bit dispatch between the I frame and P frame are more reasonable. Hence, the performance of coded sequences is improved as well. Table 6.3 shows up to 0.82dB luminance PSNR increase for the coded sequences using the proposed rate control framework.

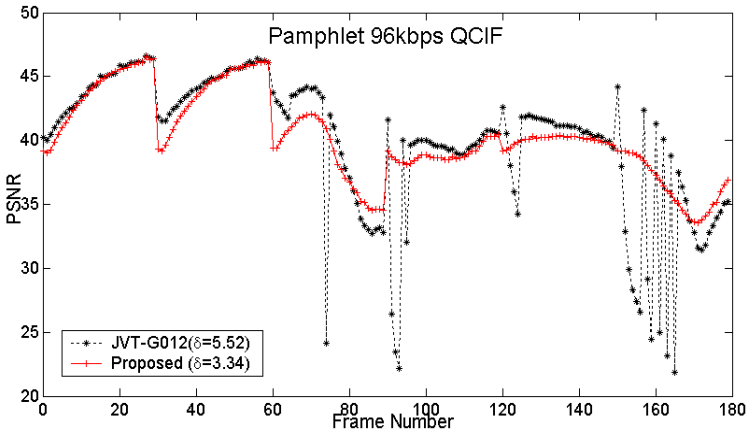
Fig. 6.5 further illustrates that the improved I frame quantization decision also generates much smoother video output in terms of smaller PSNR variation (δ). A skipped frame is of much worse quality than a normally coded frame as the concealed frame only copies from previous frame. The proposed algorithm can efficiently prevent this severe quality fluctuation from happening. This helps to provide a better viewing experience since viewers usually tend to rate the video quality along frames of worse quality.

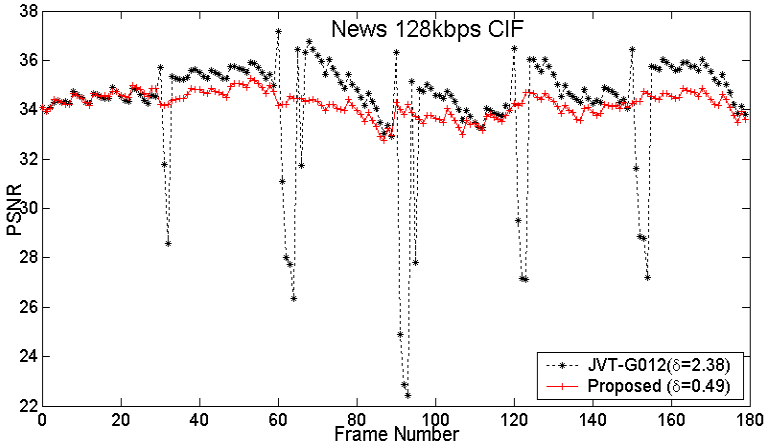
In Fig. 6.6, buffer status is also given. As seen, the proposed I frame quantization decision can efficiently adjust quantization so that the buffer level is always below the skipping threshold. As a result, no frame skipping is performed. Meanwhile, instead of over-increase the quantization, the algorithm seeks for a finest quantization that fits the buffer constraint to achieve best overall coding performance.

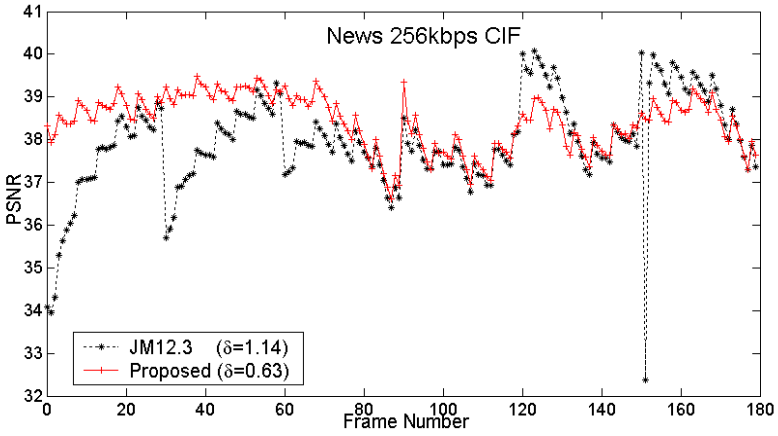

Highway 48kbps QCIF
PSNR
Frame Number
JVT-G012(δ=1.38)
Proposed (δ=0.52)


Highway 96kbps QCIF
PSNR
Frame Number
JM12.3 (δ=1.08)
Proposed (δ=0.64)


Pamphlet 48kbps QCIF
PSNR
Frame Number
JM12.3 (δ=3.64)
Proposed (δ=2.74)


Pamphlet 96kbps QCIF
PSNR
Frame Number
JVT-G012(δ=5.52)
Proposed (δ=3.34)


News 128kbps CIF
PSNR
Frame Number
JVT-G012(δ=2.38)
Proposed (δ=0.49)


News 256kbps CIF
PSNR
Frame Number
JM12.3 (δ=1.14)
Proposed (δ=0.63)

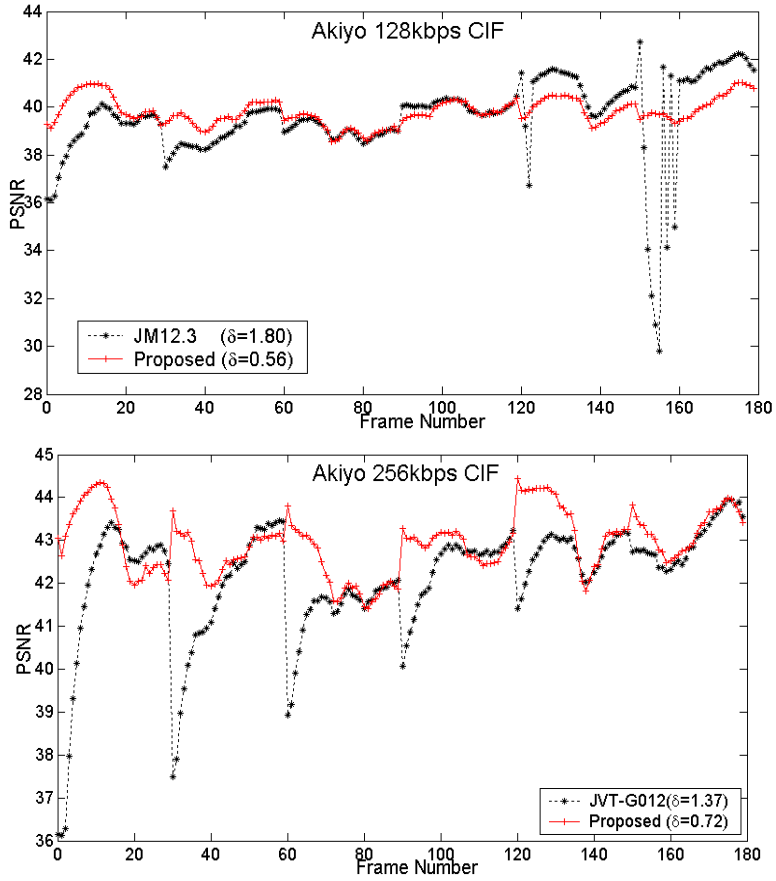
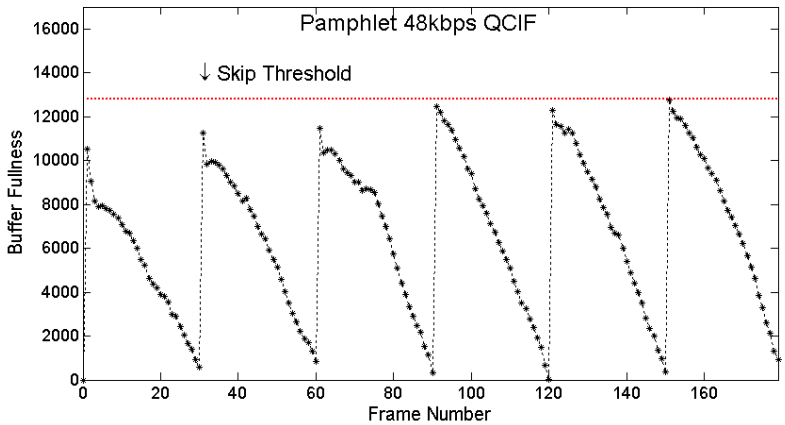


Fig. 6.5. Frame by frame PSNR comparison at different bit rates (Experiment 2)



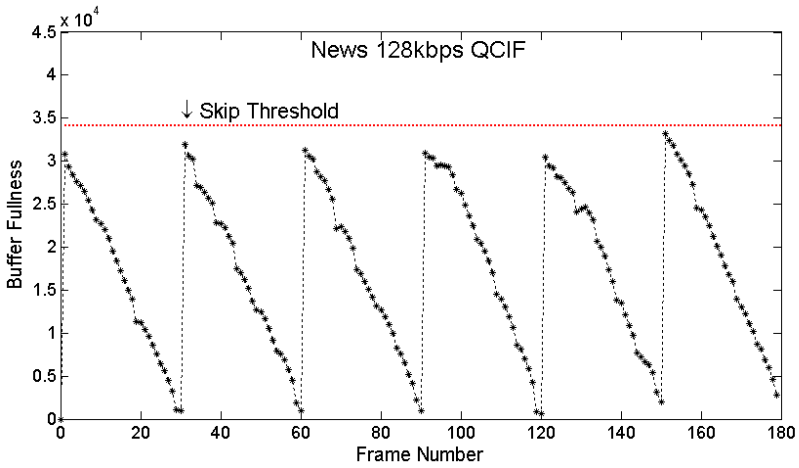


Fig. 6.6. Buffer status (Experiment 2)

6.5 Conclusions

In this chapter, an improved rate control framework was proposed to realize efficient bit allocation and accurate quantization decision for H.264/AVC encoder. The novelty of this rate control framework is of several aspects. A P frame square root rate model based on the ρ domain rate control theory accurately captures P frame R-Q relationship so that efficient P frame quantization decision can be made. The I frame source model with coding activity measurement provides a simple but efficient solution to alleviate the frame skipping problem resulting from JVT-G012 I frame quantization decision. All these new ideas contributed to a better rate control performance comparing to JVT-G012 algorithm. With the only new P frame basic unit quantization decision adopted, a PSNR increase of up to 0.47 dB was observed. When I frame bit allocation and quantization were enabled, the number of skipped frames was evidently reduced and much smoother and higher PSNR was produced by using the proposed rate control

framework. Meanwhile, the proposed algorithm also led to a more precise rate control by decreasing the mismatch between target and actual bit rates.

Table 6.3 Rate control performance comparison (Experiment 2)

Test Sequence	Format	Target Bitrate (Kb/s)	JVT-G012			Proposed				
			Actual Bitrate (kb/s)	PSNR (dB)	Frame Skipped	Actual Bitrate (kb/s)	PSNR (dB)	PSNR Gain(dB)	Frame Skipped	Reduced Frame Skipping
Highway	QCIF	48	48.17	36.11	5	48.2	36.38	0.27	0	-5
		96	96.25	38.33	0	96.25	38.44	0.11	0	0
Carphone	QCIF	48	48.27	33.43	7	48.14	33.59	0.16	1	-6
		96	96.43	36.58	2	96.17	36.69	0.11	1	-1
Mother-Daughter	QCIF	48	48.29	36.02	16	48.15	36.38	0.36	8	-8
		96	96.4	39.7	12	96.23	39.84	0.14	4	-8
Foreman	QCIF	48	48.14	30.31	0	48.03	30.46	0.15	0	0
		96	96.28	33.99	2	96.07	34.22	0.23	0	-2
Pamphlet	QCIF	48	48.32	34.84	18	48.09	35.38	0.54	0	-18
		96	96.53	39.52	26	95.98	40.27	0.75	0	-26
Akiyo	CIF	128	128.97	39.56	9	128.44	39.81	0.25	0	-9
		256	256.94	42.13	0	256.84	42.95	0.82	0	0
News	CIF	128	128.68	34.24	18	128.31	34.24	0	0	-18
		256	256.71	37.94	1	256.52	38.42	0.48	0	-1
Container	CIF	128	128.79	32.9	36	128.11	32.98	0.08	0	-36
		256	256.67	35.68	6	256.16	35.95	0.27	0	-6
Foreman	CIF	128	128.38	31.69	2	128.16	31.8	0.11	0	-2
		256	256.73	34.7	0	256.2	34.87	0.17	0	0

Chapter 7
New Techniques for next Generation Video Coding Standard

Though the state of art video coding standard H.264/AVC [15] is reported to achieve superior gains in compression efficiency to its predecessors, the increasing popularity of high definition TV, video delivery on mobile devices, and other multimedia applications never stops creating new demands for video coding standards. To face new challenges that new applications may impose on video coding standards, both MPEG and VCEG launched their next-generation video coding project, which potentially could be either an extension of H.264/AVC or a brand new standard.

In the remainder of this chapter, current standard activities are summarized in Section 7.1, and the existing new coding techniques and their performances are analyzed in Section 7.2. The last part draws the conclusions.

7.1 Current Standard Activities

At the 86th MPEG meeting in Busan 2008, MPEG determined the need for a next generation of video compression technology called High-performance Video Coding (HVC). HVC would be intended mainly for high quality applications, by providing performance in terms of coding efficiency at higher resolutions, with applicability for

entertainment-quality services such as HD mobile, home cinema and Ultra High Definition (UHD) TV [60]. A Call for Evidence was issued that allowed proponents to report about the existence of such technologies. The response to this Call for Evidence was evaluated at the 89th MPEG meeting in July 2009. By the time of this publication, a Draft call for Proposals should have been issued to start the preparation for HVC.

Comparing with HVC, the goals of ITU-T's Next Generation Video Coding (NGVC) project are similar but more detailed. In 2005 to 2008, ITU-T VCEG studied the requirement definition for NGVC, and some agreements about the goals of the NGVC project were reached during the 37th VCEG meeting [61], with primary emphasis on computational efficiency and high compression performance. For instance, in terms of coding efficiency, NGVC should be capable of providing 50% bit rate savings over H.264/AVC at the same video quality representation. To address the concern of complexity, NGVC should be capable of operating with a complexity ranging from 50% to 3 times H.264/AVC High Profile. More specifically, when operated at a complexity of 50% compared to H.264/AVC High Profile, NGVC should provide a 25% bit rate savings compared to H.264/MPEG-4 AVC High Profile at equivalent subjective quality [3].

Though MPEG and VCEG could independently create separate next generation video coding standards, two new standards of similar functionalities might not be welcomed by industry. Based on the previous success in jointly creating H.264/AVC, future

collaboration on NGVC and HVC, similar to the Joint Video Team effort, was considered during a joint meeting of MPEG and VCEG in July 2009.

7.2 Proposed new Techniques and Performance Analysis

The advances of video coding techniques were contributed by various parties. To provide a software platform to gather and evaluate these new techniques, a Key Technical Area (KTA) [62] platform was developed based on JM11 reference software, where the new coding tools are continuously added. So far, the major new coding tools added to KTA platform can be summarized as follows:

1. Intra Prediction: In [63], H.264 intra prediction is enhanced with additional Bi-directional Intra Prediction (BIP) modes, where BIP combines prediction blocks from two prediction modes using a weighting matrix. Furthermore, Mode-Dependent Directional Transform (MDDT) using transforms derived from KLT is applied to capture the remaining energy in the residual block.

2. Inter Prediction: To further improve inter prediction efficiency, finer fractional motion prediction and better motion vector prediction were proposed. Increasing the resolution of the displacement vector from 1/4-pel to 1/8-pel to obtain higher efficiency of the motion compensated prediction is suggested in [64]. In [65], a competing framework for better motion vector coding and SKIP mode is proposed, where both spatial and temporal redundancies in motion vector fields are captured. Moreover, [66]

suggests extending the macroblock size up to 64x64 so that new partition sizes 64x64, 64x32, 32x64, 32x32, 32x16, and 16x32 can be used. Instead of using the fixed interpolation filter from H.264/AVC, Adaptive Interpolation Filters (AIF) are proposed, such as 2D AIF [67], Separable AIF [68], Directional AIF [69], Enhanced AIF [70], and Enhanced Directional AIF [71].

3. Quantization: To achieve better quantization, optimized quantization decision at the macroblock level and at different coefficient positions are proposed. Rate Distortion Optimized Quantization (RDOQ), which performs optimal quantization on a macroblock, was added to the JM reference software. RDOQ does not require a change of H.264/AVC decoder syntax. More recently, [72] gives an improved, more efficient RDOQ implementation. In [73], Adaptive Quantization Matrix Selection (AQMS), a method deciding the best quantization matrix index, where different coefficient positions can have different quantization steps, is proposed to optimize the quantization matrix at a macroblock level.

4. Transform: For motion partitions bigger than 16x16, a 16x16 transform is suggested in addition to 4x4 and 8x8 transforms [66]. Moreover, transform coding is not always a must. In [74], it is proposed that for each block of the prediction error, either standardized transform coding or spatial domain coding can be adaptively chosen.

5. In-loop Filter: In KTA, besides the deblocking filter, an additional Adaptive Loop Filter (ALF) is added to improve coding efficiency by applying filters to the deblocking-filtered picture. Two different ALF techniques are adopted so far:

Quadtree-based Adaptive Loop Filter (QALF) [75] and Block-based Adaptive Loop Filter (BALF) [76].

6. Internal bit depth increase: By using 12 bits of internal bit depth for 8-bit sources, so that the internal bit depth is greater than the external bit depth of the video codec, the coding efficiency can be further improved [77].

Besides the techniques listed above, there are some noticeable contributions not added to KTA yet. For example, [78-80] proposed three methods, respectively, to use Decoder Side Motion Estimation (DSME) for B frame motion vector decision, which improves coding efficiency by saving bits on B frame motion vector coding. Also, some new techniques are under investigation and will be presented in the responses for call for proposals.

7.3 Conclusions

The Call for Evidence for HVC provided results that averaged a 15-25% gain in coding efficiency. While not enough to constitute a new standard, the evidence was sufficient to warrant issuing a call for proposals. It is expected that over the next several months, these new technologies will advance to the point where the ITU and MPEG can begin work on a new joint video coding standard.

Chapter 8
Summary

In this chapter, the completed research work is summarized first, and then a plan for future work is given.

8.1 Completed Research

In this dissertation, two problems were addressed: reduce the H.264/AVC coding complexity and provide rate control solutions for H.264/AVC. Until now, the following research work has been done.

1. Proposed a bit estimation method to reduce corresponding computational load of entropy coding. By estimating the number of bits to encode residual signal with a reasonable model, our proposed method facilitates the fast implementation of RD optimization.
2. Proposed a fast mode decision algorithm that is suitable to early eliminate possible inter mode choices for intra refresh based error resilient applications. With our proposed method, the encoder speed can be accelerated up to 10 times while similar or better reconstructed video quality can be obtained.
3. Designed a frame level bit rate allocation algorithm based on a two pass H.264/AVC encoder structure to reduce quality fluctuation resulted from

semi-constant rate bit allocation rate control. Experimental results show that the algorithm can reduce standard deviation of PSNR by up to 1/3 at trivial average PSNR impact.

4. Derived a novel basic unit square root R-Q model based on ρ-domain rate control theory and I frame source rate model with frame activity measurement. When compared with the rate control scheme JVT-G012, the proposed rate control algorithm shows good coding performance in terms of improved average luminance PSNR, reduced number of skipped frames, and smaller mismatch between target and actual bit rates.

8.2 Future Work

As introduced in Chapter 7, the dawn of the next generation video coding standard can be seen now. In the future work, the research will be continued on seeking for tools providing higher coding efficiency. Some possible directions include:

1. SOP (Second Order Prediction). Apply an additional intra coding to inter prediction residual to further exploit spatial redundancy.
2. RRU (Reduced Resolution Update). Reduce the size of residual block. RD optimization is introduced as the rule to obtain best tradeoff between distortion and bit saving.

BIBLIOGRAPHY

[1] C.E. Shannon, "Coding Theorems for a Discrete Source with a Fidelity Criterion," *IRE Nat. Conv. Rec.*, vol. 7, part 4, Mar. 1959, pp. 142–163.

[2] C.E. Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, vol. 27, 1948, pp. 379–423, 623–656.

[3] J. Lan, W. Zeng, and X. Zhuang, "Operational distortion-quantization curve-based bit allocation for smooth video quality," *Journal of Visual Communication and Image Representation*, vol. 16, 2005, pp. 527-543.

[4] D. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," *Proceedings of the IRE*, vol. 40, 1952, pp. 1098-1101.

[5] I.H. Witten, R.M. Neal, and J.G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, vol. 30, 1987, pp. 520-540.

[6] N. Ahmed, T. Natarajan, and K.R. Rao, "Discrete Cosine Transfom," *IEEE Trans. Comput.*, vol. 23, 1974, pp. 90-93.

[7] K. Ramchandran, M. Vetterli, and C. Herley, "Wavelets, subband coding, and best bases," *Proceedings of the IEEE*, vol. 84, 1996, pp. 541-560.

[8] "Information technology—coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s," *ISO/IEC 11172-2*, 1993.

[9] "Generic coding of moving pictures and associated audio information—Part 2: Video," *ISO/IEC 13818–2*, Geneva 1993.

[10] "Information technology—coding of audiovisual objects—part 2: visual," *ISO/IEC 14496–2*, Geneva, 2000.

[11] "Video codec for Audiovisual Services at p x 64 kbit/s," *ITU-T Recommendation H.261*, 1990.

[12] "Video Coding for Low bit rate Communication," *ITU-T Recommendation H.263*, November 1995.

[13] "Advanced Video Coding for Generic Audiovisual Services v5," *ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-AVC)*, July 2005.

[14] D. Marpe, T. Wiegand, and G. Sullivan, "The H.264/MPEG4 advanced video coding standard and its applications," *Communications Magazine, IEEE*, vol. 44, 2006, pp. 134-143.

[15] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, 2003, pp. 560-576.

[16] J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: tools, performance, and complexity," *Circuits and Systems Magazine, IEEE*, vol. 4, Mar. 1993, pp. 7-28.

[17] M. Sarwer and Lai-Man Po, "Fast Bit Rate Estimation for Mode Decision of H.264/AVC," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, 2007, pp. 1402-1407.
[18] Zhenyu Wei and King Ngi Ngan, "A Fast Rate-Distortion Optimization Algorithm for H.264/AVC," *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 2007, pp. I-1157-I-1160.
[19] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, 2003, pp. 620-636.
[20] G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *Signal Processing Magazine, IEEE*, vol. 15, 1998, pp. 74-90.
[21] A.N. Netravali and B.G. Haskell, *Digital Pictures: Representation and Compression*, Perseus Publishing, 1988.
[22] E. Lam and J. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *Image Processing, IEEE Transactions on*, vol. 9, 2000, pp. 1661-1666.
[23] Yao Wang and Qin-Fan Zhu, "Error control and concealment for video communication: a review," *Proceedings of the IEEE*, vol. 86, 1998, pp. 974-997.
[24] S. Wenger, G. Knorr, J. Ott, and F. Kossentini, "Error resilience support in H.263+," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 8, 1998, pp. 867-877.
[25] R. Aravind, M. Civanlar, and A. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 6, 1996, pp. 426-435.
[26] F. Kossentini, "Optimal intra coding of blocks for robust video communication over the internet," *IMAGE COMMUN*, vol. 15, 1999, pp. 25--34.
[27] P. Haskell and D. Messerschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, 1992, pp. 545-548 vol.3.
[28] J. Liao and J. Villasenor, "Adaptive intra update for video coding over noisy channels," *Image Processing, 1996. Proceedings., International Conference on*, 1996, pp. 763-766 vol.3.
[29] R. Zhang, S. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *Selected Areas in Communications, IEEE Journal on*, vol. 18, 2000, pp. 966-976.
[30] Yuan Zhang, Wen Gao, Huifang Sun, Qingming Huang, and Yan Lu, "Error resilience video coding in H.264 encoder with potential distortion tracking," *Image Processing, 2004. ICIP '04. 2004 International Conference on*, 2004, pp. 163-166 Vol. 1.
[31] T. Stockhammer, D. Kontopodis, and T. Wieg, "Rate-Distortion Optimization for JVT/H.26L Video Coding in Packet Loss Environment," *International Packet Video*

Workshop, Germany, 2002.
[32] A.J. Viterbi and J.K. Omura, *Principles of Digital Communication and Coding*, McGraw-Hill, Inc., 1979.
[33] L. Merritt and R. Vanam, "Improved Rate Control and Motion Estimation for H.264 Encoder," *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 2007, pp. V - 309-V - 312.
[34] S. Wenger, "Error patterns for Internet video experiments," *ITU-T Study Group 16 H.263+ Video Experts Group, vol. Q15I09*, October 1999.
[35] "TM5: MPEG-2 Test Model 5," *ISO-IEC AVC-491*, April 1993.
[36] "Text of ISO/IEC 14 496-2 MPEG4 video VM—Version 8.0," *ISO/IEC JTC1/SC29/WG11 W1796*, Stockholm, Sweden, 1997.
[37] "Coding of Moving Pictures and Associated Audio," *MPEG-4 Committee Draft, ISO/IEC 14496-2*, October 1998.
[38] G. Cote, B. Erol, M. Gallant, and F. Kossentini, "H.263+: video coding at low bit rates," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 8, 1998, pp. 849-866.
[39] A. Vetro, Huifang Sun, and Yao Wang, "MPEG-4 rate control for multiple video objects," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, 1999, pp. 186-199.
[40] Yu Sun and I. Ahmad, "A robust and adaptive rate control algorithm for object-based video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, 2004, pp. 1167-1182.
[41] Z. Li, Lin Xiao, C. Zhu, and Pan Feng, "A novel rate control scheme for video over the Internet," *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, 2002, pp. 2065-2068.
[42] Bo Xie and Wenjun Zeng, "Sequence-based rate control for constant quality video," *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, pp. I-77-I-80 vol.1.
[43] L. Shen, Z. Liu, Z. Zhang, and X. Shi, "Frame-level bit allocation based on incremental PID algorithm and frame complexity estimation," *J. Vis. Comun. Image Represent.*, vol. 20, 2009, pp. 28-34.
[44] J. Ribas-Corbera and Shawmin Lei, "Rate control in DCT video coding for low-delay communications," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, 1999, pp. 172-185.
[45] N. Kamaci, Y. Altunbasak, and R. Mersereau, "Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, 2005, pp. 994-1006.
[46] Zhihai He, Yong Kwan Kim, and S. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, 2001, pp. 928-940.

[47] Zhihai He and S. Mitra, “Optimum bit allocation and accurate rate control for video coding via /spl rho/-domain source modeling,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, 2002, pp. 840-849.
[48] Zhihai He and S. Mitra, “A linear source model and a unified rate control algorithm for DCT video coding,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 12, 2002, pp. 970-982.
[49] Z. G. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, “Adaptive basic unit layer rate control for JVT,” *JVT-G012*, Pattaya Thailand, March 2003.
[50] Tihao Chiang and Ya-Qin Zhang, “A new rate control scheme using quadratic rate distortion model,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 7, 1997, pp. 246-250.
[51] Hung-Ju Lee, Tihao Chiang, and Ya-Qin Zhang, “Scalable rate control for MPEG-4 video,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 10, 2000, pp. 878-894.
[52] D. Kwon, M. Shen, and C.J. Kuo, “Rate Control for H.264 Video With Enhanced Rate and Distortion Models,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, 2007, pp. 517-529.
[53] Chun-Yuan Chang, Cheng-Fu Chou, Din-Yuen Chan, Tsungnan Lin, and Ming-Hung Chen, “A q-Domain Characteristic-Based Bit-Rate Model for Video Transmission,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, 2008, pp. 1307-1311.
[54] J.L. Devore and N.R. Farnum, *Applied Statistics for Engineers and Scientists*, Duxbury Press, February 2004.
[55] Wook Joong Kim, Jong Won Yi, and Seong Dae Kim, “A bit allocation method based on picture activity for still image coding,” *Image Processing, IEEE Transactions on*, vol. 8, 1999, pp. 974-977.
[56] F. Pan, Z. Li, K. Lim, X. Lin, S. Rahardja, and D. Wu, “Adaptive intra-frame quantization for very low bit rate video coding,” *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, 2004, pp. III-781-4 Vol.3.
[57] L. Czuni, G. Csaszar, and A. Licsar, “Estimating the Optimal Quantization Parameter in H.264,” *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006, pp. 330-333.
[58] Xuan Jing and Lap-Pui Chau, “A novel intra-rate estimation method for H.264 rate control,” *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, 2006, p. 4 pp.
[59] G. Lee, H. Lin, and M. Wang, “Rate control algorithm based on intra-picture complexity for H.264/AVC,” *Image Processing, IET*, vol. 3, 2009, pp. 26-39.
[60] “MPEG Considers the Development of HVC – the Next Generation of Video,” *ISO/IEC JTC 1/SC 29/WG 11 N10117*, Busan (Korean),, October 2008.
[61] G.J. Sullivan, “Informal report of VCEG actions at SG 16 meeting Jan/Feb 2009,”

ITU-T Q.6/SG16 VCEG, VCEG-AK04, Yokohama, Japan, April 2009.
[62] http://iphome.hhi.de/suehring/tml/download/KTA/.
[63] Y. ye and M. Karczewicz, "Improved Intra Coding," *ITU-T Q.6/SG16 VCEG, VCEG-AG11*, Shenzhen, China, October 2007.
[64] J. Ostermann and M. Narroschke, "Motion compensated prediction with 1/8-pel displacement vector resolution," *ITU-T Q.6/SG16 VCEG, VCEG-AD09*, Hangzhou, China, October 2006.
[65] G. Laroche, J. Jung, and B. Pesquet-Popescu, "RD Optimized Coding for Motion Vector Predictor Selection," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, 2008, pp. 1681-1691.
[66] P. Chenn, Y. Ye, and M. Karczewicz, "Video Coding Using Extended Block Sizes," *ITU-T SG16/Q6, doc. C-123*, January 2009.
[67] Y. Vatis and J. Ostermann, "Prediction of P- and B-Frames Using a Two-dimensional Non-separable Adaptive Wiener Interpolation Filter for H.264/AVC," *ITU-T Q.6/SG16 VCEG, VCEG-AD08*, Hangzhou, China, October 2006.
[68] S. Wittmann and T. Wedi, "Simulation results with separable adaptive interpolation filter," *ITU-T Q.6/SG16 VCEG,AG10*, Shenzhen, China, October 2007.
[69] D. Rusanovskyy, K. Ugur, and J. Lainema, "Adaptive Interpolation with Directional Filters," *ITU-T Q.6/SG16 VCEG,AG21*, Shenzhen, China, October 2007.
[70] Y. Ye and M. Karczewicz, "Enhanced Adaptive Interpolation Filter," *ITU-T SG16/Q.6 Doc. T05-SG16-C-0464*, Geneva, Switzerland, April 2008.
[71] T. Arild Fuldseth, D. Rusanovskyy, K. Ugur, and J. Lainema, "Low Complexity Directional Interpolation Filter," *ITU-T Q.6/SG16 VCEG, VCEG-AI12*, Berlin, Germany, July 2008.
[72] M. Karczewicz, Y. Ye, and I. Chong, "Rate Distortion Optimized Quantization," *ITU-T Q.6/SG16 VCEG, VCEG-AH21*, Antalya, Turkey, January 2008.
[73] A. Tanizawa and T. Chujoh, "Adaptive Quantization Matrix Selection," *ITU-T Q.6/SG16 VCEG, D.266*, April 2006.
[74] M. Narroschke and H.G. Musmann, "Adaptive prediction error coding in spatial and frequency domain for H.264/AVC," *ITU-T Q.6/SG16 VCEG, VCEG-AB06*, Bangkok, Thailand, January 2006.
[75] T. Chujoh, N. Wada, and G. Yasuda, "Quadtree-based adaptive loop filter," *ITU-T Q.6/SG16 VCEG,C.181*, January 2009.
[76] G. Yasuda, N. Wada, T. Watanabe, and T. Yamakage, "Block-based Adaptive Loop Filter," *ITU-T Q.6/SG16 VCEG, VCEG-AI18*, Berlin, Germany, July 2008.
[77] T. Chujoh and R. Noda, "Internal bit depth increase for coding efficiency," *ITU-T Q.6/SG16 VCEG, VCEG-AE13*, Marrakech, MA, January 2007.
[78] S. Klomp and J. Ostermann, "Response to Call for Evidence in HVC: Decoder-side Motion Estimation for Improved Prediction," *ISO/IEC JTC 1/SC 29/WG 11 M16570*, London, UK, June 2009.

[79] T. Murakami, “Advanced B Skip Mode with Decoder-side Motion Estimation,” *ITU-T Q.6/SG16 VCEG, VCEG-AK12*, Yokohama, Japan, April 2009.
[80] S. Kamp, M. Evertz, and M. Wien, “Decoder Side Motion Vector Derivation,” *ITU-T Q.6/SG16 VCEG, VCEG-AG16*, Shenzhen, China, October 2007.

VITA

Mr. Li Liu was born in Feb. 10th, 1981, in Chenzhou City, Hunan Province, south China. He spent 6 years in Chenshi Elementary School from 1986 to 1992 and then attended Chenzhou 1st High School from 1992 to 1998, where he was ranked as the top three students in his grade.

In 1998, Mr. Liu was recommended to Tsinghua University (Beijing, China) without taking the University Entrance Exam as his excellence in high school study. He chose to study in the electronic engineering department and obtained his bachelor degree in 2002. At the same year, he was accepted into the PhD program of Tsinghua University. In 2004, he transferred to computer science department of University of Missouri as a PhD candidate and completed his PhD study in 2009.