



THE HONG KONG
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

The Hong Kong Polytechnic University
Department of Electronic and Information Engineering

Efficient Techniques for Video Rate Control

Lo Kam Fai

A thesis submitted in partial fulfillment of the requirements for
the Degree of Master of Philosophy

Feb 2005



Pao Yue-kong Library
PolyU · Hong Kong

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best to my knowledge and belief, it reproduces no material previously published or written nor material which has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

LO KAM FAI (Name of student)

Abstract

Since the first introduction of digital video compression, the development of the compression technology has never stopped. The increasing demand for digital video applications puts increasing pressure on various international standards such as MPEG-1, MPEG-2, MPEG-4, H.263 and H.264 to keep improving. Rate control is one of the most important responsibilities of a video encoder, especially for video transmission. The output rate of a video must be controlled according to the channel bandwidth and the buffer constraints; this can be achieved by adjusting the quantization parameters used to compress the video frames. These quantization parameters are used to control the trade-off between the video quality and the output bit rate. Although rate control is important, it is not defined in the international video standards since it does not affect the design of the decoders. Therefore, video applications must have their own control scheme. The reference models of the existing standards employ different rate-control schemes. In H.263+, the scheme is based on the classical logarithmic rate-distortion function. The rate model of the recent standards - MPEG-4 and H.264 - employs quadratic rate-distortion functions. These schemes rely on a statistical rate model derived from the information theory,

while some other schemes are based on fuzzy logic. Like the classical rate model, the fuzzy-logic rate controller needs the statistical data of the video in order to estimate suitable quantization parameters. However, the fuzzy-logic approach usually requires simpler implementation as compared to those based on the rate models. All these rate-control schemes are designed for real-time video coding, and their computations should be relatively low. Some rate-control schemes also consider rate-distortion optimizations, which need an iterative process to search for the optimal or sub-optimal combination of the quantization parameters for predictive coded frames. Such schemes always outperform the real-time control schemes in terms of rate and quality, but a larger amount of computations are required. Hence, these schemes are only suitable for non-real-time applications.

In this research, we explore the most popular rate-control schemes, including those adopted in the reference software of H.263+ (TMN8), MPEG-4 (VM5.0) and H.264 (JM8.1a), the fuzzy-logic rate controller, and the rate-distortion optimization schemes. We also propose a new rate-control scheme based on a distortion-rate model, which can be extended easily to a higher order for different coding algorithms without requiring much effort to determine the quantization scales. The formulation of the cubic model is derived from the quadratic rate-distortion model and from simple curve fitting. Our model is implemented in H.264 JM8.1a, and its

performance is compared to JM8.1a with different test sequences and bit rates.

Experimental results show that our proposed model outperforms the quadratic model in terms of both PSNR and bit rate.

The second part of the research aims at reducing the computations of an existing rate-distortion optimization scheme. The computational effort is reduced to half with only a slight decrease in video quality. The cubic model scheme is implemented in the Joint Model 8.1a (JM8.1a) of H.264, so that the performance can be evaluated. Experimental results show that our scheme can achieve better performance levels than JM8.1a. A study of the typical, basic hybrid coders and the new features of H.264 is also conducted in this research.

Publications

Accepted

Kam-Fai Lo, Kin-Man Lam, “A New Cubic Rate-Distortion Model for Low-Delay Video Communication”, IEEE International Symposium on Circuits and Systems, 23 – 26 May 2005, Kobe, Japan, Pages: 4586-4589

Submitted

Kam-Fai Lo, Kin-Man Lam, “Rate Control for Low-Delay Video Communication Using Cubic Model”, IEEE Transactions on Circuits and Systems for Video Technology.

Acknowledgements

I would like to appreciate my supervisor, Dr Kenneth Kin Man Lam deeply for his valuable advice on my research project. His guidance and full support are vital, especially on some parts of this thesis where the topic is not his main research area. I would also like to appreciate my co-supervisor Prof. Wan-Chi Siu for his kindly encouragement on my work.

I am grateful to Dr Yui-Lam Chan's research team for their interesting discussion on H.264. I would like to thank to my teammates Dr Wai-Pak Choi, Mr Kwok-Wai Wong, Mr. Cheung-Ming Lai and Mr. Kin-Wai Sze. I must also thank to my colleagues Mr. Wai-Kong Cheuk, Mr Kai-Yuen Cheong, Ms. Kai-Man Au, Mr. Tommy Ho-Chuen Lau and Mr. Kai-Wing Tse for their help on technical problems.

I would like to thank to the Center for Multimedia Signal Processing and Hong Kong Polytechnic University for the excellent research environment and Research Grants Council of the HKSAR, China for the kindly support on my work (Project No. PolyU 5123/01E).

Finally, I am profoundly grateful to my family with deepest love.

Table of Contents

Chapter 1	Introduction	1
1.1	A brief history of video coding	1
1.2	Hybrid video coder.....	2
1.3	Problem of video transmission.....	4
1.4	The problems of existing rate control techniques and the aim of this research project	7
1.5	A brief introduction of H.264.....	9
1.6	Summary	14
Chapter 2	Rate Control Techniques.....	16
2.1	Introduction	16
2.1.1	Rate-distortion	16
2.1.2	Operational Rate-Distortion Theory (ORDT).....	20
2.1.3	Rate-distortion curve approximation by modeling.....	22
2.1.4	Lagrangian multiplier	24
2.1.5	Typical rate control method	31
2.1.6	Rate control by prediction	32
2.1.7	Rate control by pre-analysis	32
2.2	Rate control in the Verification Model of MPEG-4.....	33
2.3	Rate control in the Joint Model of H.264.....	35
2.3.1	Group of picture (GOP) level.....	35
2.3.2	Frame level.....	38
2.3.3	Basic unit level	42

2.4	Fuzzy based rate control.....	43
2.5	Other rate control techniques	45
Chapter 3	Rate Control Based on Cubic Rate Model	46
3.1	The rate control model	47
3.1.1	The cubic rate model	47
3.1.2	The model parameters	52
3.2	Rate Control	64
3.2.1	Buffer control	64
3.2.2	Frame level rate control.....	65
3.2.3	Basic unit level rate control.....	68
3.3	Experimental results	69
3.4	Computation Analysis	76
3.5	Conclusion.....	77
Chapter 4	Simplified Pre-analysis Rate Control	80
4.1	The Problem of Dependency	80
4.1.1	Frame-wise, block-wise and object-wise optimization	83
4.1.2	Time consumption of rate control Schemes	84
4.2	Approximation of operational rate control scheme.....	85
4.2.1	Intra-frame approximation	85
4.2.2	Inter-frame dependency model.....	86
4.3	Fast approximation and modeling	89
4.3.1	Fast Intra-frame approximation.....	89
4.3.2	Fast Inter-frame dependency model	92
4.4	Conclusion.....	98
Chapter 5	Conclusion.....	100

5.1	Conclusion on the research project	100
5.2	Future development of rate-control techniques.....	103

List of Figures

Figure 1-1 Typical hybrid video coder.....	4
Figure 1-2 The close loop rate control path	7
Figure 2-1 A typical quantization matrix in MPEG-4.....	18
Figure 2-2 The quantization matrix in H.264.....	18
Figure 2-3 Typical distortion and rate curves	20
Figure 2-4 A rate-distortion curve.....	20
Figure 2-5 Operation rate distortion function.....	21
Figure 2-6 The solution of the Lagrangian optimization.	28
Figure 2-7 The Lagrangian solution.....	29
Figure 2-8	30
Figure 2-9 Typical fuzzy logic controller.....	45
Figure 3-1 The term $1/D$ in the R-D model.....	52
Figure 3-2 The term $1/R$ in the R-D model	52
Figure 3-3 The effect of the previous quantization parameter on the R-Q curve of the current frame based on the video “Grandma”.....	54
Figure 3-4 The relationship between MAD_{k-1} and a_i ’s of the current frame for the sequence Mobile.....	56
Figure 3-5 The relationship between MAD_{k-1} and a_i ’s of the current frame for the sequence Grandma.	57
Figure 3-6 The relationship between MAD_{k-1} and a_i ’s of the current frame for the sequence Hall.	58
Figure 3-7 The relationship between MAD_{k-1} and a_i ’s of the current frame for the sequence Salesman.	59
Figure 3-8 The relationship between MAD_{k-1} and a_i ’s of the current frame for the sequence Akiyo.	60
Figure 3-9 The rate threshold of the model based on Grandma.....	61
Figure 3-10 The selected quantization parameters for the first 200 frames of the video Grandma with target bit rate 64kbps.	61
Figure 3-11 The smoothed quantization parameters selected from the successive frames of Grandma at target bit rate 64kbps.	62
Figure 3-12 The rate control flow in the encoder side.	68

Figure 3-13 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video Mobile.	71
Figure 3-14 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video News.....	72
Figure 3-15 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video Akiyo.....	73
Figure 3-16 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video Grandma.	74
Figure 3-17 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video Hall.	75
Figure 3-18 The visual quality using the JM8.1a rate control scheme (left side) and using our proposed scheme (right side) based on the tested sequences (a) News, (b) Akiyo and (c) Grandma.	79
Figure 4-1 The distortion curves of a P-frame against the scaling factors of the corresponding I-frame.	83
Figure 4-2 The 3-D plane of the R-D curve.	87
Figure 4-3 The 2-D plane showing the control points used to approximate the whole dependency 3-D curve.....	88
Figure 4-4 The distortion curve of the sequence Football.....	91
Figure 4-5 The distortion curve of the sequence Mobile.	91
Figure 4-6 The distortion curve of the sequence Garden.	92
Figure 4-7 The dependency R-D curve.	93
Figure 4-8 The dependency R-D curve showing the points needed to be approximated with circles.	96
Figure 4-9 Projection from the QP_I axis.	97
Figure 4-10 The $Distortion_P-QP_I$ curve and the $Distortion_P-QP_P$ curve obtained from different points of view of the 3-D R-D dependency curve.	98

List of Tables

Table 3-1 Average relative errors of the models based on different video sequences.	51
Table 3-2 The performance on target rate and PSNR of our scheme as compared to JM8.1a. M=mobile, N=news, G=grandma and H=hall; the number that follows is the target rate (kbps).....	70
Table 3-3 Average number of multiplication and addition of RD2 and DR3	77
Table 3-4 Time consumption	77
Table 4-1 Relative modeling error of our distortion modeling of the I-frame.	90
Table 4-2 Relative modeling error of our distortion modeling scheme for P-frames.	96

Chapter 1

Introduction

1.1 A brief history of video coding

In the beginning of development of video coding, the applications were limited to wide bandwidths and low-quality picture phones. The video standard H.120 [1] was issued by CCITT (formerly ITU-T) in 1984, and is based on differential pulse code modulation (DPCM) [2] and scalar quantization. The standard introduced frame difference coding in later version in order to reduce temporal redundancy between frames. The initial prototype of nowadays hybrid video codec was formed as the standardization of H.261 [3] which includes integer-pel accurate motion compensation. However, compressed digital video was still unpopular until the MPEG-1 standard [4, 5] was introduced, which is the most well-known standard aimed at compressing a video to a 1.5Mbps bit stream and fitting onto compact disc for entertainment purpose. MPEG-1 made digital compressed video popular. In 1994, the MPEG-2 standard (H.262) [6, 7] was issued, which improves the quality of video compression in terms of bit rate and visual quality. The Digital Versatile Disc (DVD),

which adopted MPEG-2, is the medium for entertainment applications. In order to support applications such as video broadcasting and HDTV with bit rates ranging from 20Mbps to 40Mbps, the MPEG group initiated the MPEG-3 standard. However, MPEG-3 was discontinued, as such high bit-rate video communication can be supported by enhancing the initial version of MPEG-2. In 1996, H.263 [8] was launched to support the variable block-size motion compensation. The MPEG-4 standard [9] was issued in 1999, as was standardized object-based video coding. In 2003, the H.264 standard [10-12] was issued and adopted as the part 10 of MPEG-4.

1.2 Hybrid video coder

A typical current video coder contains transform, quantization, motion compensation and entropy coding. The term ‘hybrid coder’ means a combination of different compression algorithms to exploit the spatial and temporal correlation. Spatial correlation is exploited by a domain transform [13], which compacts most of the energy in a small number of transform coefficients. Temporal correlation is exploited by means of motion compensation, which searches for the corresponding best-matched macroblock in a reference frame. The displacement between the macroblocks in the current frame and the reference frame is referred to as a motion

vector. The motion vector requires a much smaller number of bits as compared to the size of the encoded bits of the whole macroblock.

In various video applications, the input raw video is converted into different formats depending on the nature of the applications. The most popular format separates a color into its Y, U and V components, where Y is the luminance component while U and V represent the chrominance components. Since the human vision is more sensitive to the luminance than to the chrominance signals, the sampling rate of the Y component is normally higher than those of the U and V components. In most of the existing video standards, the luminance and chrominance components are coded separately with different coding parameters. The raw video frames are inputted into the domain transform as shown in Figure 1-1. The most popular transform for video coding is the Discrete Cosine Transform (DCT) but it is not the best transform in terms of coding efficiency. In recent standard H.264, a DCT-like transform namely, Integer Cosine Transform (ICT) [14-16], is employed, as will be described in Section 1.5. The transform is followed by a quantization process. The quantized values are converted into a sequence of symbols, which are then further compressed using a lossless technique. Quantization is a lossy process in the encoding. The loss of information due to quantization cannot be recovered. In decoding, the quantized coefficients are dequantized and then converted to the spatial

domain using the corresponding inverse transform. The reconstruction process in the encoder is the same as that in the decoder, and hence this process is referred to as the decoding loop of the encoder. The reconstructed macroblock is then used for motion estimation by searching a macroblock in the reference frame with the smallest error. This error, known as the residual error, is coded instead of the original macroblock. The last step is the entropy coding of residual errors and motion vectors.

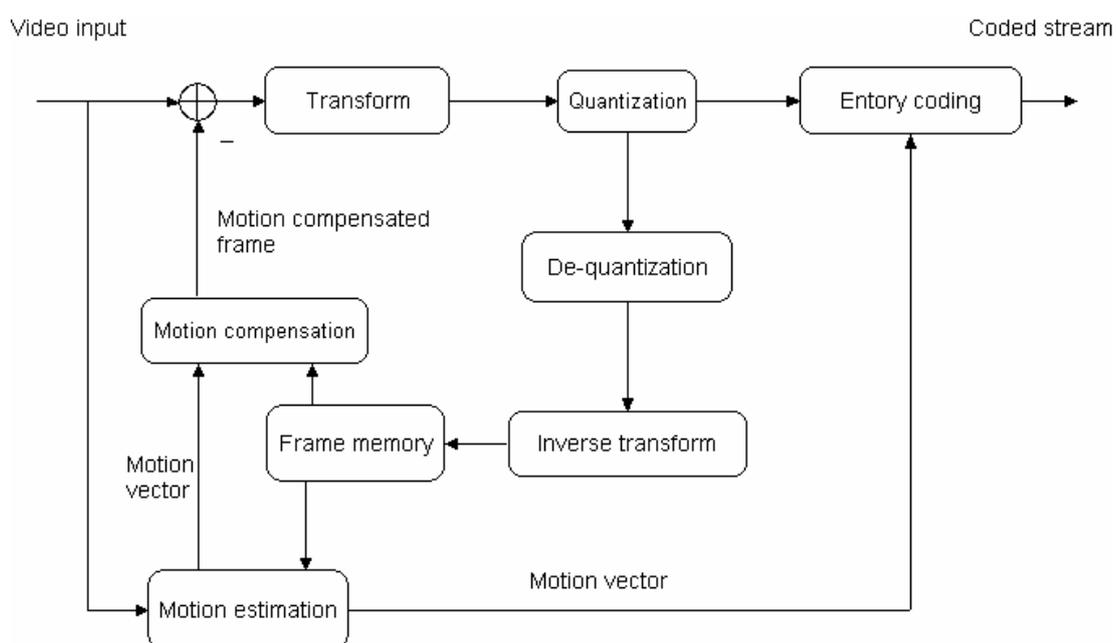


Figure 1-1 Typical hybrid video coder

1.3 Problem of video transmission

The bit-rate of a raw video sequence is linearly proportional to the number of precision levels for the pixels, the ratios of spatial sampling in the luma and chroma domains, and the resolution. It is trivial to predict the size of a raw video without

knowing the exact content and the statistics of the samples. However, digital compression makes the story different, as domain transform, quantization and entropy coding are involved in the encoder. Domain transform compacts the energy of a number of correlated coefficients into a few coefficients, and the compacted coefficients are quantized into levels with different possible precisions. The bit-rate of the quantized coefficients depends on the distribution of the coefficients. Furthermore, the quantized coefficients are entropy-coded, whereby frequent streams of bits are represented by symbols of shorter lengths, while less frequent streams are replaced by symbols of longer lengths in order to compress the coefficients effectively. Therefore, the output bit-rate of a typical hybrid video coder is a source with variable bit-rate (VBR). Unfortunately, most of the existing applications rely on constant-bit-rate channels (CBR) such as video playback, and video streaming through CBR channels. This makes rate control necessary before transmitting the video streams into the media.

Rate control is the responsibility of the video encoder, and hence it will not be included in the video standards, which specify only the agreement of the playback algorithms and the syntax of the transmitted bit-stream. As the bit-rate somehow depends on the number of frames and the resolution, even though the final encoded rate is still unpredictable, some existing algorithms are used to control the rate by

varying the frame rate [17-20] or the resolution [21]. However, the rate control adopted in the reference software of existing standards controls the peak-signal-to-noise ratio (PSNR) [22] rather than varying the frame rate or the resolution in order to control the trade-off between quality and output bit-rate. This control can be achieved by controlling the quantization step size, which yields different levels of precision of the transformed coefficients [23].

Although the coded bit-rate can be controlled by varying the frame rate, resolution or quantization step, some fluctuations in bit-rate still exist. A buffer is generally employed at the output of the encoder so as to smooth the bit-rate fluctuations that a CBR channel or media transmission bus cannot digest. Hence a typical design of current rate control algorithms contains a close-loop control and an output buffer, as shown in Figure 1-2. The controller can employ any one of the three above-mentioned rate control schemes. The control scheme can be based on a statistical model (to be described in Section 2.1.3), the fuzzy logic controller (to be described in Section 2.4), as well as the non-real-time optimization controller (to be presented in Section 2.1.2).

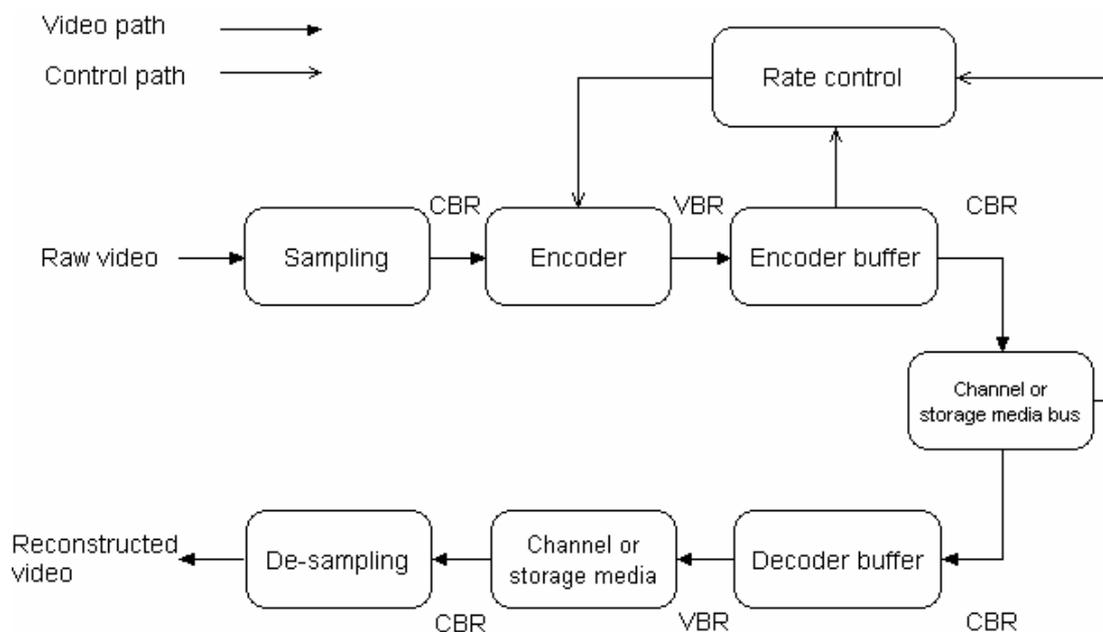


Figure 1-2 The close loop rate control path

Besides being used for CBR channel transmission, video applications with variable bit-rate (VBR) output aim to maintain the perceptual quality of a video constant, and allow the bit-rate to fluctuate. The characteristic of ATM is suitable for VBR transmission [24].

1.4 The problems of existing rate control techniques and the aim of this research project

In the existing work on predictive rate control, the rate model is usually described by means of the logarithmic or quadratic statistical models. The performances of such models are very reliable for different video coding standards. Although the models are derived through statistical derivation with a reasonable theoretical background, they are unable to model all rate-distortion curves effectively.

For a quadratic model, its parameters must be solved in order to determine the corresponding quantization step size for a specific bit-rate. The order of the model is usually limited to two, as the process of solving a higher-order model requires many more computations. Therefore, in this thesis, one of our objectives is to develop a reliable rate-distortion model that allows us to employ a higher-order model with only a slight increase in computational complexity. As the computational power of processors, especially the MCUs of embedded systems, is increasing continuously, the slightly increased complexity is not significant. Another problem that makes the development of a new rate-distortion model difficult is the requirement that the parameters of a new model can be determined in a simple way, and probably have a linear relationship with the statistics of the previous coded frames or macroblocks. A model that can fit the actual rate-distortion curve is unreliable if it requires a difficult algorithm to determine its parameters. The most popular technique to determine such model parameters is linear regression. Hence, we also employ this technique in the calculation of the parameters of the model that we propose in this thesis.

The other main stream of rate-control algorithms is based on pre-analysis. The rate-distortion model is determined by coding the same frame or macroblock with several quantization steps in order to obtain the corresponding bit-rates and

distortions. This kind of technique can be employed by non-real-time applications in which time is not a tight constraint, as the pre-analysis process involves a huge number of computations. Examples of these applications include video broadcasting and stored video for media with a tight size. However, the demand on cable TV is increasing. The encoding process of the broadcasting media has been becoming frequent. Some movie channels encode certain number of new movies for broadcasting every season. Decreasing the complexity of pre-analysis rate control technique is therefore useful. Some existing algorithms reduce the computational complexity in such a way that a sub-optimal combination of the quantization steps is found. However, such algorithms still require a significant amount of computation. Another aim of this research is therefore to develop a fast approximation model for an existing pre-analysis rate-control algorithm. This fast model requires less computation but with a slight decrease in accuracy for bit-rate control.

1.5 A brief introduction of H.264

In early 1998, the Video Coding Expert Group (VCEG) called for proposals for a project called H.26L, targeted at halving the output bit-rate of the existing video coding standards, with a given quality requirement. In late 2001, VCEG and Moving Picture Experts Group (MPEG) formed the Joint Video Team (JVT) to work on the

project specifically. The project was approved as a joint standard H.264 and as Part 10 of MPEG-4 14496-10 in early 2003. Its common name is Advanced Video Coding (AVC). The details of the development history of H.264 are given in [25-27]. The highlighted differences between H.264 and other existing video standards are described in [28].

In H.264, a number of new coding techniques are employed. Also, some existing techniques in the previous standards are enhanced and adopted as part of the standard in order to make the codec meet the target performance. Following are some of the important techniques employed in the standard.

Intra spatial prediction

The pixels of the current macroblocks are intrapolated by their neighboring decoded macroblocks in the current frame. This technique can further compact the energy of the pixels in the macroblocks.

Context-adaptive entropy coding

Two powerful entropy-coding techniques are employed in H.264 [29]. For the baseline profile, context-adaptive variable-length coding (CAVLC) is employed, while context-adaptive binary arithmetic coding (CABAC) is used in the higher profiles. Both techniques are based on the context adaptive technique, in which the

symbols vary depending on the context in order to further compact the information contained in the bit-stream than was possible with the coding techniques adopted in those previous standards.

4 × 4 transform block

The transform size in H.264 is set at 4×4 [14, 15] rather than 8×8 in the previous standards. A smaller transform size can compact the energy of the block in a more local manner. This benefits video sequences with finer and local texture information.

Variable block size

In previous standards, the block size for motion compensation is fixed to 16×16 . In the classic hybrid codec, the transform size is not equal to the block size for motion compensation. This makes the blocky effect occur very often and more easily. H.264 allows the motion-compensation block size to be equal to the transform block size [30, 31] in some profiles.

Integer cosine transform

In the classic Discrete Cosine Transform (DCT), the forward transform and the inverse transform are not necessary to reconstruct the original coefficients with the same precision since the decoder may have a different precision definition to the

encoder. This problem results in a mismatch between the forward transform and the inverse transform, and the quality of the decoded sequence cannot achieve the expected performance level. The kernel of the Integer Cosine Transform (ICT) is based on integers, and therefore the transformed coefficients are also integers [14-16]. Hence, the forward transform and the inverse transform match each other.

The cyclic quantization

In order to implement the ICT with less computation, some values of the kernel are factorized to a scale matrix and combined with the quantization matrix. On the other hand, the quantization table is actually repeated for every increment of 6 in the quantization parameters (QP) [14, 15, 32]. Cyclic quantization is used to reduce the size of the quantization table.

Deblocking filter

Prediction and residual coding always result in blocky artifacts that can be iteratively propagated. An in-loop deblocking filter is included in the prediction as to how H.264 will improve the quality of the inter-coded units and the later related units [33].

Improved skipped and direct mode

Skipped macroblocks in the previous standards do not allow motion compensation. However, the performance of the skipped mode is not robust if the skipped regions actually contain a global motion. Therefore, H.264 allows skipped MBs to have their motion vectors. In addition, the direct mode is an enhancement of MPEG-4 and H.263+ while it is a basic feature in the profiles.

Quarter-pel motion compensation

The precision of motion vectors in H.264 is as fine as quarter pel, which is one feature in the MPEG-4 advanced profile. The quarter-pel motion compensation in H.264 is improved to have less computation.

Multiple frame reference

In the previous standards, the number of reference frames for motion estimation is restricted. However, H.264 allows the motion compensation to be based on at most 32 reference frames.

Flexible macroblock ordering (FMO)

The macroblocks are grouped as independent decodable flexible-size slice groups. The ordering of the slices is highly flexible to enhance the robustness against data loss by managing the spatial location of the slice groups.

SI/SP frames

This is an option that allows for a smooth change between sequences with different target bit-rates merely by inserting a relatively small-size frame as a switching frame. Such a switching frame is not necessarily intra-mode. The switching frame can also be a predicted frame in order to allow flexible switching in case the switching point is between two intra frames [34-37].

Data partitioning

This feature separates the more important information from the less important by placing them into different packets. This allows applications to provide different levels of error protection for packets of differing importance [38].

1.6 Summary

In this chapter, we have presented a brief introduction to video coding and the importance of the video transmission algorithms. The structure of a video codec and the basic ideas for bit-rate control have also been described. As the major work of this thesis, we have also addressed the major problems of the existing rate-control algorithms. H.264 is the latest video standard, and the rate-control algorithms proposed in this thesis will be compared to the one in the reference model. Hence, a brief introduction to the new features in H.264 has also been given. A brief

introduction to the organization of this thesis is also provided. In the following chapters, we will give a detailed review of existing rate-control algorithms and the rate-control algorithms proposed in this thesis.

Chapter 2

Rate Control Techniques

2.1 Introduction

As introduced in Section 1.3, there are many different rate control approaches. The control schemes can be categorized into two main groups, prediction and pre-analysis. The former is to predict the quantization parameter based on statistical characteristic of the video sequence while the latter group decides the quantization parameter (QP) based on iterative analysis and coding the sequence actually with different combination of QP and probably with different choice of coding modes.

2.1.1 Rate-distortion

Various standards have different quantization processes. In MPEG-4, the quantization of the DC coefficients and AC coefficients are very different. DC quantization is divided by a scaling factor which value depends on the quantization parameter (QP) while the AC coefficients are divided by a quantization matrix before the scaling factor [14, 15, 32]. The quantization matrix can be a flat matrix or defined

by the encoder. The encoder must include the quantization matrix in the encoded video bit-stream in case a non-flat quantization matrix is used. The pattern of the quantization matrix reveals that low-frequency coefficients are more important. A general pattern is shown in Figure 2-1 in which the region of low-frequency coefficients are assigned lower quantization steps to reserve higher perception because human visual system (HVS) is believed to be more sensitive on low-frequency signal. On the other hand, the scaling factor is to control the overall effect of the quantization matrix preserving the relative coarseness between coefficients in the quantization matrix. The scaling factor is non-linearly proportional to the quantization parameters (QP). The relationship between QP and scaling factor in these standards are fixed. This quantization parameter has different names in different standards, for example, it is known as *MQQUANT*, *Q_level* and *QP* in MPEG-1, MPEG-2, and MPEG-4, respectively. We will use the term, *QP*, whenever we discuss the quantization parameter.

16	17	18	19	20	21	22	23
17	18	19	20	21	22	23	24
18	19	20	21	22	23	24	25
19	20	21	22	23	24	26	27
20	21	22	23	25	26	27	28
21	22	23	24	26	27	28	30
22	23	24	26	27	28	30	31
23	24	25	27	28	30	31	33

Figure 2-1 A typical quantization matrix in MPEG-4

However, the quantization process of H.264 as mentioned in Section 1.5 is very different from those of other standards [14, 15, 32]. The weighting matrix of the ICT is combined with the quantization matrix. Hence the pattern of the matrix is fixed as shown in Figure 2-2 to preserve the pattern of the factor of the ICT kernel. Similar to previous standards, the quantization step (QStep) is non-linearly proportional to the QP in H.264 and QP is used to control the trade-off between output rate and perceptual quality.

P_1	P_3	P_1	P_3
P_3	P_2	P_3	P_2
P_1	P_3	P_1	P_3
P_3	P_2	P_3	P_2

Figure 2-2 The quantization matrix in H.264

The rate control schemes of the reference models in existing coding standards control the output bit rate and the output buffer fullness by choosing different QP .

Different QP will result in different bit rate and distortion. Generally, bit rate is

inversely proportional to QP , while distortion is proportional to QP . Figure 2-3 shows a typical example for the rate curve and the distortion curve. It is also possible to plot the rate and distortion together to form the so-called rate-distortion curve and quantization function (QF). The curve may either be a plot of the rate against the distortion, or the distortion against the rate. This only affects the notation of a rate control algorithm, but has no impact on the algorithm itself. Therefore, it should be made clear in this section. Figure 2-4 shows an example of an R-D plot. There are many different ways to represent the distortion. Subjective visual/perceptual distortion and statistical error measurement [22] are two major approaches to measure the distortion. Visual distortion measurement is much more meaningful against error measurement however, most of the existing rate control schemes use error measurement as subjective measurement always include a complicated human visual system model. Error measurement includes sum of square differences (SSD), mean squared error (MSE), and peak signal-to-noise ratio (PSNR). They are defined as follows:

$$SSD = \sum |F - G|^2 \quad (2.1)$$

$$MSE = \frac{1}{N} \sum |F - G|^2 \quad (2.2)$$

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \quad (2.3)$$

where F is the original frame, G is the decoded frame and N is the number of pixels in F .

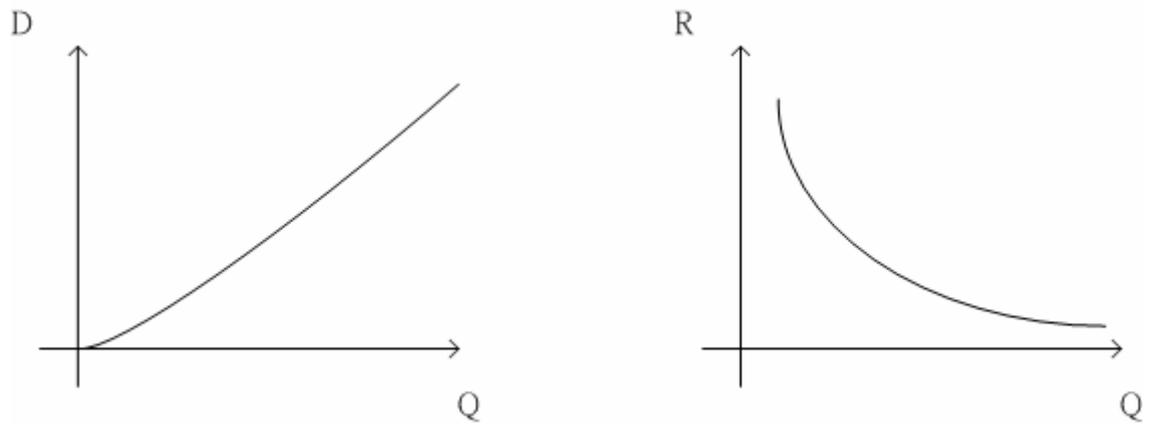


Figure 2-3 Typical distortion and rate curves

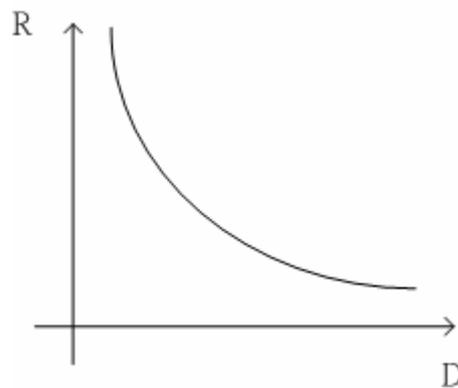


Figure 2-4 A rate-distortion curve

2.1.2 Operational Rate-Distortion Theory (ORDT)

The rate-distortion curve can be obtained by coding a frame repeatedly with different quantization parameters. Let $Q = \{q_1, q_2, \dots, q_m\}$ be the set of all m possible

quantizers, $R(q_i)$ be the rate resulted from quantizer q_i , and $D(q_i)$ be the corresponding distortion. Then, the quantization function can be defined as follows:

$$QF = \{(R(q_i), D(q_i))\}_{i=1}^m \quad (2.4)$$

Then, the operational rate distortion function can be defined as follows:

$$Q_{ORDF} = \{q : q \in Q, R(q) \geq R(p) \Rightarrow D(q) < D(p), \forall p \in Q\} \quad (2.5)$$

This definition means that, for a quantizer based on ORDT, there is no other quantizer that can result lower rate and lower distortion. Note that this does not mean Q_{ORDF} lying on the convex hull of the sets quantization function (QF). Figure 2-5 shows an example in which $m = 31$ (i.e. the crossed points) and only 10 (i.e. the circled points) points are used to form the Q_{ORDF} curve (i.e. the solid line).

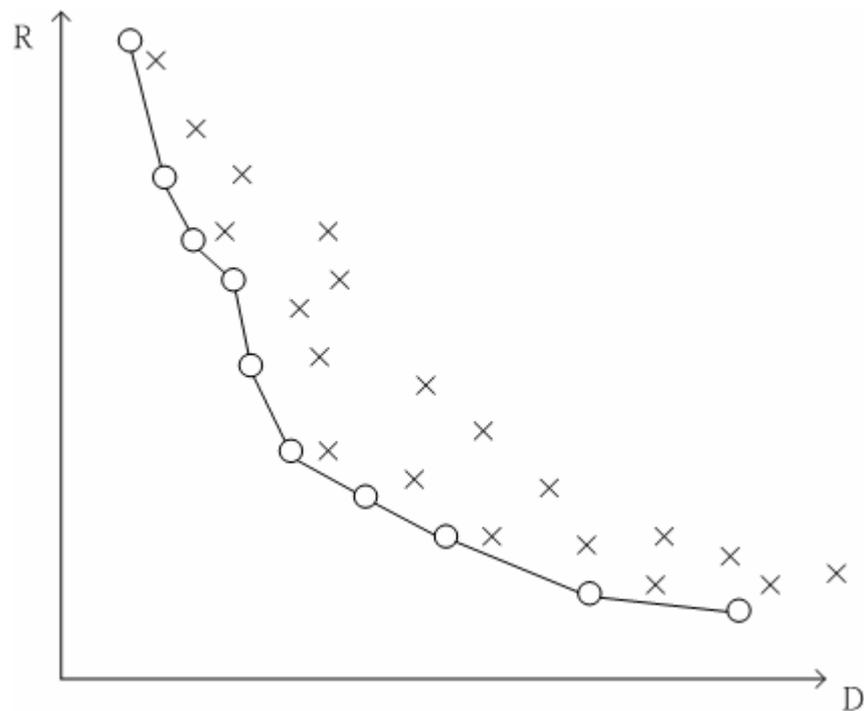


Figure 2-5 Operation rate distortion function

2.1.3 Rate-distortion curve approximation by modeling

Although operational rate distortion function can reveal the best R-D relationship for different QP , it requires significant computations for measuring the rate-distortion points as all possible values of QP are used to code a sequence. Therefore, approximation schemes are used to model the R-D curves. In modeling the R-D curves, the information source is modeled as a specific discrete-time memoryless source [39, 40]. Based on the source model, the rate and distortion relationship is derived to a closed form.

Gaussian Source

Discrete memoryless Gaussian source is the most common source model in communication. The characteristics of communication model are quite similar to the situation of video coding in which the distortion due to the quantization process is similar to the channel error in communication applications. If the source outputs are independent Gaussian random variable with zero mean and variance σ^2 , and the squared-error distortion measure is used, the distortion function can be represented as follows,

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad (2.6)$$

$$R(D) = \frac{1}{2} \ln \frac{\sigma^2}{D} \quad (2.7)$$

where x and μ is the actual value and mean value of the source and D is the distortion within the range $1 \leq D \leq \sigma^2$. By mapping Q and D , the rate-distortion model can be rewritten as,

$$R(Q) = u_1 + u_2 \log \frac{1}{Q} \quad (2.8)$$

where u_1 and u_2 are the model parameters function by empirical results or statistical analysis. The proof of the equation can be found on [39].

Laplacian Source

The distribution of the DCT coefficients is modeled as a Laplacian source in [41-43]. However, it can also be a special case of Gaussian distribution [44]. The closed form of the Laplacian distribution [45] is given as follows,

$$P(x) = \frac{\alpha}{2} e^{-\alpha|x|} \quad (2.9)$$

$$R(D) = \ln \left(\frac{1}{\alpha D} \right) \quad (2.10)$$

where D is the distortion given by $D(x, \bar{x}) = |x - \bar{x}|$ with $0 < D < 1/\alpha$, R is the rate and x and \bar{x} are the actual value and quantized value of the source, respectively.

Exponent Function

A model function based on exponent relationship was proposed in [46-49].

The definition is given as follows,

$$R(Q) = v_1 + \frac{v_2}{Q^g} \quad (2.11)$$

where $1 \leq \mathcal{G} \leq 2$. The performance of this model is accurate than that of the Gaussian source model as it uses one more parameter \mathcal{G} . The model is fitted into the actual R-D curve by calculating the corresponding values at the 2 points: $Q = 1$ and $Q = 31$.

The accuracy of a model depends on how good the model function can fit into the actual situation. The model parameters can be preset, but a global setting cannot fit into all sequences. If the parameters are set adaptively, the statistical analysis requires additional computational effort. Further models can be found in [50] and [51] for TM5 [52].

2.1.4 Lagrangian multiplier

Lagrangian multiplier was first proposed by Everett [53] to solve constrained continuous optimization problems. It was first used in [54] for discrete problem of optimal bit allocation and the selection of quantizers. A clear review and explanation on Lagrangian optimization can be found in [22], [55] and [56]. The following provides some important theorems for Lagrangian optimization.

2.1.4.1 Lagrangian Optimization Theorem

Theorem 1: Let S_Q be a finite set, $Q \in S_Q$ be a member of the set, let $R(Q)$ and $D(Q)$ be real valued functions defined over S_Q . Then $\forall \lambda \geq 0$, the optimal solution $Q^*(\lambda)$ of the unconstrained problem,

$$\min_{Q \in S_Q} (D(Q) + \lambda * R(Q)) \quad (2.12)$$

is also the optimal solution to the constrained problem,

$$\min_{Q \in S_Q} D(Q) \quad (2.13)$$

subject to:

$$R(Q) \leq R(Q^*(\lambda)) \quad (2.14)$$

Equation (2.13) is so-called the objective function and (2.14) is the constraint.

Theorem 2:

$$\text{If } R(Q^*(\lambda_1)) > R(Q^*(\lambda_2)), \text{ then } \lambda_2 \geq -\frac{D(Q^*(\lambda_1)) - D(Q^*(\lambda_2))}{R(Q^*(\lambda_1)) - R(Q^*(\lambda_2))} \geq \lambda_1 \quad (2.15)$$

Theorem 3: $R(Q^*(\lambda))$ is a monotonic decreasing function of the Lagrangian multiplier λ . The proof of these three theorems can be found in [56].

2.1.4.2 Search for the Lagrangian multiplier λ

Since $R(Q^*(\lambda))$ is a non-increasing function of the Lagrangian multiplier λ , so the bisection method is normally used for solving the optimal Lagrangian multiplier λ . Suppose that the two initial points of λ are denoted as λ_H and λ_L , such that $\lambda_H > \lambda_L$. Therefore $R(Q^*(\lambda_L)) \geq R_{max} \geq R(Q^*(\lambda_H))$ where R_{max} is the maximum number of bits available for the channel or the buffer. As the maximum number of bits R_{max} lies between the two initial guesses, it is obvious that the solution of the optimization problem must lie between λ_H and λ_L . Then set $\lambda_M = \frac{\lambda_H + \lambda_L}{2}$,

and examine whether $R(Q^*(\lambda_L)) \geq R_{\max} \geq R(Q^*(\lambda_M))$ or $R(Q^*(\lambda_M)) \geq R_{\max} \geq R(Q^*(\lambda_H))$. If the former is true, the solution must lie between λ_M and λ_L , otherwise the solution will be between λ_H and λ_M . Then, the range of the solution is further divided, and the process is repeated until a rate that is very close to R_{\max} is achieved. Note that this is a discrete problem, therefore an exact solution may not be found. If the solution λ satisfies $R_{\max} + e \geq R(Q^*(\lambda)) \geq R_{\max} - e$, where e is a threshold, then the iteration can be terminated.

2.1.4.3 Other methods to find the Lagrangian Multiplier λ

In the previous section, the R-D points are calculated by selecting the scaling factor QP of all blocks in a frame. In other words, a frame is a coding unit. In this section, the coding unit is generalized to be block-wise. Consider that we have N coding units where each coding unit has M different possible quantizers. Let r_{ij} and d_{ij} be the rate and the distortion, respectively, of the i th coding unit with quantizer j . A higher j results in a higher distortion level while a lower j can preserve the visual quality of the coding unit. Note that the selection of the quantizer for a particular coding unit does not affect the selection of the quantizer for other coding units. Quantizer j can be further notated by $x(i)$, where x is a mapping function depending on the choice of the quantizers. For instance, there is four coding units and three

possible quantizers say $(q_1, q_2$ and $q_3)$ and the quantizers used for the four coding units are $q_2, q_2, q_3,$ and $q_1,$ respectively. Then $x(1) = 2, x(2) = 2, x(3) = 3,$ and $x(4) = 1.$ Thus, x depends on the choice of the combination of the quantizers. The budget constraint can be formulated by,

$$\sum_{i=1}^N r_{ix(i)} \leq R_{\max} \quad (2.16)$$

and the mapping $x(i)$ is the optimal solution to the budget constrained problem if it minimizes the following formulation,

$$\sum_{i=1}^N d_{ix(i)} + \lambda r_{ix(i)} \quad (2.17)$$

We denote the solution as $(r_{ix'(i)}, d_{ix'(i)}),$ which satisfies,

$$\sum_{i=1}^N r_{ix'(i)} \leq R_{\max} \quad (2.18)$$

and,

$$\sum_{i=1}^N d_{ix'(i)} \leq \sum_{i=1}^N d_{ix(i)} \quad (2.19)$$

The constrained optimization problem can be solved by the following unconstrained optimization problem,

$$\min \left(\sum_{i=1}^N d_{ix(i)} + \lambda r_{ix(i)} \right) = \sum_{i=1}^N \min(d_{ix(i)} + \lambda r_{ix(i)}) \quad (2.20)$$

Therefore, we can minimize the Lagrangian cost of the individual coding unit and finally accumulate the total cost instead of plotting all the R-D points for all the possible combinations. If we employ the above example and plot all the R-D points for all possible combinations, then we need $3^4 = 81$ points and have to code the

sequence 81 times. The most concerned disadvantage of Lagrangian multiplier for optimal rate distortion control is that it may not reach some points which are practically better than those achieved by the Lagrangian multiplier. Figure 2-6 shows a case in which the Lagrangian optimization cannot reach all circled points resulted from operational rate-distortion curve. The Lagrangian optimization can only reach those points lying on the convex hull. Thus, the solution resulted from Lagrangian optimization is always referred as convex-hull solution.

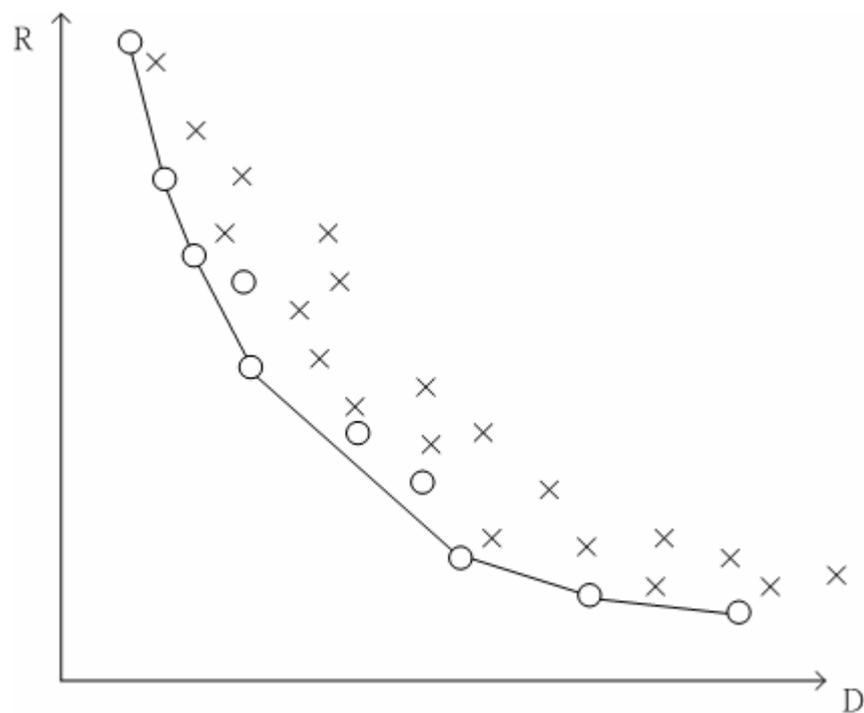


Figure 2-6 The solution of the Lagrangian optimization.

Obviously, the unreachable points, in some situations, may be much appropriate than the convex-hull solution resulted from the Lagrangian optimization.

Consider the case as shown in Figure 2-7, the optimal solution should be the R-D

point nearest to the maximum budget. However, as (2.12) is a line moving upward from the origin of the graph until it touches the first circle point, it can never reach those points located higher than the convex hull. In the next section, a block-wise optimization technique proposed in [57] will be discussed.

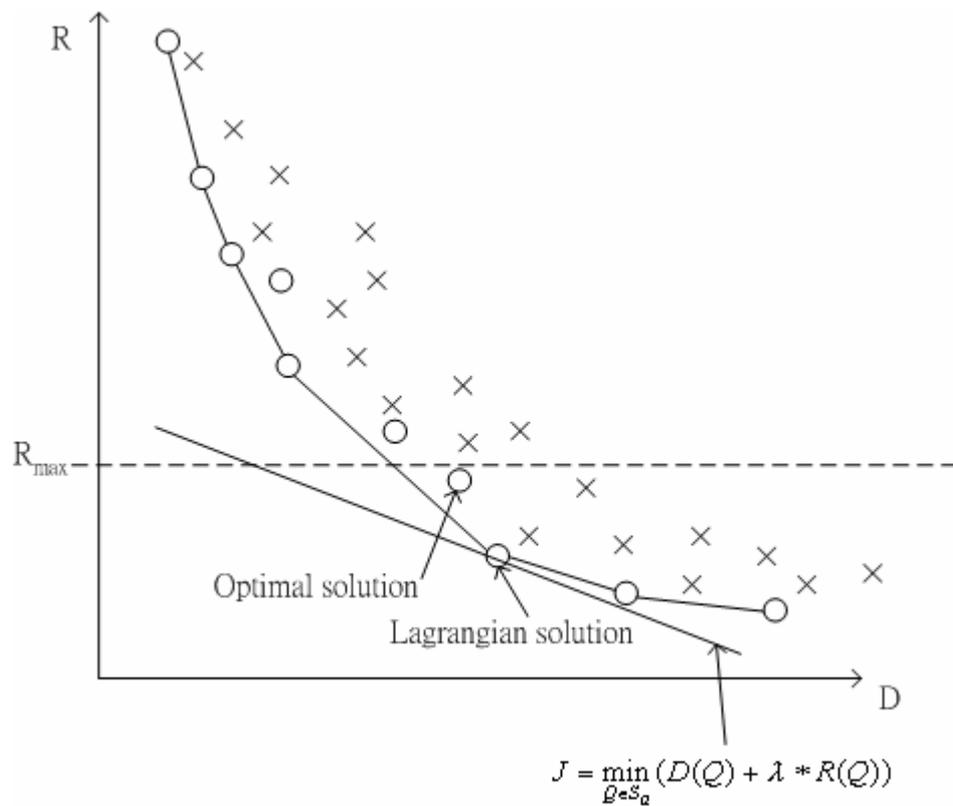


Figure 2-7 The Lagrangian solution

As mentioned above, Lagrangian techniques cannot reach points that do not reside on the convex hull of the R-D curve. An alternative formulation is then proposed in [57], which is based on deterministic dynamic programming called trellis-based bit allocation. The possible combination of quantizers of the coding units is represented by a tree diagram with corresponding distortion for each branch,

as shown in Figure 2-8. The Viterbi algorithm is employed for choosing the best path, i.e. the best combination of quantizers.

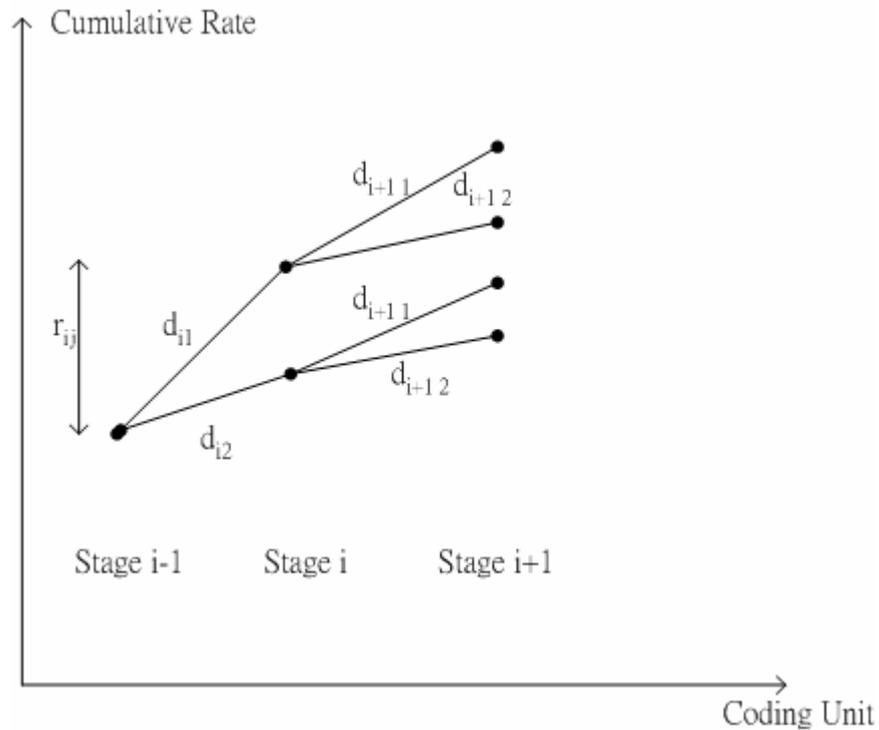


Figure 2-8 The Viterbi algorithm

Each stage represents a coding unit. The corresponding rate of each possible quantization choice is added to the cumulative rate in the y-axis. In each stage, these branches which result a rate exceeding R_{max} will be eliminated. In addition, for each stage, the branches with the lowest distortion will be taken while the branches having higher distortion up to that point should be removed or pruned from the tree.

After the whole tree has been worked out, the optimal path will be searched by the Viterbi algorithm which is widely used in trellis-based coding in

communications. At the final stages of the trellis tree, there should be only a few branches exist. Then, travel along the paths reversely from the final stage backward to the first stage. There is only one path that can reach the first stage in the inverse path. And, this is the optimal solution of the particular coding unit. Since the size of the tree is very large as the number of stages increases, a sliding window technique has also been proposed in [57] to limit the number of previous stages in the trellis tree. Other optimization techniques can be found in [58].

2.1.5 Typical rate control method

In order to estimate QP for a coding unit, rate controller is always given the current channel bandwidth. Based on the channel bandwidth (the target output bit rate), the controller estimates the allocated target rate for a particular coding unit and the target buffer fullness. Besides, the buffer fullness is varied based on channel state such that the start-up delay is under estimate [59]. The controller predicts the appropriate distortion or similar information based on the statistic of the current unit and previously coded unit. The QP is then estimated with the predicted distortion and the target bit rate for the current unit.

2.1.6 Rate control by prediction

Popular prediction schemes in the reference model of H.263+ [60], MPEG-4 [61, 62] and H.264 [63] are based on statistical rate-distortion model together with a bundle of buffer control algorithms. The computational afford required to determine the QP is normally low and hence such schemes always aim at real-time or low-delay video communications. In Section 2.2 and 2.3, the rate control schemes mentioned above are discussed where Joint Model 8.1a is explored in details. Besides those algorithms using rate-distortion statistic model, there are some other rate models that are not a function of distortion. Instead, the rate model is a function of “number of zeros” in the transformed bit stream [64, 65]. The advantage of such function is that the relationship between the number of zeros and the coded bit rate is always linear and hence the parameters of the linear rate model can be determined reliably.

2.1.7 Rate control by pre-analysis

In contrast with predictive rate control schemes, pre-analysis target on non-real-time video coding such as stored video and pre-recorded video transmission. Pre-analysis schemes consider different combinations of QP along consequent frames or MBs to yield the best QP set in particular channel situations. Some pre-

analysis rate control also aim at rate-distortion optimization [57, 66, 67] by iterating different coding modes.

2.2 Rate control in the Verification Model of MPEG-4

In VM5.0 of MPEG-4, the employed rate distortion model is derived from the expansion of Taylor series of a Laplacian distributed source as shown in (2.9). The rate-distortion model can then be derived as (2.10) and the Taylor series of the model can be expanded as,

$$R(D) = -\frac{3}{2} + \frac{2}{\alpha} D^{-1} - \frac{1}{2\alpha^2} D^{-2} + R_{remainder} \quad (2.21)$$

where $R_{remainder}$ is the remainder term of the series. This series can be approximated by using a quadratic model [61] as follows:

$$R(Q) = b_1 Q^{-1} + b_2 Q^{-2} \quad (2.22)$$

where Q is the quantization step, and b_1 and b_2 are the model parameters found by linear regression which determines the model parameters best fitting into the rate-distortion curves of previous frames. The formulation of the linear regression is shown as,

$$b_1 = \frac{n \sum_{i=1}^n R_i - \left(\sum_{i=1}^n Q_i^{-1} \right) \left(\sum_{i=1}^n Q_i R_i \right)}{n \sum_{i=1}^n Q_i^{-2} - \left(\sum_{i=1}^n Q_i^{-1} \right)^2} \quad (2.23)$$

$$b_2 = \frac{\sum_{i=1}^n Q_i R_i - b_1 \sum_{i=1}^n Q_i^{-1}}{n} \quad (2.24)$$

where n is the number of coded frames, Q_i and R_i are the average actual coding quantization step and bit rate respectively. Similar to typical rate control techniques, the number of bits available in the GOP is determined a priori. The total number of remaining bits in the current GOP is defined as,

$$R = T_I + N_P T_P + N_B T_B \quad (2.25)$$

where T_I , T_P and T_B are the target bit rate of the I, P and B frames respectively, N_P and N_B are the number of frames coded as P and B frames respectively. Furthermore, the number of I, P and B frames follows a certain ratio as follows,

$$\frac{K_P}{K_B} = \frac{Q_P}{Q_B} \quad (2.26)$$

$$\frac{K_P}{K_I} = \frac{Q_P}{Q_I} \quad (2.27)$$

where K_I , K_P and K_B are the constant ratio between I, P and B frames respectively. Q_I , Q_P and Q_B are the average quantization steps for I, P and B frames respectively.

Based on (2.22), the rate-distortion relationship can be described by,

$$T_I = b_1 Q_I^{-1} + b_2 Q_I^{-2} \quad (2.28)$$

$$T_P = b_1 Q_P^{-1} + b_2 Q_P^{-2} \quad (2.29)$$

$$T_B = b_1 Q_B^{-1} + b_2 Q_B^{-2} \quad (2.30)$$

By solving the six equations (2.25) - (2.30), the target bit rate T_I , T_P and T_B , the quantization steps Q_I , Q_P and Q_B can be determined.

2.3 Rate control in the Joint Model of H.264

The rate control technique [63] adopted in the reference model JM8.1a [68] of H.264 is based on a quadratic rate-distortion model similar to those adopted in MPEG-4 VM5.0 as mentioned in Section 2.2. Although both rate control techniques employ similar rate-distortion models, their overall approaches are quite different. In this rate control technique, the quantization steps and parameters of bi-directional frames are computed through interpolation methods rather than using the rate-distortion model. Furthermore, the rate control process in JM is divided into Group of Picture (GOP) level, frame level and optional basic unit (BU) level where GOP level is a must level and either the frame level or the BU level can be chosen to cooperate with the GOP level.

2.3.1 Group of picture (GOP) level

In this level, the bit rate available for the rest frames in the GOP is estimated as shown in (2.31) and (2.32). The number of bits available for the rest of the GOP is updated according to the current bandwidth as shown in the lower equation in (2.31).

$$B(k) = \begin{cases} \frac{R_c(k)}{f} \times N - V(k), & k = 1 \\ B(k-1) + \frac{R_c(k) - R_c(k-1)}{f} \times (N - k + 1) - b(k-1), & k = 2, 3, \dots, N \end{cases} \quad (2.31)$$

where $R_c(k)$ is the instant available channel bit rate when coding the k^{th} frame, N is the number of frames in the current GOP, $V(k)$ is the current occupancy of the buffer and f is the pre-defined frame rate, b is the bit generated in the frame $(k - 1)$. Equation (2.31) sets the initial number of available bits of the current GOP by multiplying the average bit rate per frame by the number of the rest frames in the GOP. This estimates the number of bits available for each frame in that GOP. $V(k)$ is the bits from the previous GOP that cannot be sent within the period of that GOP and hence it will occupy some bandwidth of the current GOP. In (2.31), the available bits $B(k)$ for the GOP is updated according to the channel bit rate by adding the channel bit rate difference between the instants of frame k and $(k - 1)$ and subtracting the actual coded bit rate of frame $(k - 1)$. In case the channel is CBR, the term $\frac{R_c(k) - R_c(k-1)}{f} \times (N - k + 1)$ in (2.31) is equal to zero and the equation is simplified as follows,

$$B(i) = B(k-1) - b(k-1) \quad (2.32)$$

The initial quantization parameter $QP(1)$ of the first GOP, can be defined through the configuration file in JM8.1a. If the initial $QP(1)$ is not defined (indicated by setting $QP = 0$), then the system assigns the initial QP as follows,

$$QP(1) = \begin{cases} 35 & bpp \leq l1 \\ 25 & l1 < bpp \leq l2 \\ 20 & l2 < bpp \leq l3 \\ 10 & bpp > l3 \end{cases} \quad (2.33)$$

where $l1 = 0.1$, $l2 = 0.3$ and $l3 = 0.6$ for QCIF format, $l1 = 0.2$, $l2 = 0.6$ and $l3 = 1.2$ for CIF format. $l1 = 0.6$, $l2 = 1.4$ and $l3 = 2.4$ for resolution larger than CIF. Note that the definition of the numbers shown above is different from those in [63]. The numbers shown here are based on JM8.1a. For the following GOPs, the $QP(1)$ is set as follows,

$$QP(1) = \max \left\{ QP'(1) - 2, \min \left[QP'(1) + 2, \overline{QP} - \min \left(2, \frac{N}{15} \right) \right] \right\} \quad (2.34)$$

$$QP(1) = QP(1) - 1 \quad \text{if} \quad QP(1) > QP'(N - L) - 2 \quad (2.35)$$

where \overline{QP} is the mean of QP of the stored frames in the previous GOP, QP' is the QP of the previous GOP and L is the number of non-stored frame between two stored frames. These two equations limit the QP fluctuation of the first stored frame (intra frame), in other words, the quality fluctuation is limited. Stored frames are referred to those frames that are stored as reference frames for motion compensation. Therefore, the quality fluctuations of stored frames always propagate to other non-stored frames.

2.3.2 Frame level

2.3.2.1 Non-stored frames

Non-stored frames are referred as those frames that will not be referenced in motion compensation and hence the frames are not stored in the cache while encoding or decoding. Therefore, the effect of QP does not propagate to other frames and the quality of the decoded non-stored frames does not affect other frames. As a consequent fact, the choice of QP for non-stored frames is less important and the process of choosing the QP can be simplified. In JM8.1a, the QPs of non-stored frames are estimated by interpolation method based on QPs of previous and later stored frames. Suppose $QP(k)$ and $QP(k + L + 1)$ are the QPs of the k^{th} and $(k + L + 1)^{\text{th}}$ frames and they are stored frames. For the non-stored frames between these two stored frames, the QPs are estimated as follows,

Case 1: There is only one non-stored frames between k^{th} and $(k + L + 1)^{\text{th}}$ frames, i.e. $L=1$.

$$QP(k+1) = \begin{cases} \frac{QP(k) + QP(k+2) + 2}{2} & \text{if } QP(k) \neq QP(k+2) \\ QP(k) + 2 & \text{Otherwise} \end{cases} \quad (2.36)$$

Case 2: There are more than one non-stored frames between k^{th} and $(k + L + 1)^{\text{th}}$ frames, i.e. $L > 1$.

$$QP(k+s) = QP(k) + \phi + \max \left\{ \min \left[\frac{(QP(k+L+1) - QP(k))}{L-1}, 2 \times (s-1) \right], -2 \times (s-1) \right\} \quad (2.37)$$

where $s = 1, 2, \dots, L$ and

$$\phi = \begin{cases} -3 & (k+L+1) - QP(k) \leq -2 \times L - 3 \\ -2 & QP(k+L+1) - QP(k) = -2 \times L - 2 \\ -1 & QP(k+L+1) - QP(k) = -2 \times L - 1 \\ 0 & QP(k+L+1) - QP(k) = -2 \times L \\ 1 & QP(k+L+1) - QP(k) = -2 \times L + 1 \\ 2 & \text{Otherwise} \end{cases} \quad (2.38)$$

After all, QP is further bounded to be within 1 and 51.

2.3.2.2 Stored frames

As mentioned in Section 2.3.1, stored frames are those frames will be referenced by other frames during motion compensations. The choices of the QPs of such frames always affect the quality of the other frames in the same GOP. Therefore, the estimation of such QPs is relatively important. In order to estimate an appropriate QP and quantization step size with the rate-distortion model, the target rate of the current frame must be known. The methods estimating the target rate can be very different in various video standards and other rate control techniques. In JM8.1a, the rate is estimated by jointly considering the target buffer level and the weighted complexity of the previous frames.

The target buffer level of the second frame of the current GOP is updated as follows,

$$S(k+1) = S(k) - \frac{S(2)}{N_p(k)-1} + \frac{\bar{W}_p(k) \times (L+1) \times R_c(k)}{f \times (\bar{W}_p(k) + \bar{W}_b(k) \times L)} - \frac{R_c(k)}{f} \quad (2.39)$$

where \bar{W}_p and \bar{W}_b are the average complexity of the stored frames and non-stored frames respectively. The complexity factors are determined by,

$$\begin{aligned} \bar{W}_p(k) &= \frac{W_p(k)}{8} + \frac{7 \times \bar{W}_p(k-1)}{8} \\ \bar{W}_b(k) &= \frac{W_b(k)}{8} + \frac{7 \times \bar{W}_b(k-1)}{8} \\ W_p(k) &= b(k) \times QP_p(k) \\ W_b(k) &= \frac{b(k) \times QP_b(k)}{1.3636} \end{aligned} \quad (2.40)$$

The final target bit rate is combined by two components, the first component is,

$$\tilde{R}(k) = \frac{R_c(k)}{f} + \varphi \times (S(k) - V(k)) \quad (2.41)$$

where φ is set to 0.5 if no non-stored frames presented otherwise, it is set to 0.25.

The second one is,

$$\hat{R}(k) = \frac{W_p(k-1) \times B(k)}{W_p(k-1) \times N_p^r + W_b(k-1) \times N_b^r} \quad (2.42)$$

where N_p^r and N_b^r are the number of remaining stored and non-stored frames in the GOP respectively. The final target bit rate is the weighted sum of (2.41) and (2.42),

$$R(k) = \eta \times \hat{R}(k) + (1 - \eta) \times \tilde{R}(k) \quad (2.43)$$

where η is a constant set to 0.5 if no non-stored frame present and 0.9 otherwise.

2.3.2.3 The quadratic rate-distortion model

As mentioned in Section 2.1.5, the reason for determining the quantities above is to find the quantization step $QStep$ and QP to code the current frame. A quadratic model similar to those adopted in VM5.0 is employed in JM8.1a. A detail of the rate-distortion model of VM5.0 is given in Section 2.2. Both of them are derived from the Taylor series expansion of the logarithm rate-distortion model of a Laplacian source as shown below,

$$T(k) = d_1 \times \frac{\tilde{\theta}(k)}{Q(k)} + d_2 \times \frac{\tilde{\theta}(k)}{Q^2(k)} - m_h(k) \quad (2.44)$$

where d_1 and d_2 are determined based on linear regression, Q is the quantization step and $\tilde{\theta}$ is the mean absolute difference (MAD) of the current stored frames. The MAD is predicted by a linear model as follows,

$$\tilde{\theta}(k) = g_1 \times \theta(k-1-L) + g_2 \quad (2.45)$$

where θ is the actual MAD of the previous stored frame, g_1 and g_2 are the model parameters determined by linear regression. The $QStep$ is mapped into QP once the $QStep$ is estimated. The $QStep$ - QP mapping is as discussed in Section 2.1.1. Finally, the mapped QP is further bounded by,

$$QP(k) = \min \{QP(k-L-1) + 2, \max [QP(k-L-1) - 2, QP(k)]\} \quad (2.46)$$

in order to limit the fluctuation of the QP.

2.3.3 Basic unit level

The basic unit level rate control is another important idea in which every basic unit (BU) can have its own QP. A basic unit is a consequence of MBs. The number of MBs in a BU is denoted by n_{unit} . The BU size can be as small as one MB and as large as the total number of MBs in the frame. The size of the basic unit is formulated as,

$$n_{unit} = \frac{n_{mbframe}}{n_{mbunit}} \quad (2.47)$$

where $n_{mbframe}$ is the number of MBs in a frame and n_{mbunit} is the number of MBs in a basic unit. The target bits to be assigned to the l^{th} basic unit are defined as follows:

$$\tilde{b}_l = R_r \cdot \frac{MAD_l^2}{\sum_{m=l}^{n_{unit}} MAD_m^2} \quad (2.48)$$

where R_r is the number of remaining bits for the remaining basic units in the current frame, and MAD_m is the estimated MAD value of the m^{th} MB. The number of texture bits b_l for the l^{th} basic unit is then given by subtracting the estimated header bits of the current basic unit from \tilde{b}_l . Based on this target bit rate for the current basic unit, the quantization step can be found by (2.44). In JM8.1a, a constraint, similar to (2.46), is employed to control the variation of the QP, and the selected quantization parameter $QP_{k,l}$ of the l^{th} MB of the k^{th} frame is bound as shown,

$$QP_{k,l} = \max(0, \overline{QP_{k-1}} - 6, \min(51, \overline{QP_{k-1}} + 6, QP_{k,l})) \quad (2.49)$$

After this rate control algorithm has been adopted to the JM, there are also other techniques proposed for the JM such as [69].

2.4 Fuzzy based rate control

Fuzzy logic is popular in various of system control applications. Some applications involve a huge amount of information and parameters. The systems are possibly too complicated to be analyzed such that a controller can yield the best resultant solution for a given set of input parameters. In these cases, fuzzy logic is a popular solution which makes the control easier and “fuzzy”. Typical fuzzy logic controller [70, 71] includes fuzzification, decision-making and de-fuzzification as shown in Figure 2-9. Fuzzification is a process to classify data into linguistic categories that are referred as fuzzy sets. The function defines the range of the fuzzy set is called membership function. The x -axis is the range of the input value and the y -axis of the membership function is the degree of membership that tells us how true the value should fall into a membership.

The shape of the membership function depends on the application and there is no prior rule to follow. Some system designers may design the shape and range of the membership function by including trained neural network [71].

There are two inputs to a fuzzy controller typically, one is the error of the quantity we are going to control and the other one is the derivative of time (DOT) of the error i.e. the speed of change of the error. The error is the difference between the actual output value of the quantity and our target value of the quantity. It describes how far the value is from the target value. The DOT gives the information about the direction of tendency of the quantity. These two inputs have their own membership functions, and the functions may have different definitions. Once the two inputs are fuzzified, the corresponding fuzzy sets are inputted to the decision-making process. A pre-defined rule matrix is a functional description of all possible combinations of the two fuzzified inputs. It also defines the decisions for different combinations of the inputs. Like the fuzzy sets, the decisions are also described by linguistic language. The last process is de-fuzzification in which the linguistic decision is mapped into a specific value. This value is the final output of the fuzzy logic controller. The content of the output value depends on the characteristic of the system.

Rate control is a problem to adjust the quantization parameter to control the buffer fullness, output bit rate and the perceptual quality. Hence the inputs to the fuzzy logic controller can be any one of these quantities or all of them. The final output of a fuzzy rate controller is, like the rate control techniques in Section 2.3, typically an appropriate quantization parameter which is expected to make the output

bit rate, buffer fullness and the perceptual quality get closer to the expected values [72-75]. Some other methods smooth the output bit rate by taking the benefits of CBR and VBR control [72].

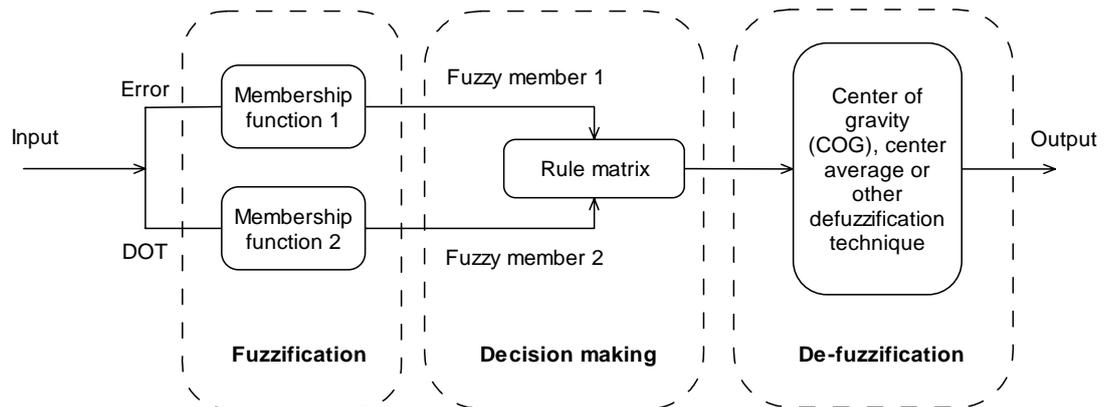


Figure 2-9 Typical fuzzy logic controller

2.5 Other rate control techniques

Besides controlling the output bit rate by controlling the quantization parameters, there are also some other approaches controlling the frame rate [17], resolution as well as the PSNR altogether [17-21].

Another kind of rate control is to sacrifice the perceptual quality in some region which is claimed to be less concerned or less interested by the human visual system (HVS). These techniques control the quantization parameters of the macroblocks in the region of interest (ROI) to be smaller while the other parts are with higher quantization parameters [76-78].

Chapter 3

Rate Control Based on Cubic Rate Model

In this section, we present a new rate-distortion model, which is based on a cubic function and its control constraints. The proposed algorithm is implemented in JM8.1a [63], the joint model of H.264. The performance of our algorithm is compared to the control algorithm proposed by Ma *et al.* [63]. Experimental results show that our algorithm outperforms the joint model control method with a variety of video sequences and bit rates.

This section is organized as follows. In Section 3.1.1, our cubic rate model and its control constraints will be presented. In Section 3.2.1, we will describe the use of the buffer fullness control scheme [63] in our algorithm. Experimental results will be shown and discussed in Section 3.3. Finally, we will conclude this chapter and give some remarks in Section 3.4.

3.1 The rate control model

3.1.1 The cubic rate model

The classical logarithmic rate-distortion model can be deduced from the rate-distortion theory if a source is known to be in a specific form of distribution [39]. A variety of different source distribution has been discussed in Section 2.1.3. The source distribution in video coding is believed to be Gaussian and Laplacian popularly. However, the distribution of the DCT coefficients of intra- and inter-coded pictures has been studied thoroughly, and is generally approximated by a Laplacian distribution [42]. The close form solution [45] of the rate-distortion function $R(D)$ for a Laplacian distribution source is given in (2.10). By expanding (2.10) using the Taylor series in (3.1), we have a formulation as in (2.21) and the approximation is (2.22).

$$R(D) = R(\tilde{d}) + R'(\tilde{d})(D - \tilde{d}) + \frac{R''(\tilde{d})}{2!}(D - \tilde{d})^2 + \dots + \frac{R^{(n)}(\tilde{d})}{n!}(D - \tilde{d})^n + \dots \quad (3.1)$$

In our proposed scheme, the model is a quantization-rate model rather than a rate-quantization model as in (2.44) and (2.22). The advantages of using a distortion-rate model are that a higher-order model can be employed easily, and when we perform rate control, the required quantization steps Q can be determined without much additional effort in order to achieve a targeted bit rate R .

For a rate-distortion model similar to (2.44), R is modeled as a function of Q ; it is computational to solve Q for a given R if the order of the model is increased further. Suppose the quadratic model is extended to be a cubic model. To solve the extended cubic R-D model, it requires iterative numerical methods in order to determine the appropriate quantization step for a given target rate R . Such iterative numerical methods require high computational effort. If the number of order is going to be increased, the process involves further heavy computation. This limits the order of the rate-distortion model to be used, as does the accuracy of the model. Therefore, we introduce a new modeling concept in which the order of the rate-distortion model can be increased easily with limited computational efforts.

The expansion of the order of the rate model is based on the inverse of classical rate model. The inverse of the rate-distortion model of a Laplacian source (2.10), i.e. the expression of distortion D in terms of the coding rate R , can be written as follows:

$$D(R) = \frac{1}{\alpha} e^{-R} \quad (3.2)$$

By using the Taylor series as shown in (3.1), the distortion-rate function can be expressed as follows:

$$Q(R) = c_1 R + c_2 R^2 + c_3 R^3 + \dots \quad (3.3)$$

where c_i are the model parameters, and (3.3) can be approximated by a second order model as follows:

$$Q(R) = c_1R + c_2R^2 \quad (3.4)$$

However, the quadratic model (3.4) cannot describe the distortion-rate curve efficiently. The shape of the individual terms of (3.3) and (3.4) is different from the exponential function in (2.22). The shape of the individual term in (2.22) is shown in Figure 3-1. Consequently, a much higher order function is required to model the curve in order to perform comparative accuracy to classical quadratic rate distortion model (2.22). Therefore, we model the distortion-rate curve using a form similar to (2.22) as follows:

$$Q(R) = a_1R^{-1} + a_2R^{-2} \quad (3.5)$$

where a_1 and a_2 represent the model parameters. The term of the model (3.5) is shown in Figure 3-2. Clearly the shape is much closer to those shown in Figure 3-1.

Higher-order model of the form is expected to be able to fit the rate-distortion curve better. With our proposed distortion-rate model, its order can be increased to any value without requiring much effort to solve the quantization scale Q . To determine the quantization step of the model for a given target rate, only simple substitution is required. For each increment of model order in (3.5), only one more addition and two more multiplications are required in order to determine the

quantization step. In our algorithm, a third-order term is included in the model as follows:

$$Q(R) = a_1R^{-1} + a_2R^{-2} + a_3R^{-3} \quad (3.6)$$

Note that (3.5) and (3.6) is not necessarily equivalent to the Taylor series (3.3)

. The model can be considered as a tool for curve fitting. To evaluate the performance of the cubic model, we obtain a number of rate-quantization (R - Q) pairs by inter-coding two successive frames F_{k-1} and F_k with quantization parameters ranging from 0 to 51 in JM8.1a. For each quantization parameter used to encode F_{k-1} , we can form the corresponding R - Q curve. Therefore, a total of 52 R - Q curves can be constructed. The quadratic model in (2.44) denoted as RD2, and our models (3.5) and (3.6), denoted as DR2 and DR3, respectively, are fitted to the 52 R - Q curves by the least-squares method, and the corresponding relative error of an estimated quantization step x' is measured as follows:

$$e = \frac{1}{n} \sum_j^n \frac{|x_j - x'_j|}{x_j} \quad (3.7)$$

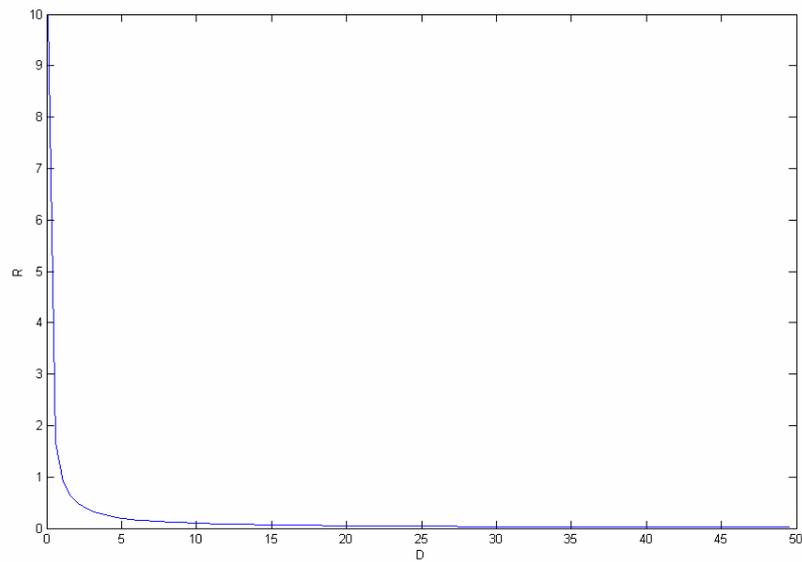
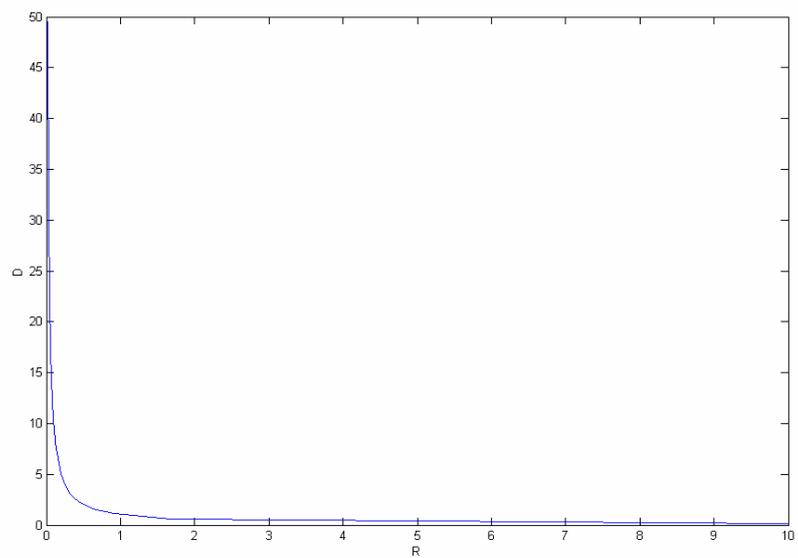
where x_j is the quantization step derived from the R - Q curves, x'_j is the corresponding estimated quantization step from the models, and n is the number of R - Q points used in the model.

In our experiments, our cubic model can fit into the $R-Q$ curves with lower relative errors for different video sequences. Table 3-1 tabulates the relative errors of the model RD2, and our proposed models DR2 and DR3 based on the video sequences “Mobile”, “Grandma” and “Foreman”.

Table 3-1 Average relative errors of the models based on different video sequences.

	RD2	DR2	DR3
Mobile	1.4785	1.4299	1.0475
Grandma	6.5319	5.7445	5.968
Foreman	5.24	6.254	5.01

The results show that our third-order model can achieve the lowest relative errors when compared to the second-order models. The two second-order models have similar performances; our second-order distortion-rate model performs better for the video sequences “Mobile” and “Grandma”, while the classical rate-distortion model achieves a lower relative error with the sequence “Foreman”.

Figure 3-1 The term $1/D$ in the R-D modelFigure 3-2 The term $1/R$ in the R-D model

3.1.2 The model parameters

The results depicted in Table 3-1 were obtained by best-fitting using the cubic model with the model parameters a_i estimated by using linear regression [79,

80]. The performance of the model depends on the estimation of the parameters. In this section, we show how to estimate the model parameters a_i based on prior knowledge from the previous coded frames.

The 52 R - Q curves of the video “Grandma” are plotted in Figure 3-3, which can show the influence of the quantization parameters QP_{k-1} of a previous reference frame F_{k-1} on the R - Q curves of the current frame F_k . In general, the curve will move upward when QP_{k-1} is higher, and downward when QP_{k-1} is smaller. Therefore, the model parameters of the current frame can be estimated from QP_{k-1} , or by another related quantity – the distortion of frame $k-1$ resulted from QP_{k-1} . In our experiments, the mean absolute difference (MAD) is employed as the distortion measure. Figure 3-4 to Figure 3-8 shows the relationships between the MAD, denoted by MAD_{k-1} , of the reference frame and the model parameters of the current frame. Although the relationships are not perfectly linear, a linear model, as shown in (3.8), can be employed to describe the relationships.

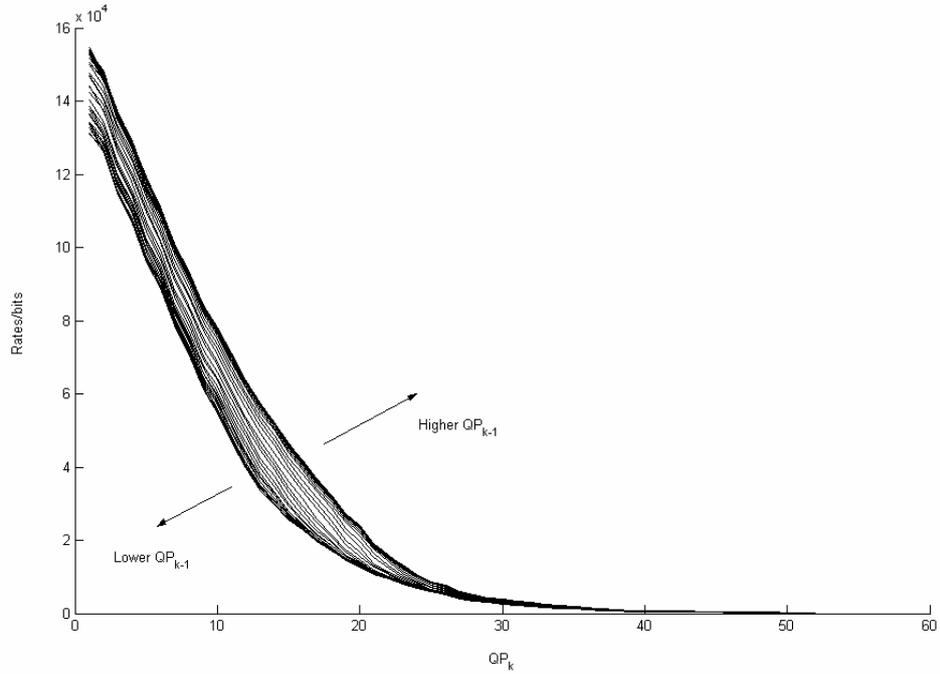


Figure 3-3 The effect of the previous quantization parameter on the R-Q curve of the current frame based on the video “Grandma”.

$$\hat{a} = \beta MAD_{k-1} + \gamma \quad (3.8)$$

where \hat{a} is a common factor of the three parameters, while β and γ are computed by linear regression. The use of the common factor allows us to perform the linear regression once in the computation of the three model parameters a_1 , a_2 and a_3 . To determine the value of the respective parameters, \hat{a} is multiplied by another parameter \bar{a}_i , i.e.

$$a_i = \hat{a} \bar{a}_i \quad (3.9)$$

and the cubic model can be written as follows:

$$Q(R) = \hat{a}(\bar{a}_1 R^{-1} + \bar{a}_2 R^{-2} + \bar{a}_3 R^{-3}) \quad (3.10)$$

where \bar{a}_i is estimated by linear regression in such a way that R is an independent variable while Q is the dependent variable. As a 3×3 matrix is required to be solved, the computation will become very intensive if the basic unit for rate control is small. The basic unit here refers to a group of MBs having the same quantization step, as defined in [63]. Therefore, in our algorithm, we propose a pseudo-regression method to greatly reduce the required computation. In our method, \bar{a}_3 is first set to zero, and a regression is done based on the first two terms of the model. The resulting \bar{a}_1 and \bar{a}_2 are then substituted into (3.10) and one more regression is performed to estimate \bar{a}_3 . Therefore, \bar{a}_3 can be considered as a refinement term in the model, and the computation is simple, as either one or two parameters are estimated in each of the regressions.

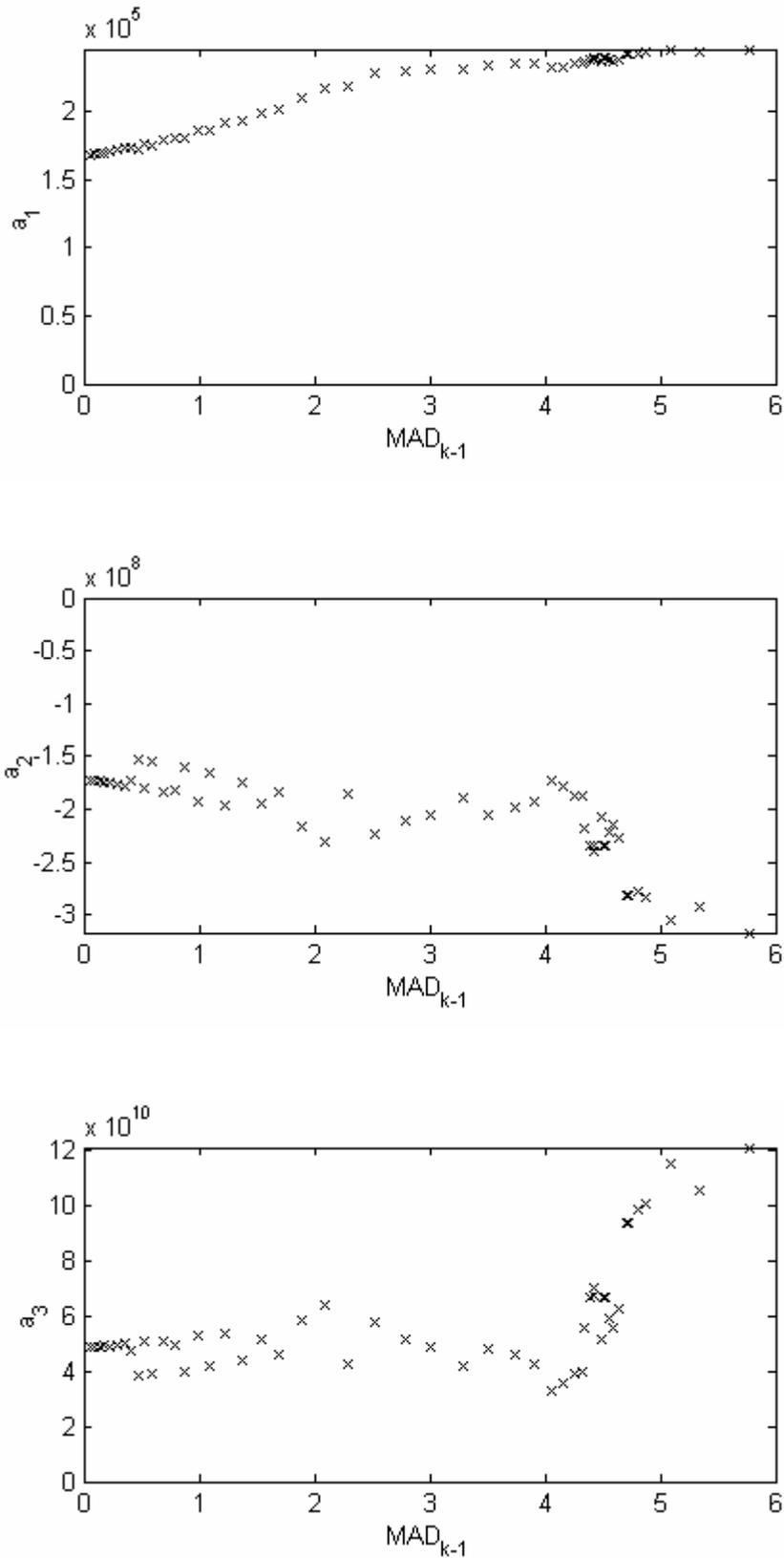


Figure 3-4 The relationship between MAD_{k-1} and a_i 's of the current frame for the sequence Mobile.

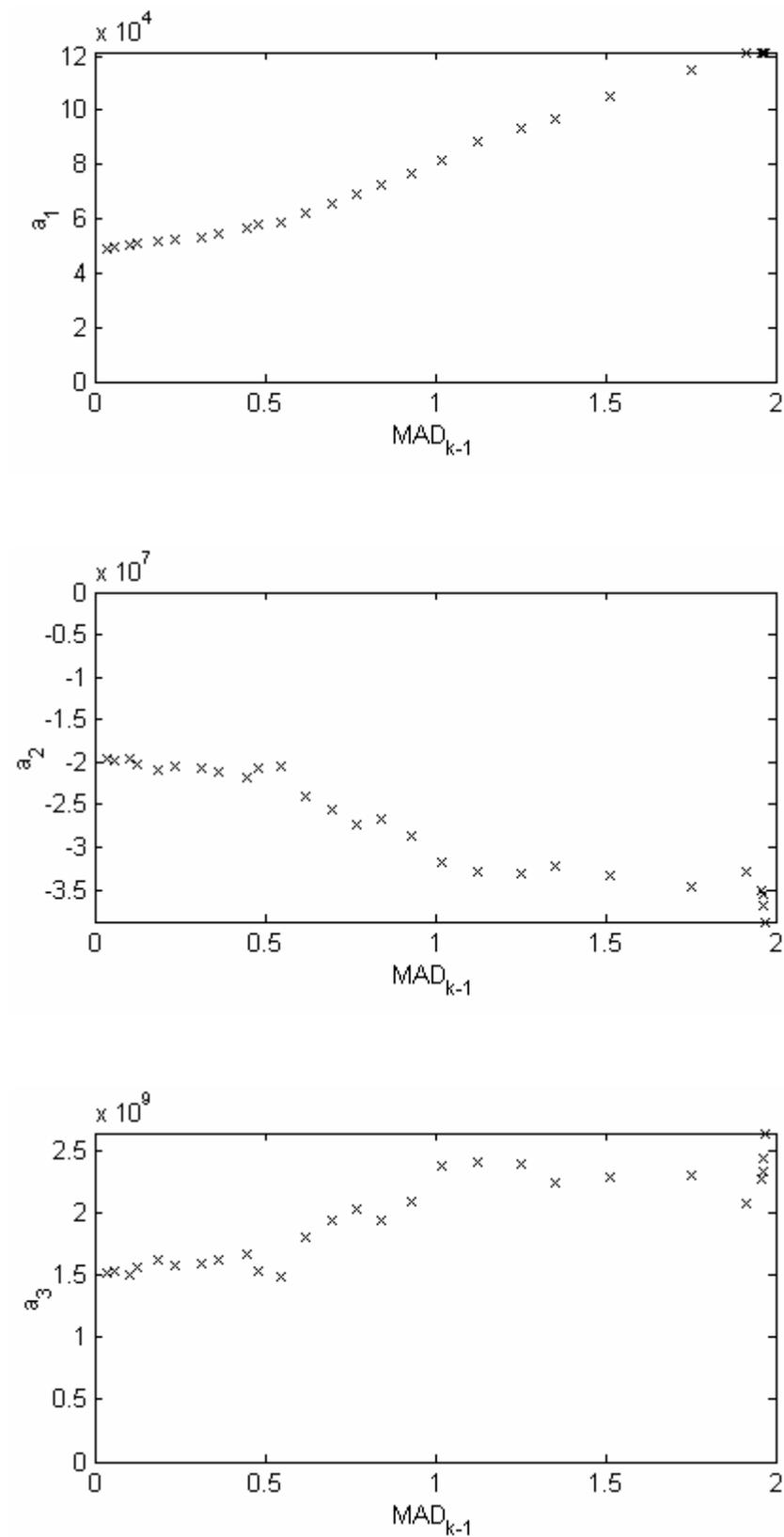


Figure 3-5 The relationship between MAD_{k-1} and a_i 's of the current frame for the sequence Grandma.

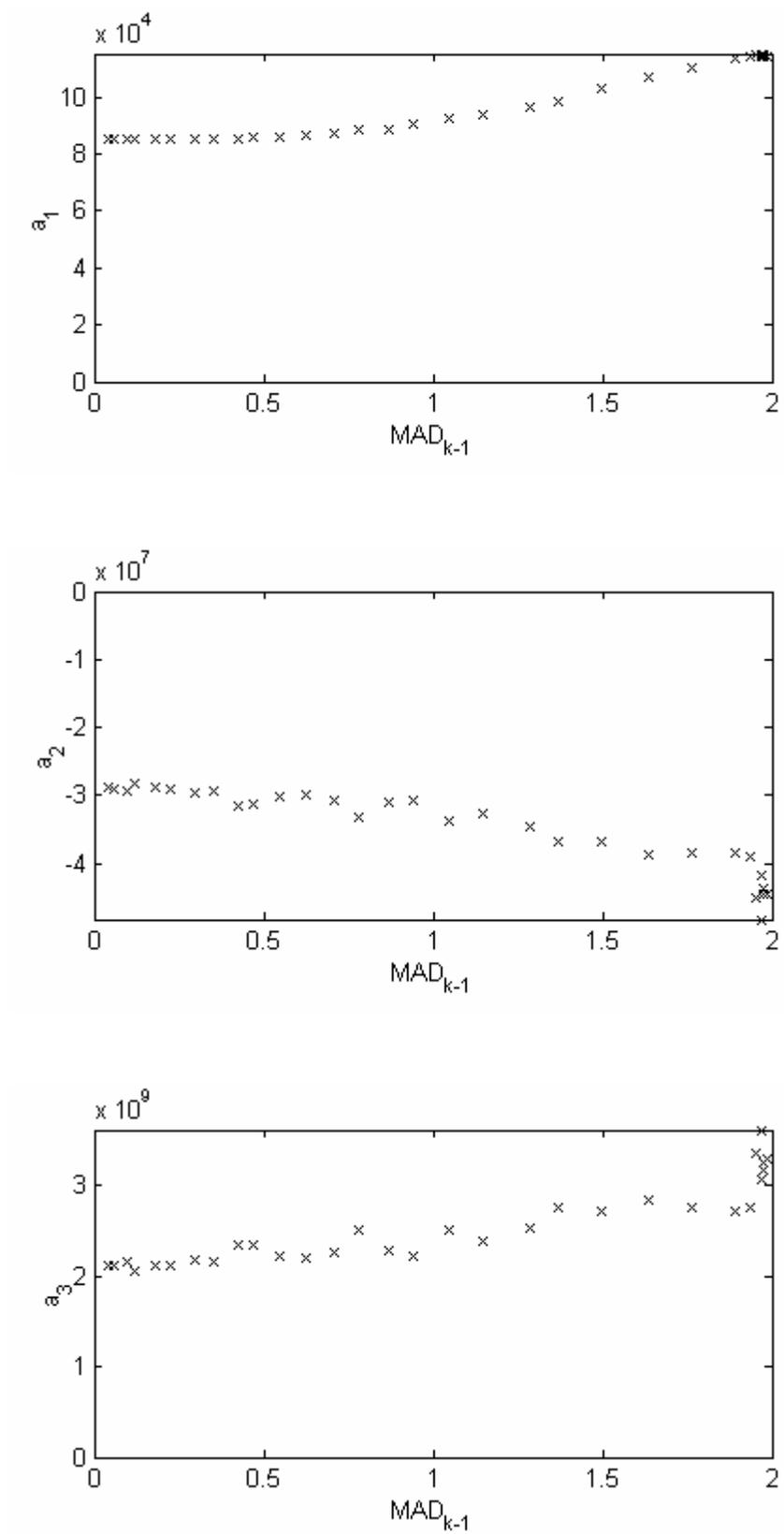


Figure 3-6 The relationship between MAD_{k-1} and a_i 's of the current frame for the sequence Hall.

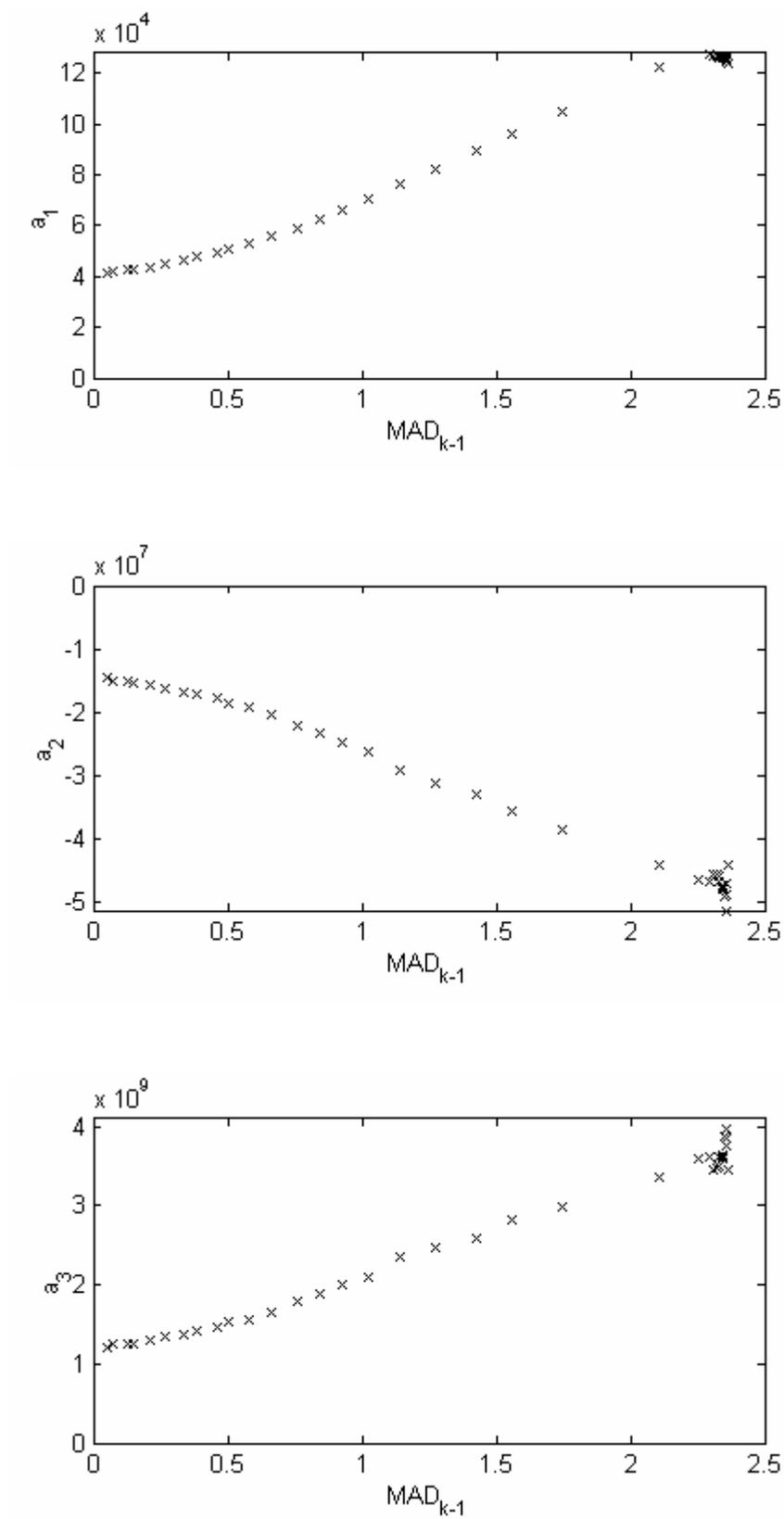


Figure 3-7 The relationship between MAD_{k-1} and a_i 's of the current frame for the sequence Salesman.

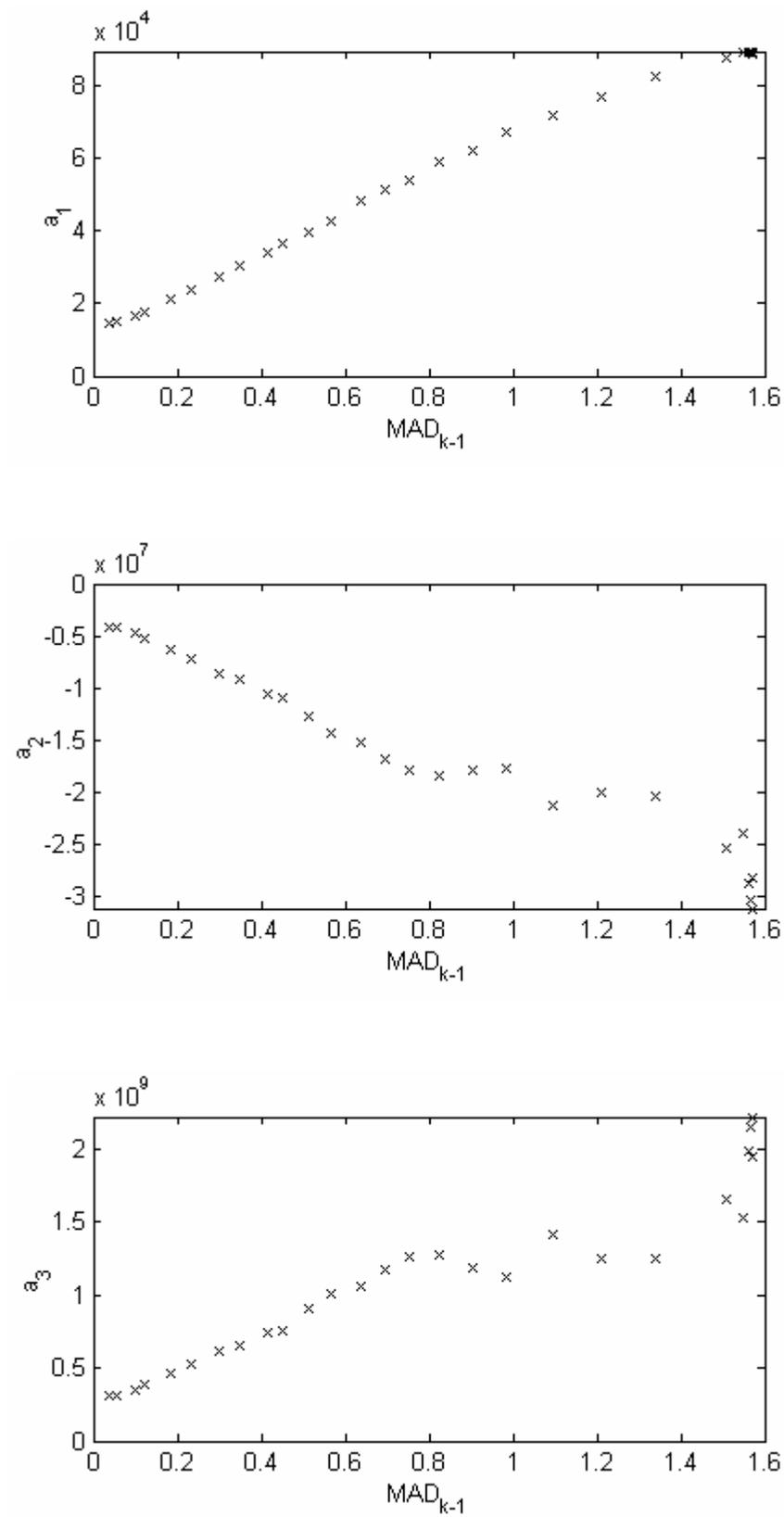


Figure 3-8 The relationship between MAD_{k-1} and a_i 's of the current frame for the sequence Akiyo.

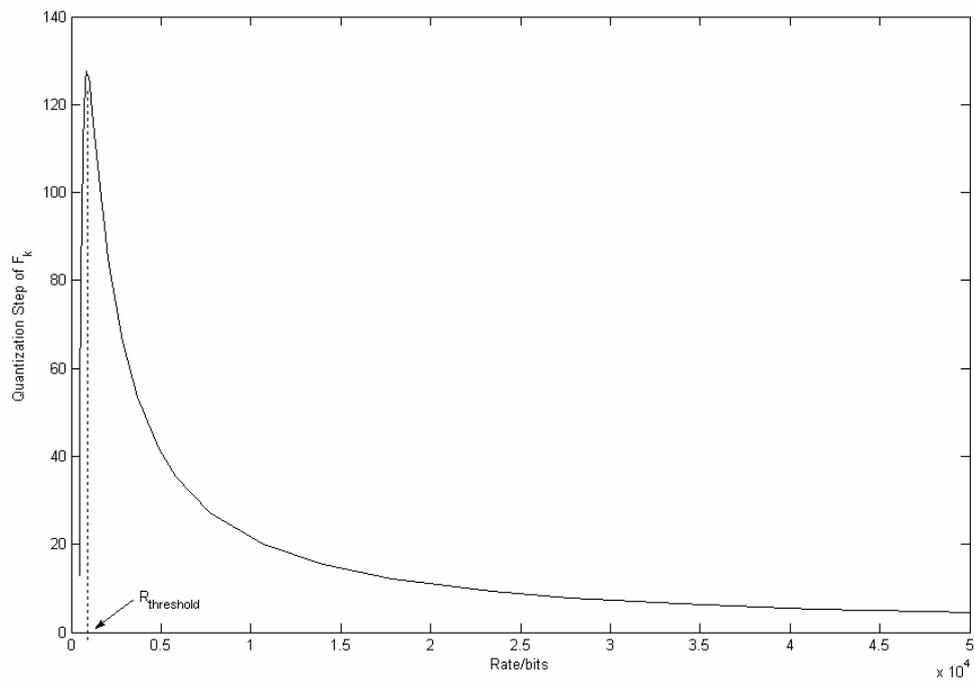


Figure 3-9 The rate threshold of the model based on Grandma.

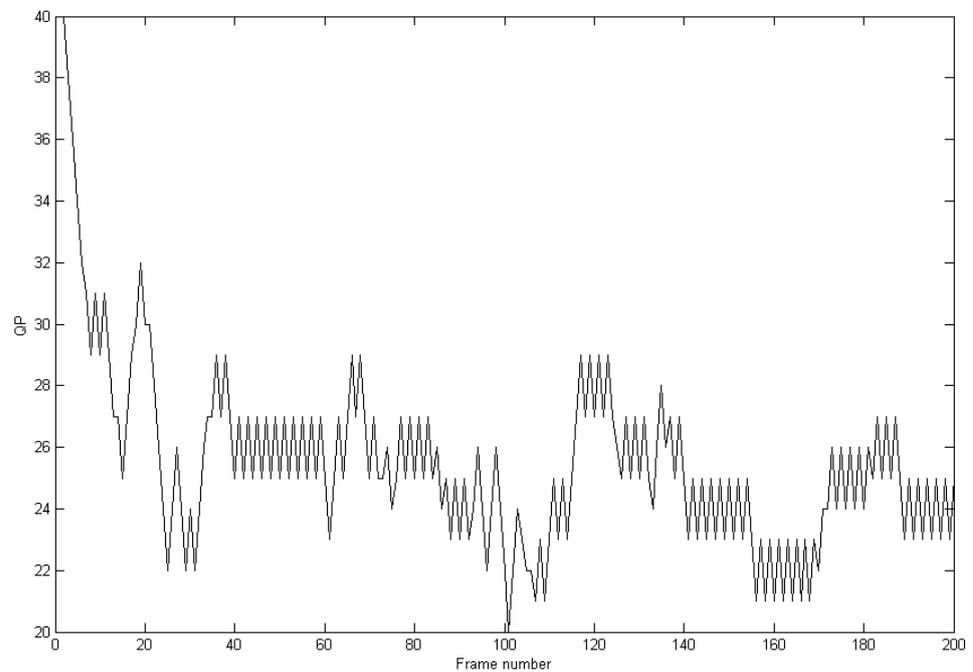


Figure 3-10 The selected quantization parameters for the first 200 frames of the video Grandma with target bit rate 64kbps.

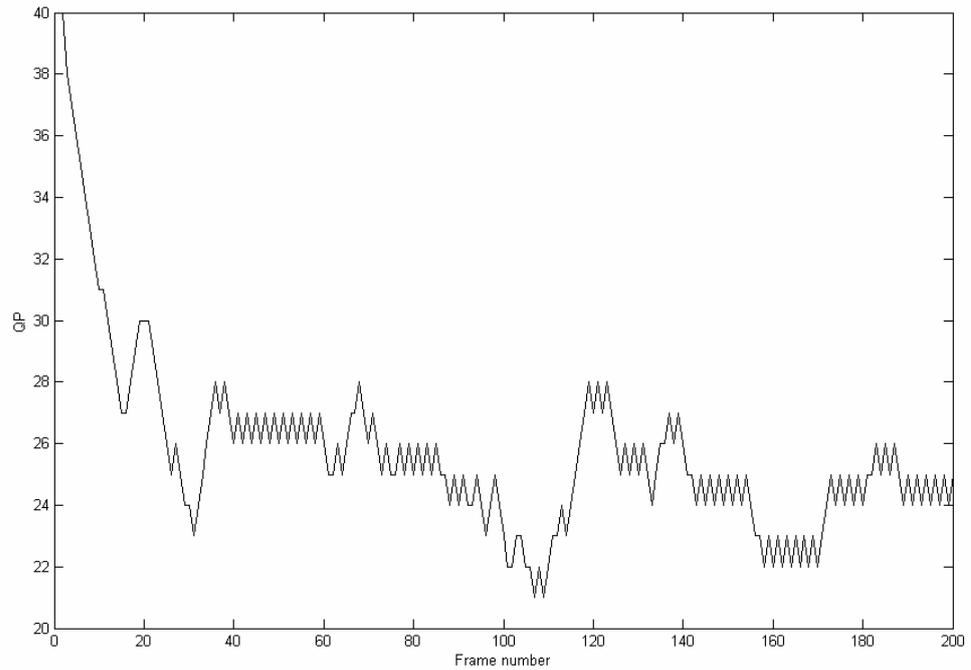


Figure 3-11 The smoothed quantization parameters selected from the successive frames of Grandma at target bit rate 64kbps.

As only the first two terms in (3.10) are considered in the first regression, the curve may be monotonic or non-monotonic, depending on its parameters.

Considering the first two terms only, (3.10) can be written as follows:

$$Q(R) = a_1 R^{-1} + a_2 R^{-2} \quad (3.11)$$

If this model is non-monotonic, a maximum may occur on the $Q(R)$ curve even if the R^{-3} term is included. In other words, a solution exists when we differentiate (3.10) and set it to zero. As a_1 , a_2 and a_3 may be positive or negative, we define a threshold equal to the more positive one of the following two solutions:

$$R_{threshold} = \frac{-2a_2 \pm \sqrt{\Delta}}{2a_1} \quad (3.12)$$

where $\Delta = 4a_2^2 - 12a_1a_3$. If $\Delta < 0$, the curve is monotonic decreasing and our proposed pseudo-regression can be employed. Otherwise, the threshold $R_{threshold}$ exists, which may be either positive or negative. Any negative solution will be replaced by 0. Figure 3-9 shows the $Q(R)$ curve when we consider the first two terms only based on the video “Grandma”. The model cannot reflect the appropriate quantization step to be used for a given target bit rate lower than the rate threshold $R_{threshold}$. If a given target rate of a frame or basic unit is lower than $R_{threshold}$, then the conventional 3×3 matrix regression is employed instead of the pseudo-regression. The $R_{threshold}$ will then be updated according to the set of new \bar{a}_i . However, it is possible that the target rate is still lower than the new $R_{threshold}$ with the 3×3 conventional regression. In this case, both \bar{a}_2 and \bar{a}_3 will be set to zero, and the model becomes a first-order one, in which case a further simple regression is required. Our experiments based on “Grandma” show that 12.6% and 1.4% of frames or basic units on average shifted to the use of the 3×3 conventional regression and the first-order model, respectively. The experiments also show that Δ is negative most of the time. In other words, the model is usually monotonic decreasing with the conventional regression.

3.2 Rate Control

In this section, the model proposed in Section II is applied for rate control on the platform JM8.1a. Our rate control scheme includes buffer control, target bit rate control, and quantization step refinement. Our proposed model is applicable to all the conventional hybrid video coders.

3.2.1 Buffer control

A rate control scheme is required to control the buffer level so that it neither overflows nor underflows. The buffer level at the beginning of coding the k^{th} frame is given as follows:

$$V_k = V_{k-1} - \frac{R_c}{f} + b_{k-1} \quad (3.13)$$

where V_{k-1} is the buffer level at the beginning of coding the previous frame, R_c is the current channel rate, f is the frame rate, b_{k-1} is the actual bit rate for the previous frame, and $k \geq 2$. At the beginning of coding a sequence, the buffer level, V_1 , is initialized to zero, and is updated after encoding each frame.

Buffer delay is caused by bits from the previous frames accumulating in the buffer. Some strategies can be employed to control the buffer fullness, and thus the buffer delay. Ribas-Corbera *et al.* [60] has proposed to use a frame-skipping algorithm to avoid further increase in buffer fullness when the buffer level reaches a

certain value. Any buffer delay control schemes such as frame skipping can be easily applied to our scheme. However, in our experiments, buffer delay control is not used so that we can compare the performance of our scheme to that in JM8.1a in a fair manner.

3.2.2 Frame level rate control

In JM8.1a, the target bit rate of the current frame is estimated by the current buffer fullness and the complexity weight of the previous frames for stored frames. The complexity weight is an estimation to the degree of complexity of the previous frames, and is estimated as follows:

$$W_{k-1} = r_{k-1} \times QP_{k-1} \quad (3.14)$$

where r_{k-1} and QP_{k-1} are the actual coded bit rate and the quantization parameter of the previous frame, respectively. The complexity weights of those previous frames in a GOP are averaged and used to predict the complexity of the current frame. The number of bits required to encode the current frame can then be estimated based on the current complexity weight.

With our model implemented in JM8.1a, the quantization step is estimated by substituting the target bit rate for the current frame as the variable R in the cubic model (3.10). The quantization-step-to-quantization-parameter mapping function is

then applied, and the final quantization parameter QP_k for the current frame is obtained

For the rate control in JM8.1a, the quantization parameter is bound by a constraint in order to maintain visual quality and smoothness along the successive frames. The simplified equation is as shown as follows,

$$QP_k = \min(QP_{k-1} + 2, \max(QP_{k-1} - 2, QP_k)) \quad (3.15)$$

The constraint bounds the variation of the quantization parameters along successive frames by ± 2 of QP_{k-1} . However, this constraint allows the variation of quantization to fluctuate continuously, as shown in the frames between frame 40 and frame 60 of “Grandma” at 64kbps in Figure 3-10. In our scheme, the constraint (3.15) is modified to be an adaptive constraint as

$$QP_k = \min(QP_{k-1} + \Delta QP, \max(QP_{k-1} - \Delta QP, QP_k)) \quad (3.16)$$

where ΔQP is an allowable variation of the quantization parameter along successive frames and ΔQP is defined by

$$\Delta QP = \max\left(\min(\sigma^2, \Delta QP_{\max}), \Delta QP_{\min}\right) \quad (3.17)$$

where ΔQP_{\max} and ΔQP_{\min} are the upper bound and lower bound of the quantization parameter, respectively, and σ^2 is the variance of the previous quantization parameters inside a sliding window of length w , as shown below:

$$\sigma^2 = \frac{\sum_j^w (QP_j - \overline{QP})^2}{w} \quad (3.18)$$

This constraint limits the successive quantization steps to fluctuating in a narrower range when the step sizes tend to be steady. Hence, this can achieve a smooth visual quality, as shown in Figure 3-11. In our experiments, ΔQP_{\max} and ΔQP_{\min} are set at 2 and 1, respectively, and $w = 3$. Note that both Figure 3-10 and Figure 3-11 are generated by the rate model in JM8.1a, but the latter is bounded by (3.16) to show the effect of the constraint under the same rate model. The resulting QP in (3.16) is then used to code the current k^{th} frame.

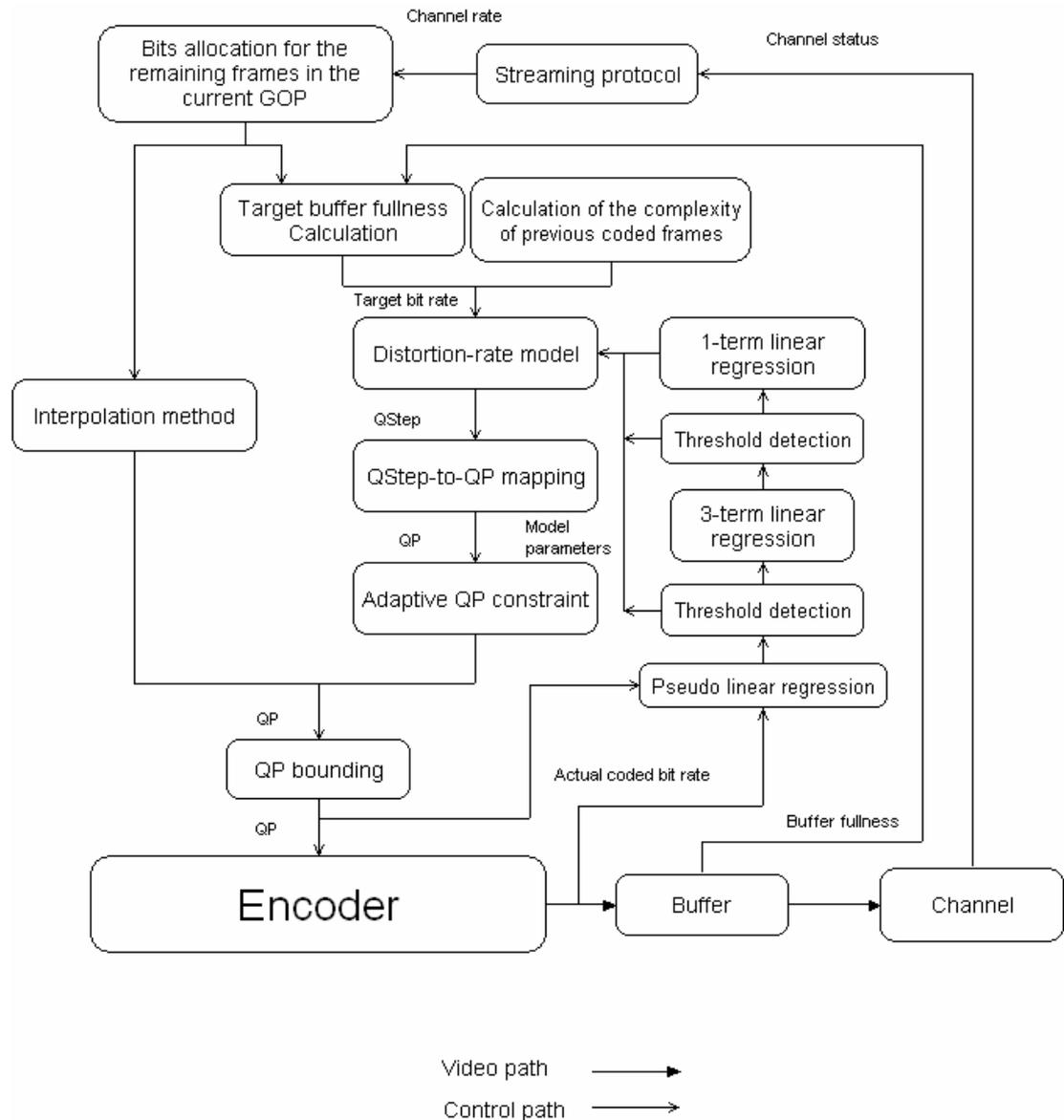


Figure 3-12 The rate control flow in the encoder side.

3.2.3 Basic unit level rate control

As mentioned in 2.3.3, the basic unit level of JM8.1a is defined a group of continuous MBs. By using our model (3.10) with $R = b_i$, the estimated quantization step can be found for the current basic unit. In JM8.1a, a constraint is employed to control the variation of the quantization steps, and the selected quantization

parameter $QP_{k,l}$ of the l^{th} MB of the k^{th} frame is bound as shown in (2.49). Similarly, the constraint can be modified as follows:

$$QP_{k,l} = \max(0, \overline{QP}_{k-1} - \Delta QP, \min(51, \overline{QP}_{k-1} + \Delta QP, QP_{k,l})) \quad (3.19)$$

where \overline{QP}_{k-1} is the mean quantization parameter among the basic units of the previous frame, and ΔQP is defined by (3.17) and (3.18) with QP as the quantization parameter of the co-located basic units in previous frames within the sliding window.

In our experiments, ΔQP_{\max} and ΔQP_{\min} are set at 6 and 3, respectively, and $w = 3$.

3.3 Experimental results

As mentioned in the previous sections, our scheme is implemented in H.264 JM8.1a and compared to the joint model [68]. In our experiments, the baseline profile is employed and we set the maximum search range to 16 and the number of B frames “NumberBFrames” to 0. The format of the tested sequence is YUV 4:2:0 QCIF at 30fps. The sequences are coded with different target bit rates: 24kbps, 48kbps and 64kbps in order to test the performance of the proposed scheme at different bit rates. The average PSNR and coded bit rates of both JM8.1a and our scheme are tabulated in Table 3-2. In the table, the average coded bit rate is compared to the target bit rate so that the closeness of the target bit rates achieved by the two schemes can be compared. The results show that our scheme achieves a

closer bit rate on average. The last column also shows the gain in PSNR based on our scheme. The average increased rate and PSNR gain are summarized in the last row of the table. The results show that our scheme can improve the PSNR for a wide variety of video sequences, and can achieve more accurate coded bit rates on average. The adaptive constraint (3.16) bounds the fluctuating QP in some occasions and stabilizes the QP in the frames that follow. Therefore the distortion propagation due to the QP fluctuation can be terminated. Some sequences such as N24 can achieve a significant PSNR gain. In general, if the best QStep is located on somewhere between two consequent QSteps provided by JM8.1a, then the QP will fluctuate around the optimal QStep.

Table 3-2 The performance on target rate and PSNR of our scheme as compared to JM8.1a. M=mobile, N=news, G=grandma and H=hall; the number that follows is the target rate (kbps).

Seq.	JM8.1a			Our scheme (DR3)			PSNR gain (dB)
	PSNR(dB)	Bit rate	Rate inc(%)	PSNR(dB)	Bit rate	Rate inc(%)	
M24	25.19	27.46	14.41	25.19	25.46	6.08	0.0017
N24	33.81	24.1	0.42	34.23	23.99	-0.04	0.4229
G24	37.15	24.17	0.70	37.96	24.03	0.13	0.8125
H48	38.83	47.88	-0.25	38.92	48.11	0.23	0.0953
Aver.			2.53			1.15	0.39

The generated bit rates are as stable as JM8.1a, but with a higher PSNR, as shown in Figure 3-13 to Figure 3-17. Some coded sequences are shown in Figure 3-18 to compare the visual quality based on the two schemes.

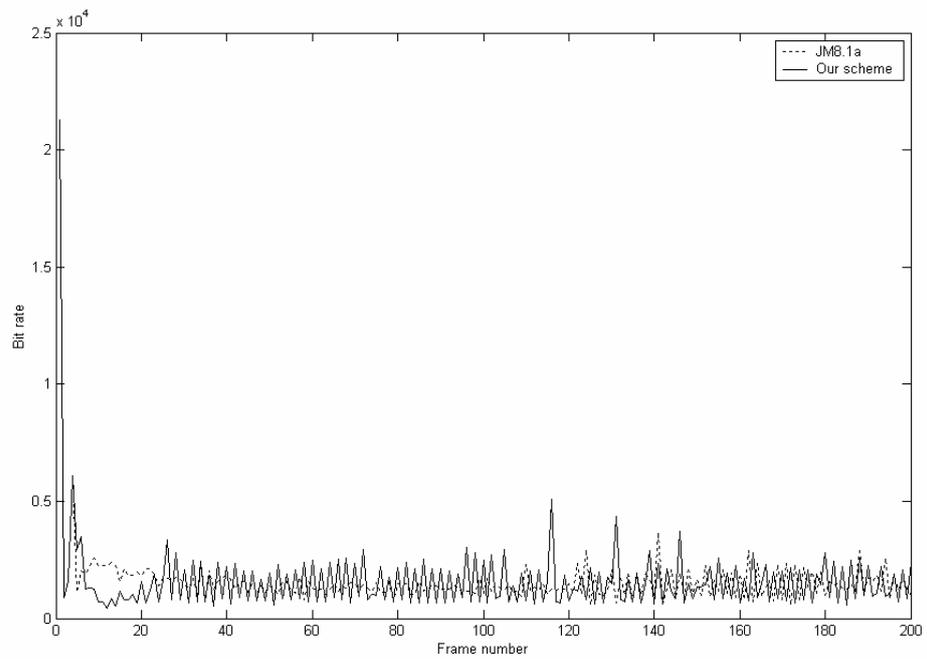
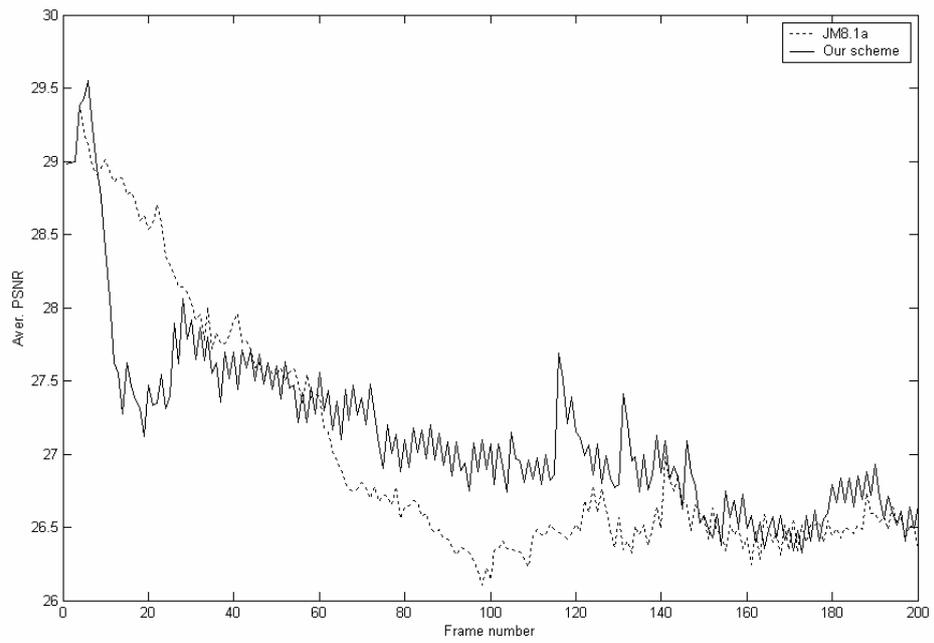


Figure 3-13 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video Mobile.

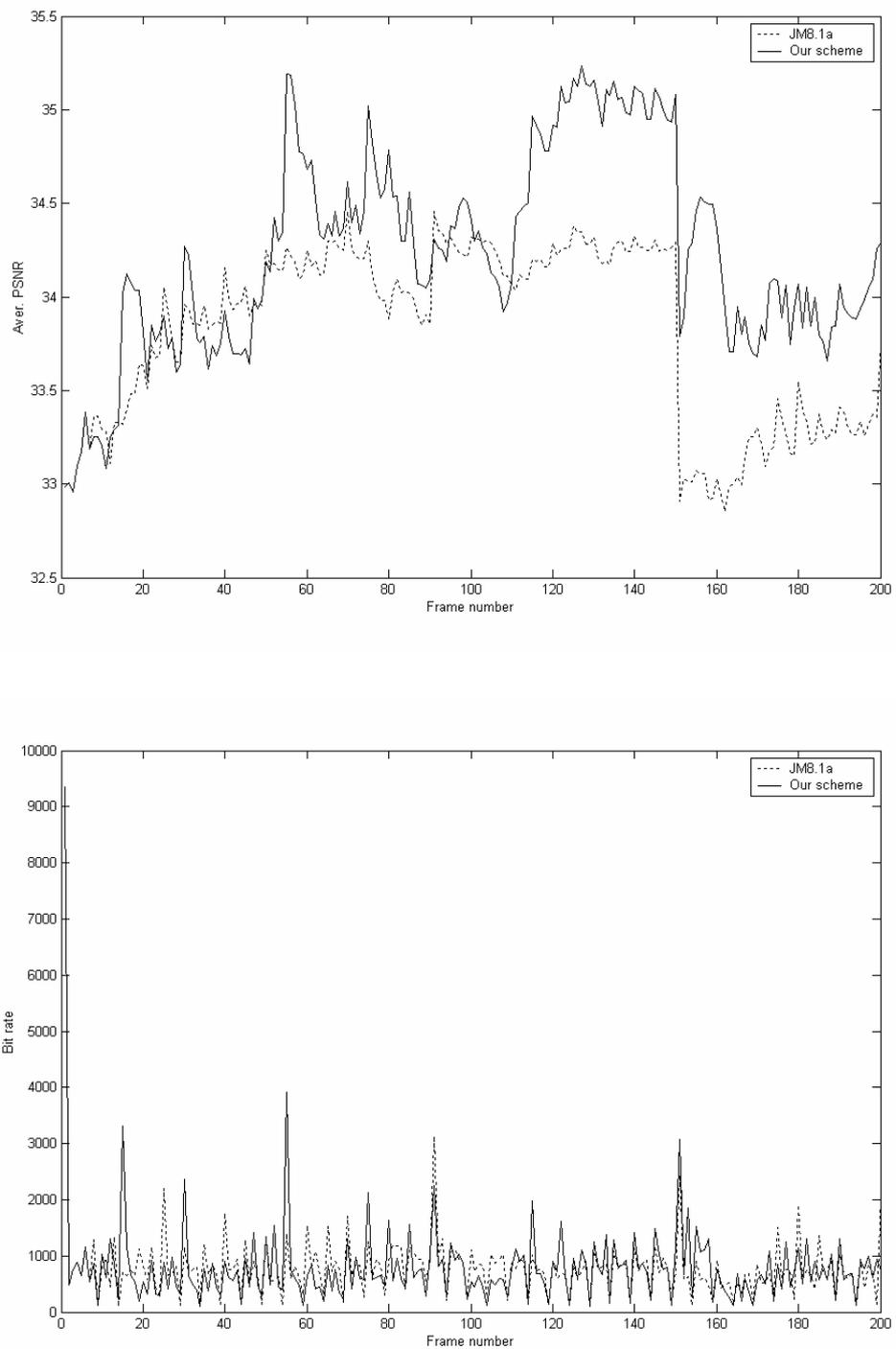


Figure 3-14 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video News.

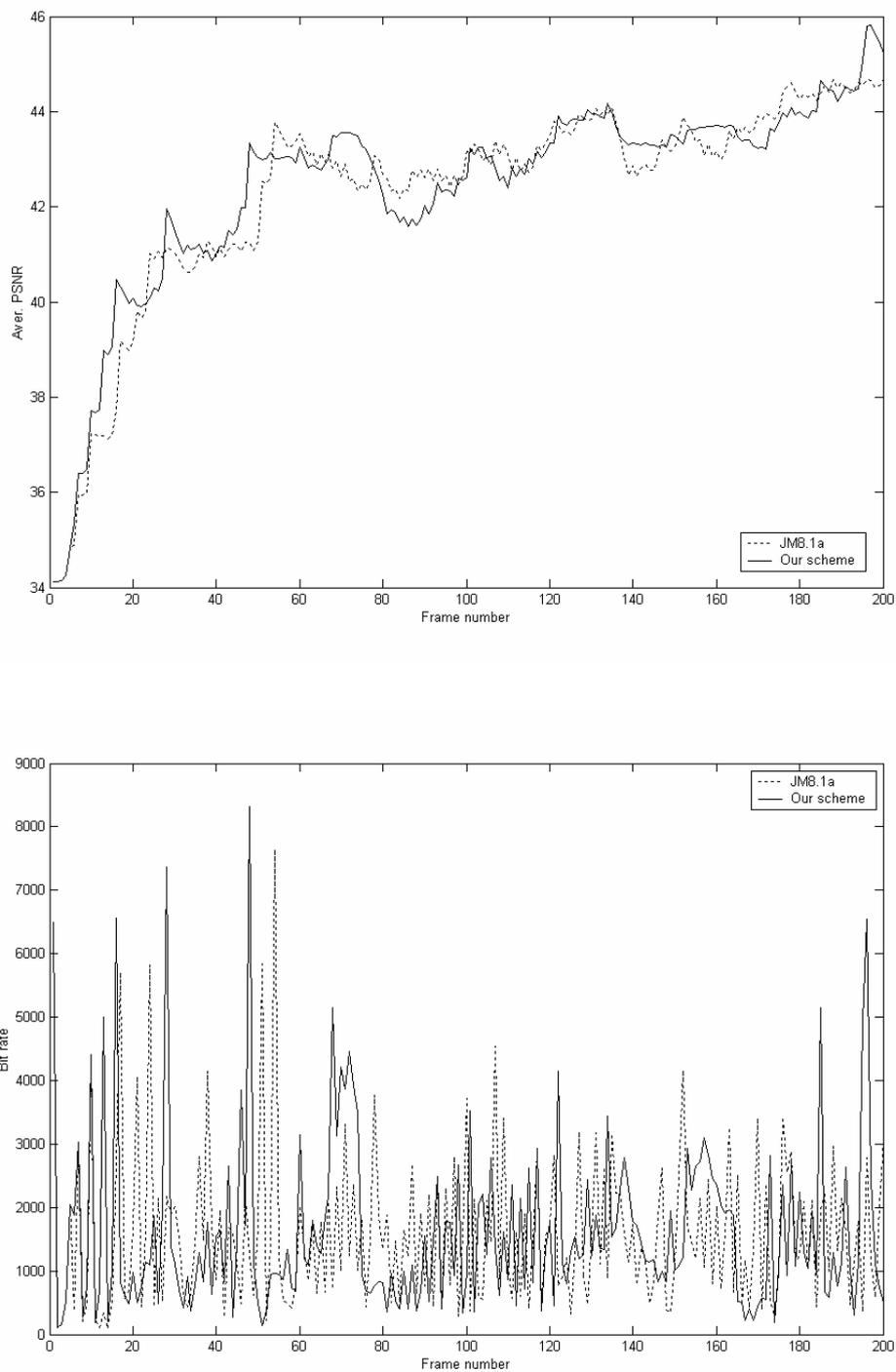


Figure 3-15 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video Akiyo.

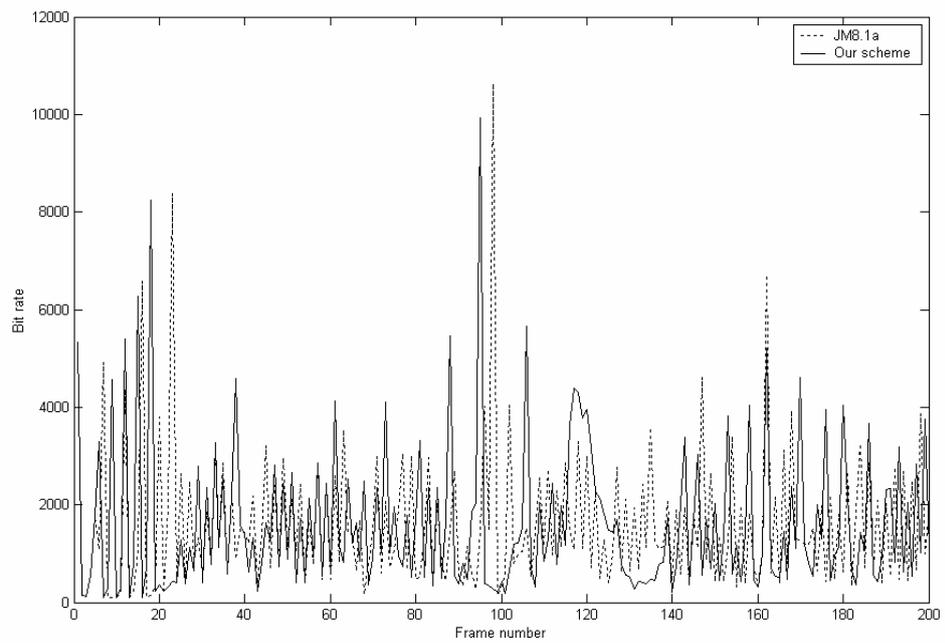
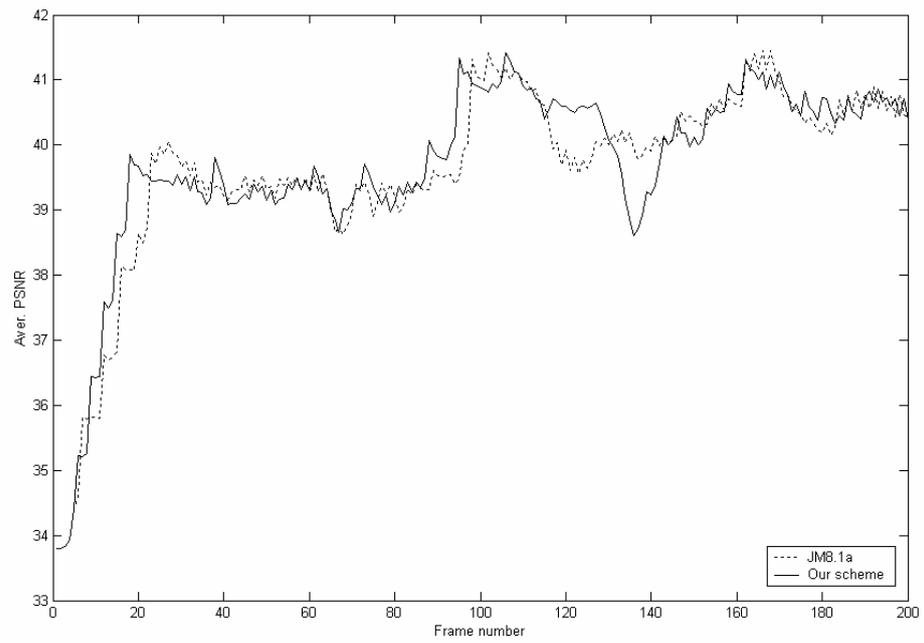


Figure 3-16 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video Grandma.

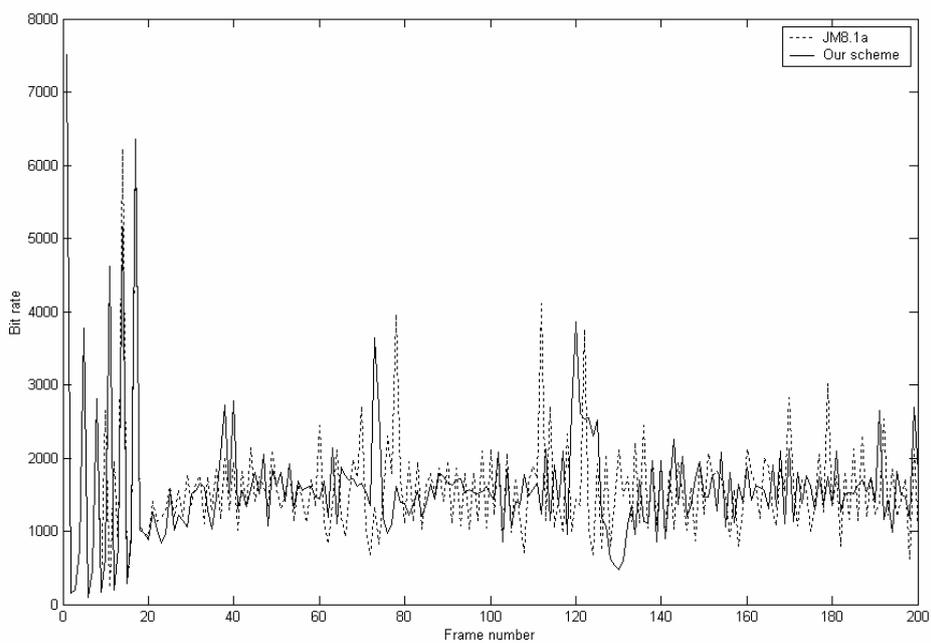
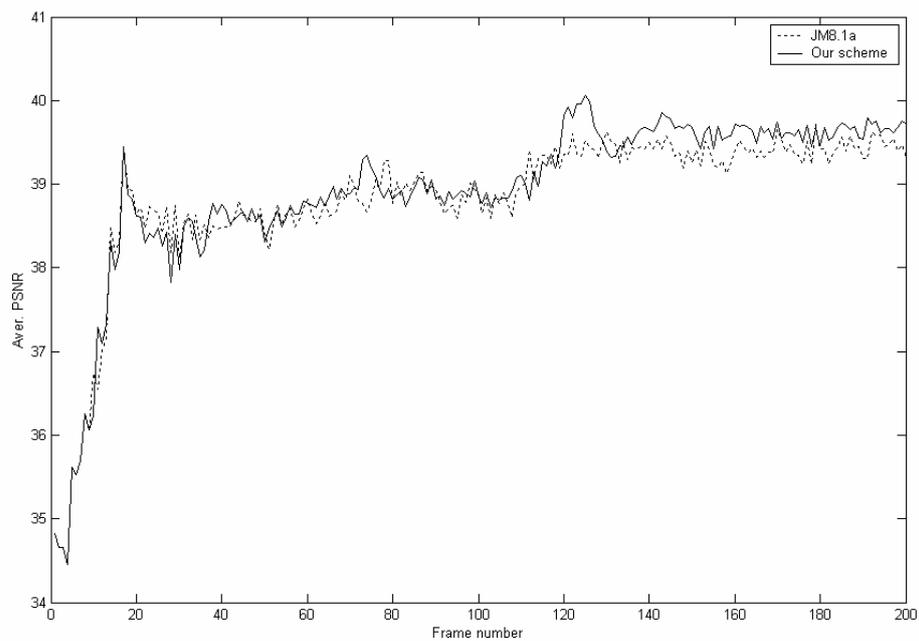


Figure 3-17 The PSNR in (dB) and the coded bit rates for the JM8.1a and our proposed scheme. The upper figure shows the PSNR and the lower figure shows the generated bit rates based on the video Hall.

3.4 Computation Analysis

The proposed rate control algorithm requires a small amount of computation for calculating the model parameters. The major portion of the increased complexity is resulted from the additional model parameters $\overline{a_1}$, $\overline{a_2}$ and $\overline{a_3}$. Although a pseudo approach is employed, a certain amount of calculation is required especially when the 3-term linear regression is employed. However, not all the coded units require the 3-term linear regression or further downscale to the one-term quadratic model as shown in Figure 3-12. The three model parameters can be solved by means of the Cramer rules in matrix form, as shown in (3.20),

$$\begin{bmatrix} n & \sum_i^n R_i^{-1} & \sum_i^n R_i^{-2} \\ \sum_i^n R_i^{-1} & \sum_i^n R_i^{-2} & \sum_i^n R_i^{-3} \\ \sum_i^n R_i^{-2} & \sum_i^n R_i^{-3} & \sum_i^n R_i^{-4} \end{bmatrix} \begin{bmatrix} \overline{a_1} \\ \overline{a_2} \\ \overline{a_3} \end{bmatrix} = \begin{bmatrix} \sum_i^n Q_i R_i \\ \sum_i^n Q_i \\ \sum_i^n Q_i R_i^{-1} \end{bmatrix} \quad (3.20)$$

In order to analyze the additional time and computation of our algorithm, the average numbers of addition and multiplication are considered as a reference. By coding the sequences “Mobile”, “News” and “Grandma” with QCIF resolution, 30fps at 48kbps, the average computational effort is obtained and the results are tabulated in Table 3-3.

Table 3-3 Average number of multiplication and addition of RD2 and DR3

	No of multiplication	No of addition
RD2	11	4
DR3	14.4	5.58

The computational effort of the proposed algorithm is slightly higher than that of RD2. In addition, the overall increase of coding time is another important indication of complexity. The coding processes, including motion estimation, mode selection and entropy coding, are the most time-consuming parts of JM8.1a. As a result, the increased runtime due to the additional computation required in our proposed algorithm is not significant. The runtimes are given in Table 3-4.

Table 3-4 Time consumption

	RD2	DR3
Mobile	151441ms	153582 ms
News	139086 ms	141212 ms
Grandma	128883 ms	131021 ms

3.5 Conclusion

In this chapter, we have presented a new cubic rate model together with its control schemes for typical hybrid video coders. Our proposed rate model is based on the classical quadratic model with extended terms. As the terms in our model no longer follow the Taylor series, the model is not necessarily monotonically increasing. Hence, we have also proposed a control scheme to avoid the coder operating in the range where the model fails. For the rate control part, a conventional

buffer fullness formulation is used such that existing frame rate control schemes can be employed. The overall performance of our scheme is better than the existing quadratic model based on the platform JM8.1a. The quadratic model outperforms other models at the time when H.264 was proposed and therefore adopted in the reference software. Thus, our model should outperform other classical rate-distortion models. The scheme can be implemented in other similar hybrid coders for low-delay video communications, as the computations required are low.





Figure 3-18 The visual quality using the JM8.1a rate control scheme (left side) and using our proposed scheme (right side) based on the tested sequences (a) News, (b) Akiyo and (c) Grandma.

Chapter 4

Simplified Pre-analysis Rate Control

In this chapter, the judgment criteria and existing problems of rate-control schemes are discussed. The judgment criteria to determine whether an algorithm can be based on frame-wise, block-wise, or object-wise optimization, and the computational complexity and time consumption of the algorithms will be described and analyzed. In the middle of this chapter, we will discuss the advantages of pre-analysis, which the predictive analysis of rate-control techniques cannot really include. Based on an existing predictive rate-control technique, we have developed a new model that can dramatically reduce the computation required based on a simplified version of the Verification Model (VM) as a testing platform.

4.1 The Problem of Dependency

Compression is the process of reducing the correlation among data such that most of the information can be compacted into fewer elements, symbols or coefficients. However, this will also increase dependency among the compressed

data. In other words, decoding a coefficient or symbol will depend on its previous coefficients. In any MPEG standard, a predicted frame depends on one or more reference frames. Any error occurring in a reference frame will therefore distort the predicted frame. Any distortion resulting from quantization in a reference frame will also be propagated in its corresponding predicted frames. This problem makes the design of rate-control schemes much more difficult, since choosing different quantization step $QStep$ in a reference frame can also change the R-D curve of the predicted frame. Hence, non-adaptive pre-analysis of the R-D curves of a predicted frame may not be useful in actual situations. Lin et al. [67] has proposed a rate-control algorithm for MPEG-1 and MPEG-2 video. Their experiments were conducted based on the Test Model 5 (TM5). They used the piecewise spline interpolation by considering 6 R-D samples rather than 31 samples. Each choice of the QP for the reference frame (denoted as QP_I) has its own corresponding R-D curve for the predicted frame. Figure 4-1 shows an example of the dependency curve. Since each line in Figure 4-1 corresponds to one QP_I , their dependency no longer has to be considered. The only action that the encoder needs to do is to choose the correct R-D curve for the P-frame according to the QP_I chosen in the I-frame. Chiang et al. [61] and Lee et al. [62], [23] proposed a quadratic rate distortion scheme, as discussed in Section 2.2. Their idea is based on modeling the R-D curves rather than

using the concept of operational rate control, as discussed in Section 2.1.2. They assume that the source is Laplacian and that the rate-distortion model is approximated by a quadratic equation that contains two parameters. These two parameters are computed using linear regression with the parameters used in the previous frames of the same GOP. The scheme provides a simple calculation by considering the dependency problem. However, similar to other modeling schemes, the source is assumed to be specified. All the derivations of the formulations have to be recalculated if other modeling functions are going to be used. The advantage of their schemes is that the parameters to be used are based on the previous parameters, and hence the encoder can estimate or find the R-D points without having to encode the frames.

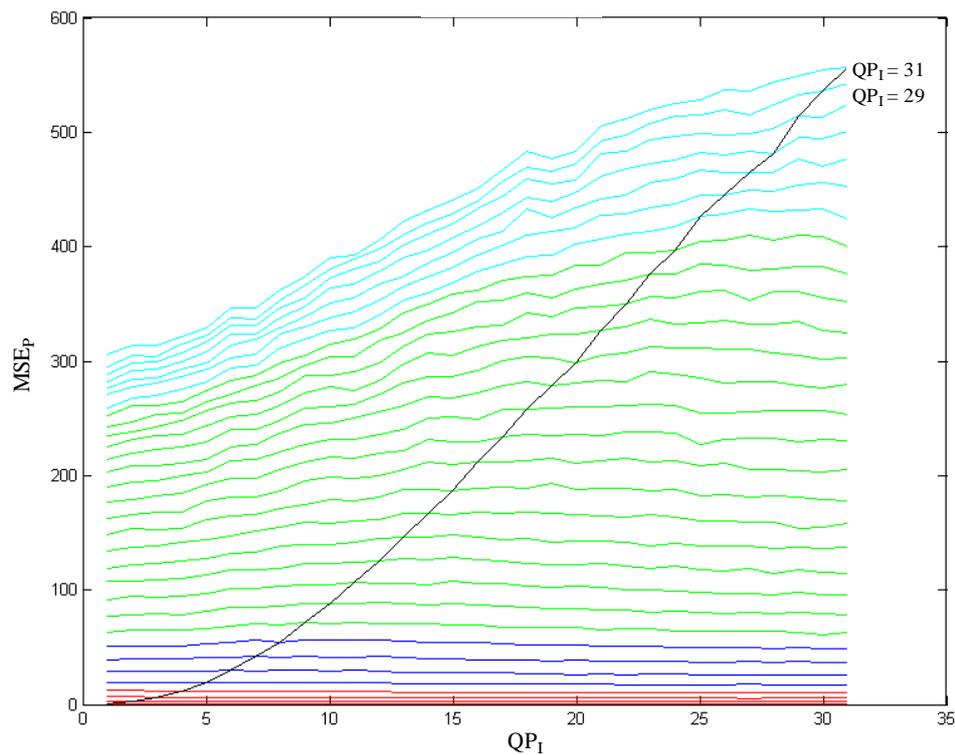


Figure 4-1 The distortion curves of a P-frame against the scaling factors of the corresponding I-frame.

4.1.1 Frame-wise, block-wise and object-wise optimization

Ortega et al. [57] proposed a technique based on trellis coding to find the optimal combination of scaling factors, as described in Section 2.1.1, for all the blocks in a frame. A block-wise technique can be used before a frame-wise technique. The abovementioned schemes are generally based on frame-wise optimization like [67]. The algorithm in [61] can be used for frame-wise, block-wise and object-wise (for MPEG-4) optimization. Of course, a scheme that is suitable for

frame-wise, block-wise and object-wise optimization is an advantage, but this depends on the specific applications and the system resources.

4.1.2 Time consumption of rate control Schemes

No matter that an operational rate-control scheme or an approximation model is used for video rate control, a significant number of operations are required. Operational rate control requires encoding a video sequence many times depending on the possible number of quantizers and the algorithm itself. Hence, this type of rate-control scheme cannot be applied to real-time video coding applications, and instead is suitable for stored video streaming in which the R-D curves are pre-analyzed. The most suitable quantizer is chosen based on the channel capacity to encode the sequence when a receiver or client makes a request. For some interactive applications and live broadcasting, operational rate-control schemes are not suitable. A typical example of a live broadcast is a football match transmitted over the Internet. The encoder has a stringent time constraint in which to analyze the sequence. Otherwise, the live broadcast is not appropriate for a live football match. The delay problem is much more serious for interactive applications, such as video conferencing, in which the round-trip delay is much more perceptible to end-users. R-D curve modeling is recommended for such applications, as it requires relatively

low computational effort as compared to the operational rate control. However, it may also need some adaptive statistical analysis to calculate the model parameters.

4.2 Approximation of operational rate control scheme

Operational rate-distortion schemes are generally more accurate, but with higher computational efforts. Therefore, some existing techniques still employ the operational rate-control scheme.

4.2.1 Intra-frame approximation

As mentioned in the previous sections, the operational rate-control scheme requires the whole R-Q set from all possible quantizers. Such a scheme is very time-consuming if no approximation is applied. Lin et al. [67] proposed a fast approximation technique which can preserve the accuracy of the R-D curves with lower computation efforts. In the scheme, they chose only some control points instead of all 31 points as in the MPEG. The control points can be found by the equation $QP_i = QP_{i-1} + QP_{i-2}$ i.e. the control points are (1, 2, 3, 5, 8, 13, 21, 31). Selecting the QP in this way can fit the rate curves and the distortion curves, as they are generally exponential functions. Cubic interpolation is then employed to approximate the points between any two adjacent control points. For example, to find

the R-D point when $QP = 7, (5, 8, 13, 21)$ are used to determine the four unknowns of the spline polynomial as follows:

$$f_j(X) = a_j X^3 + b_j X^2 + c_j X + d_j \quad (4.1)$$

4.2.2 Inter-frame dependency model

Another essential idea of the algorithm in [67] is the inter-frame dependency model. Since QP_I affects the R-D curve of the predicted frame, the scheme is to find out all possible effects from a QP_I . Let the rate and distortion functions for the P-frame be $D_p(QP_I, QP_p)$ and $R_p(QP_I, QP_p)$, respectively. $D_p(QP_I, QP_p)$ can be plotted as a 3D plot, as shown in Figure 4-2. If we set QP_I as a constant and plot $D_p(QP_I, q)$ for $q = (1, 2, \dots, 31)$ in the same figure, we will have the relationship as shown in Figure 4-7. Each line corresponds to one possible choice of QP_I , and $D_p(QP_I, q)$ does not increase further when QP_I and QP_p are equal. This linear-constant model is partly based on the selection of the skip-mode in MPEG. The distortion increases linearly as QP_I increases. The proof can be found in [66]. The final derivation in [66] is given below:

$$EL_p \approx F_I - F_p + EL_I \quad (4.2)$$

The relationship of the distortion of the P-frame is roughly, linearly proportional to the distortion of the I-frame (i.e. the distortion of the P-frame is

roughly proportional to the QP_I). Therefore, a linear-constant model is used to represent the curve as follows:

$$\begin{aligned} \text{If } QP_I \leq q, \text{ then } D_p(QP_I, q) &= \xi - \psi[D_I(q) - D_I(QP_I)] \\ \text{otherwise } D_p(QP_I, q) &= \xi \end{aligned} \quad (4.3)$$

where ξ and ψ are the model parameter and $q = QP_P$.

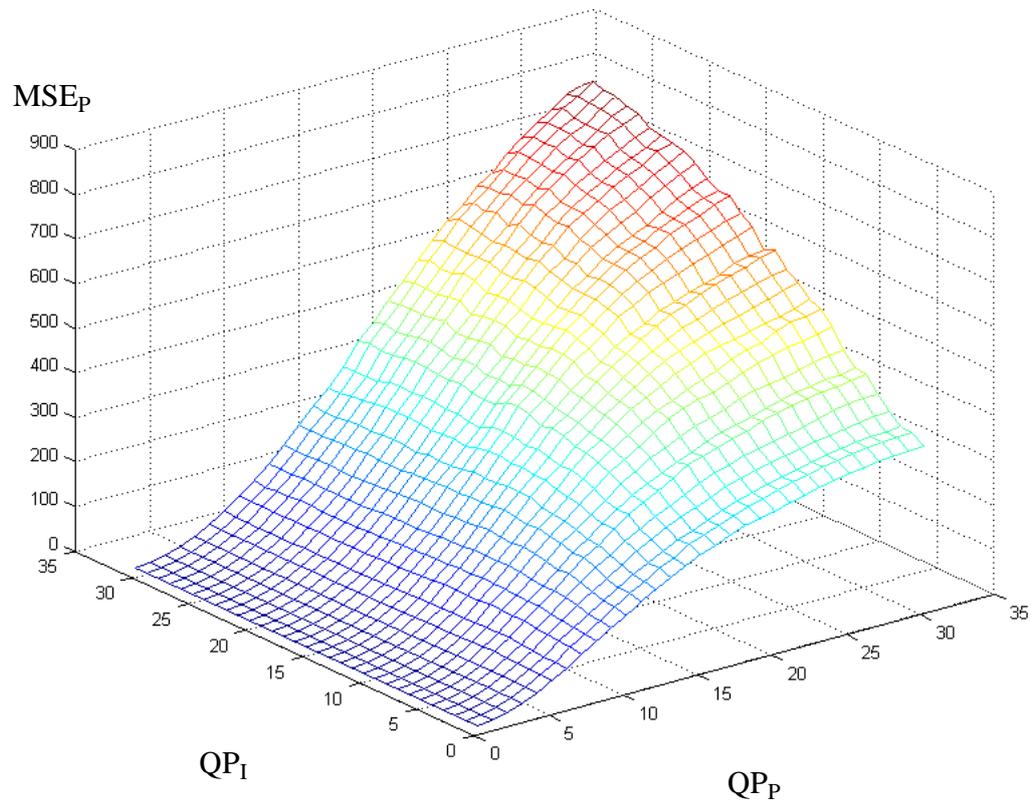


Figure 4-2 The 3-D plane of the R-D curve.

ξ and ψ can be solved by encoding and measuring the distortion at two values of QP_I . If the spline model uses 6 control points, then 12 control points are necessary for the inter-frame dependency model. An inter-frame model can be represented by a 2D space, as shown in Figure 4-3. Note that both the spline interpolation and the

linear-constant model can be used to find any R-D points. Let us use the example in [67], the two values for finding ξ and ψ are 5 and 13, as shown in Figure 4-3. If the point to be considered is the arbitrary point $(Q_P, Q_I) = (10, 10)$, then use the linear-constant model to find the four points $(10, 5)$, $(10, 8)$, $(10, 13)$, $(10, 21)$ and these four points are used in (4.1) to determine $(10, 10)$. The total number of control points needed is 8 for $(10, 10)$, but to cover all the 2D space, 12 points are required.

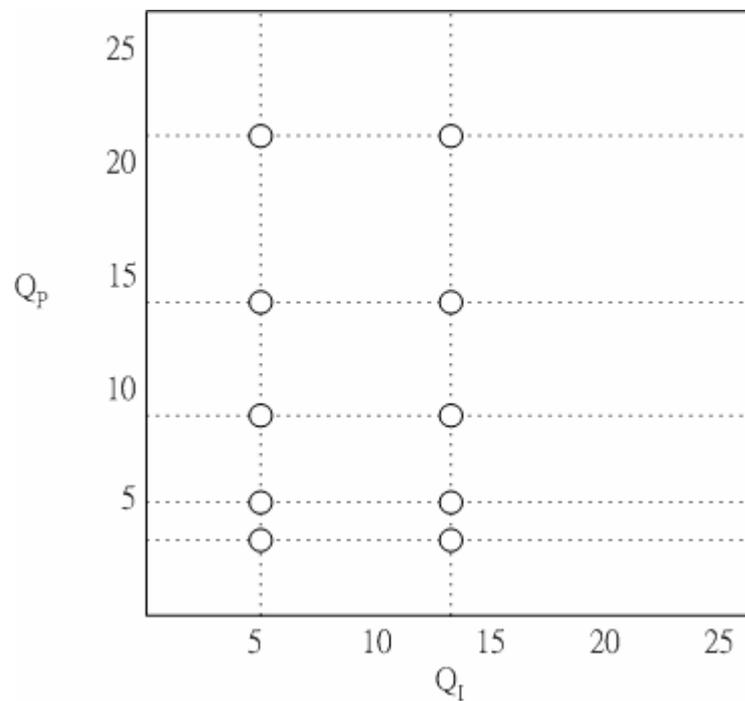


Figure 4-3 The 2-D plane showing the control points used to approximate the whole dependency 3-D curve.

4.3 Fast approximation and modeling

A faster approximation method is devised in this thesis, which requires only half of the computation of [67]. Our method adaptively selects the number of control points instead of using a global choice.

4.3.1 Fast Intra-frame approximation

In most of the existing modeling rate-control schemes, the modeling functions are usually assumed to be a linear distortion function. However, in most situations, this assumption is not true. The linear characteristic happens only when the scale parameter is sufficiently large, for example, higher than 15. Figure 4-4, Figure 4-5 and Figure 4-6 show the R-D curves of the sequences: Football, Mobile and Garden, respectively. When the scaling factors are greater than 15, the R-D curves tend to be straight lines. Based on this characteristic, we can use a bisection algorithm, which is adaptive to the individual sequence. Our algorithm includes only those necessary R-D points. If the R-D of a sequence is a straight line, this algorithm requires 2 points only to approximate the R-D curve. The details of the algorithm are given as follows,

Set the distortion, denoted as MSE_I^1 , at $QP_I = 0$ to be 0.

Encode the I-frame with $QP_I = 31$ and set $QP_I^{temp} = 31$. Compute the distortion MSE_I^{31} . Then use spline interpolation to approximate the R-D curve using two points.

Encode the I-frame with $QP_I = 16$, which is derived by $\lfloor QP_I^{temp} / 2 \rfloor$. Compute MSE_I^{16} and compare the value of MSE_I^{16} from the approximated R-D curve to the actual MSE_I^{16} . If the error is within a preset threshold, then form the final approximated curve by using the three points (QP_I^1, MSE_I^1) , (QP_I^{16}, MSE_I^{16}) , (QP_I^{31}, MSE_I^{31}) . Otherwise, set $QP_I^{temp} = 16$ and encode the point $\lfloor QP_I^{temp} / 2 \rfloor$ until the threshold is met.

Note that the point (QP_I^1, MSE_I^1) is not a result of encoding the point. Therefore, at most, six points are used, and the points chosen are (31, 16, 8, 4, 2, 1). However, in most of the cases, 4 points are sufficient. Table 1 shows the experimental results. The relative error is determined as follows:

$$\text{relative error} = \left| \frac{\text{original value} - \text{estimated value}}{\text{original value}} \right| \quad (4.4)$$

Table 4-1 Relative modeling error of our distortion modeling of the I-frame.

	Average error	Number of points
Garden	2.15%	4
Football	2.6%	5
Mobile	2.2%	4

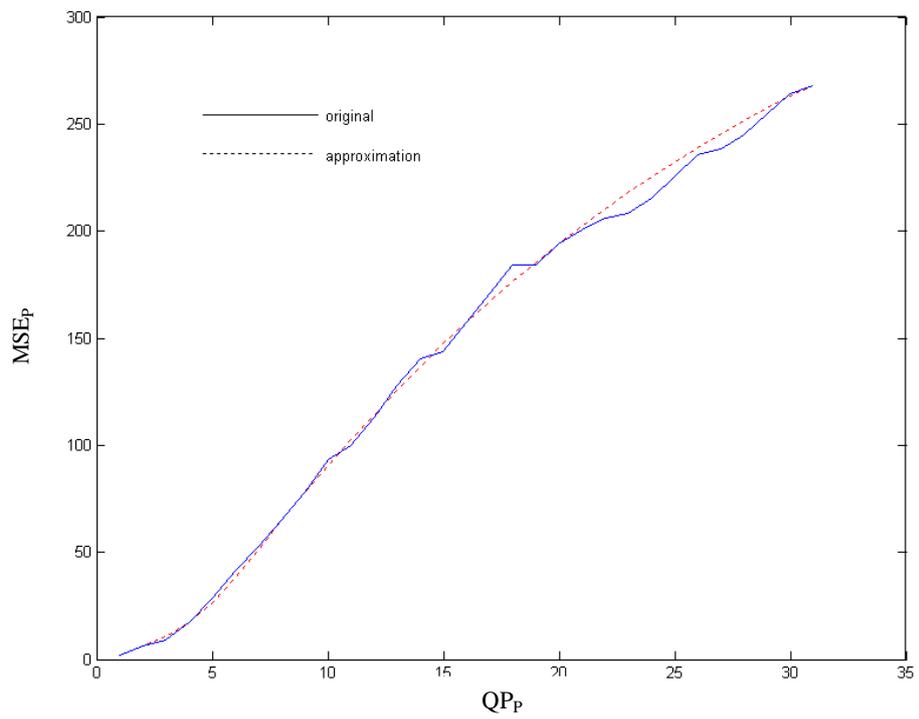


Figure 4-4 The distortion curve of the sequence Football.

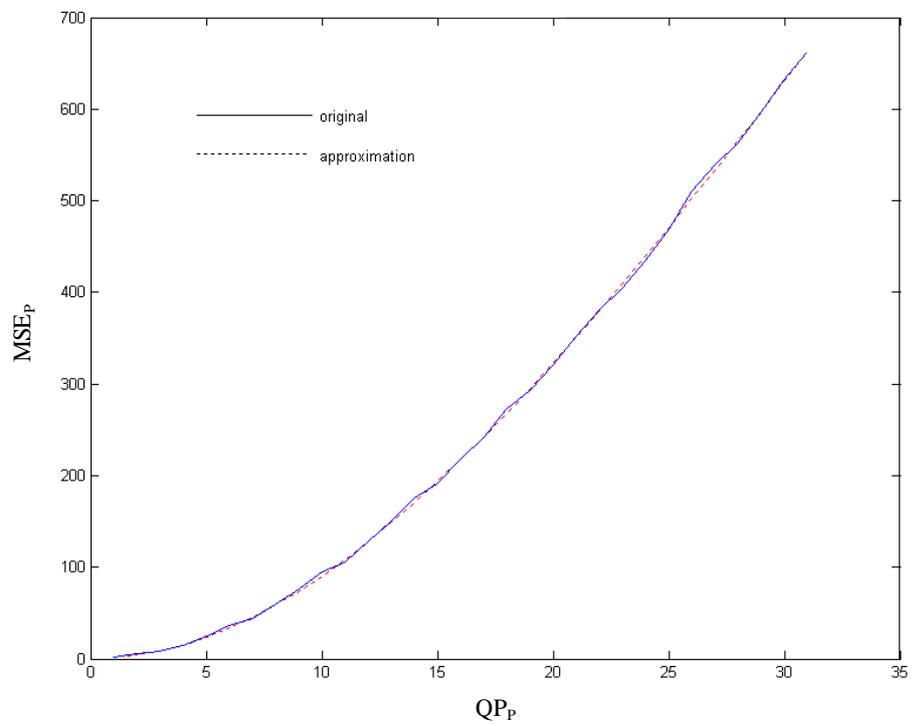


Figure 4-5 The distortion curve of the sequence Mobile.

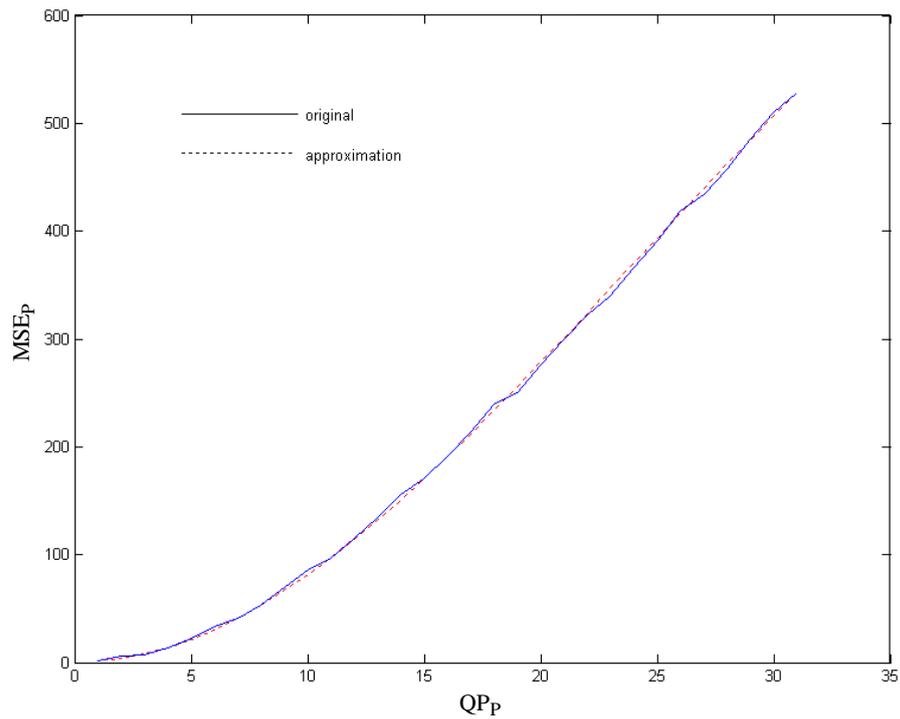


Figure 4-6 The distortion curve of the sequence Garden.

4.3.2 Fast Inter-frame dependency model

A fast inter-frame dependency model based on the linear-constant model proposed in [67] will be developed in this section. However, our technique is quite different to the previous method as it does not require spline interpolation for the predicted frame. Figure 4-7 shows that the distortion does not increase further after the line $QP_I = QP_P$. Therefore, this line can be used to approximate the whole R-D plot with all combinations of the scaling factors of the I-frame and the P-frame. The first step to determine this dependency model is to calculate the MSE for $QP_I = 1$

and for all QP_P , i.e. the circles in Figure 4-8. By projecting the QP_I values (1, 2, ..., 31) from the x -axis up to the line satisfying $QP_I = QP_P$ as shown in Figure 4-9, the point marked with a triangle can be found. The whole R-D plot can then be found by connecting the circled points and the triangle-marked points. Note that the circle points are found by using spline interpolation, as these points are on the R-D curve of the P-frame at $QP_I = 1$. See Figure 4-10, it is clear to show the relationship by a 3D space.

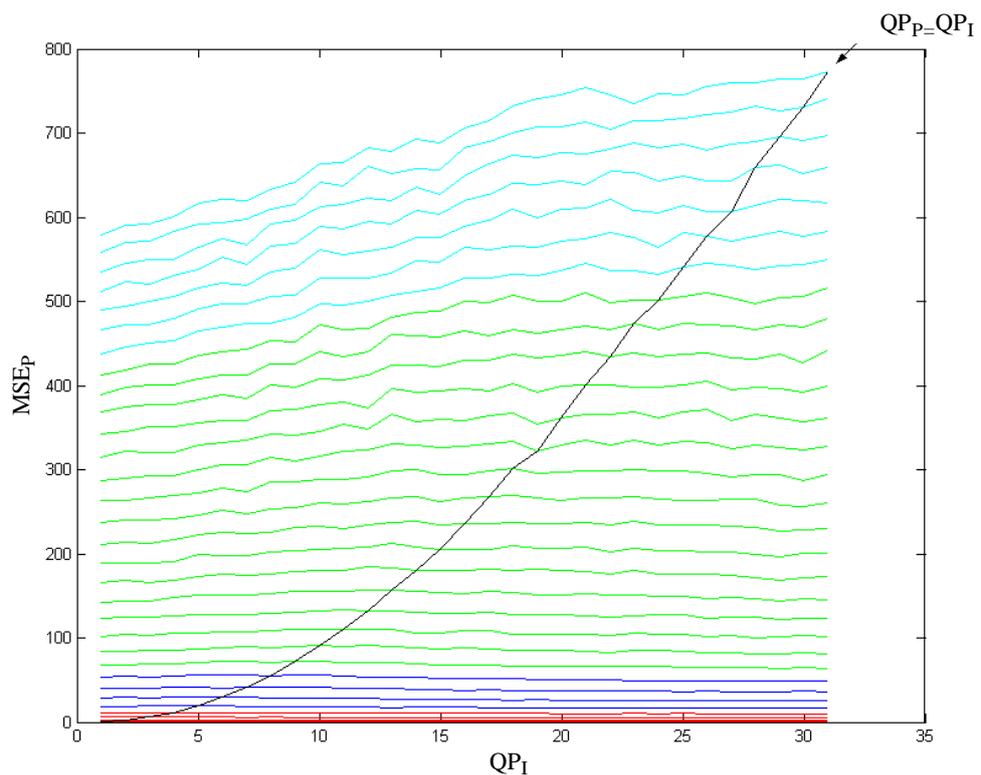


Figure 4-7 The dependency R-D curve.

The technique used to model the line satisfying $QP_I = QP_P$ is derived in this section. Based on the model shown in Figure 1-1, let F_p , R_p , R_p^Q , EL_p and QP_p , be the P-frame, the residue, the quantized residue, the enhancement layer (i.e. the quantization distortion) and the scaling factor of the P-frame, respectively. We also assume that an I-frame has a similar set of notations, that is, F_I , R_I , R_I^Q , EL_I and QP_I . As the residue and the enhancement layers are affected by the scaling factors, we will further represent the parameters by $R_p(QP_I, QP_P)$, $R_p^Q(QP_I, QP_P)$, $EL_p(QP_I, QP_P)$ and $R_I(QP_I)$, $R_I^Q(QP_I)$, $EL_I(QP_I)$. The P-frame can be formulated as follows,

$$F_p = MC(F_I - EL_I) + R_p \quad (4.5)$$

where MC means motion compensation. Then, we have

$$\begin{aligned} F_p &= MC[F_I - EL_I(QP_I, QP_P)] + R_p \\ &= MC(F_I) - MC[EL_I(QP_I, QP_P)] + R_p \\ &= MC(F_I) - MC[EL_I(QP_I, QP_P)] + R_p^Q(QP_I, QP_P) + EL_p(QP_I, QP_P) \\ &= MC(F_I) - MC[EL_I(QP_I)] + R_p^Q(QP_I) + EL_p(QP_I), \text{ by setting } QP_I = QP_P \\ EL_p(QP_I) &= MC[EL_I(QP_I)] - R_p^Q(QP_I) + F_p - MC(F_I) \\ EL_p(QP_I) &= MC[EL_I(QP_I)] - \nu \bullet R_p^Q(QP_I) + [F_p - MC(F_I)] \end{aligned} \quad (4.6)$$

where ν is a matrix and \bullet represents the dot product operation.

$$EL_p(QP_1) = MC[EL_l(QP_1)] - \nu \bullet R_p^o(QP_1) + \tau \quad (4.7)$$

where τ is a constant matrix equal to $[F_p - MC(F_l)]$. If the motion between the I-frame and the P-frame is very small, then $MC[EL_l(QP_1)] \approx EL_l(QP_1)$ and τ becomes a zero matrix. We therefore have,

$$EL_p(QP_1) \approx EL_l(QP_1) - \nu \bullet R_p^o(QP_1) \quad (4.8)$$

If $\nu \bullet R_p^o(QP_1)$ is ignored, we will have,

$$EL_p(QP_1) \approx EL_l(QP_1) \quad (4.9)$$

As the enhancement layer can actually reflect the distortion, we can use MSE to represent the distortion. Therefore,

$$MSE[EL_p(QP_1)] \approx MSE[EL_l(QP_1)] \quad (4.10)$$

It is interesting to note that $MSE[EL_p(QP_1)]$ is the line satisfying $QP_l = QP_p$, as shown in Figure 4-7, and $MSE[EL_l(QP_1)]$ is actually the R-D curve of the encoded I-frame. The difference between (4.7) and (4.9) is $[F_p - MC(F_l)]$ and $\nu \bullet EL_l(QP_1)$.

Of course, if these items are available, the exact result can be found. However, obtaining these items - which are actually the motion compensation and quantization process for the enhancement layer, respectively - requires many computations.

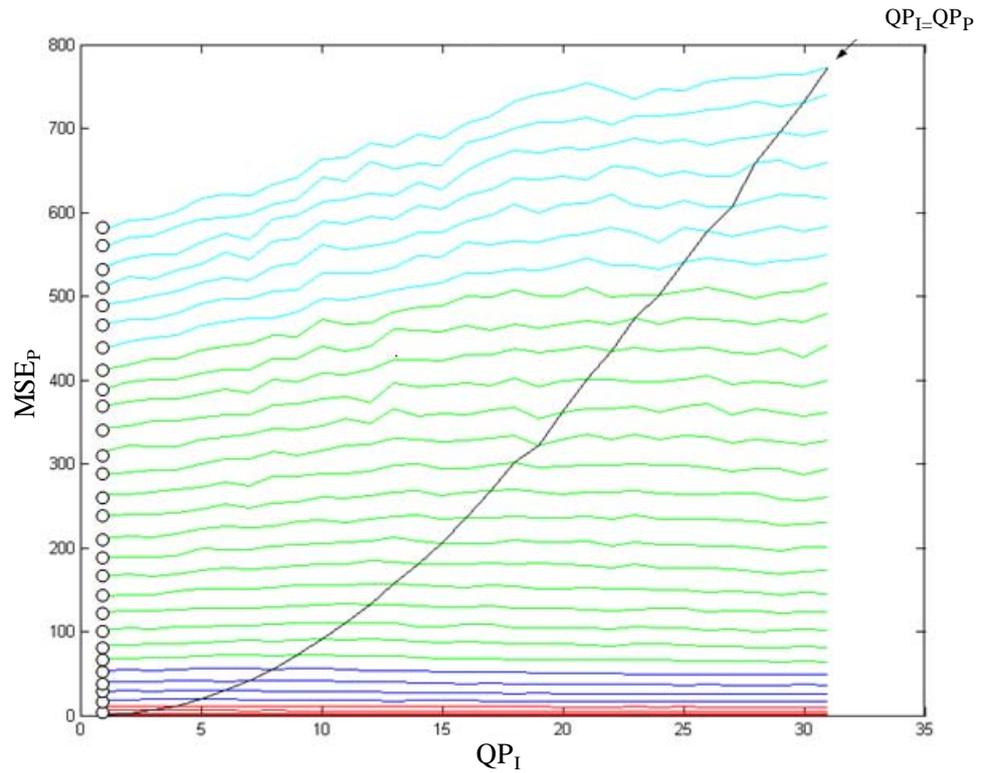
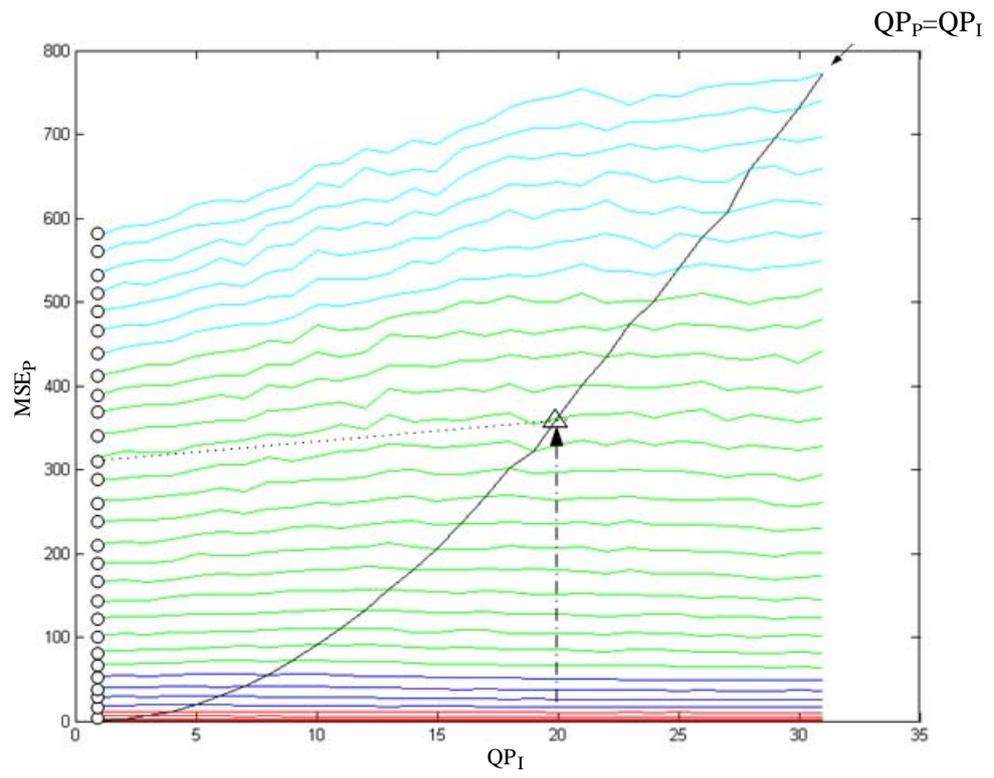


Figure 4-8 The dependency R-D curve showing the points needed to be approximated with circles.

Table 4-2 tabulates the experimental results of our algorithm. [67] generally requires 12 points for inter-frame modeling while our algorithm needs generally 5 points.

Table 4-2 Relative modeling error of our distortion modeling scheme for P-frames.

	Average error	Number of points
Garden	4.2%	5
Football	6.3%	5
Mobile	7.8%	5

Figure 4-9 Projection from the QP_1 axis.

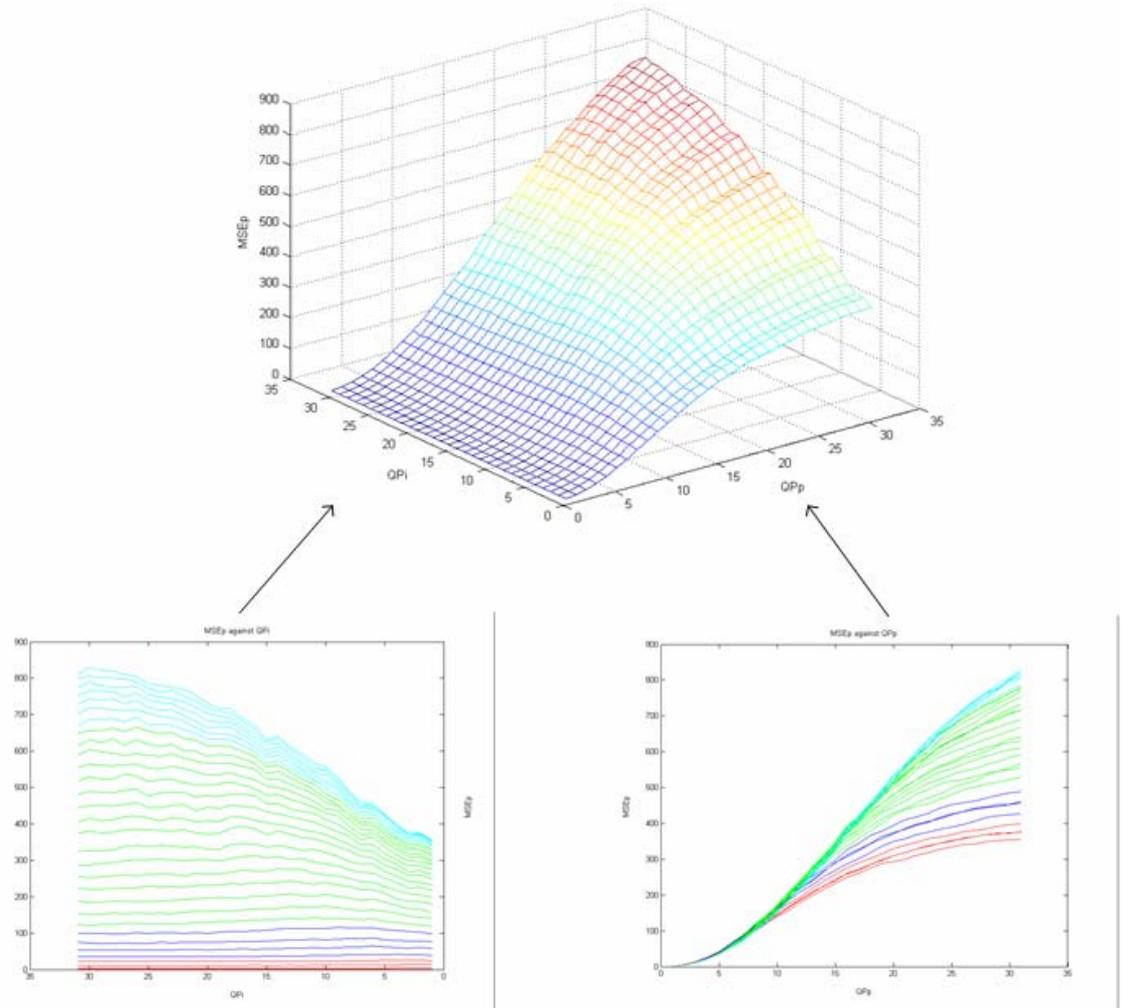


Figure 4-10 The $Distortion_P-QP_I$ curve and the $Distortion_P-QP_P$ curve obtained from different points of view of the 3-D R-D dependency curve.

4.4 Conclusion

We have developed a model for an intra-coded frame and a model for an inter-coded frame. Both models can achieve a comparable performance to the original spline model. Experiments show that the errors based on our models increase slightly, by 1.12%, compared to those using the spline model on average, but the

computation required is reduced to only 45% of the spline model. This can significantly reduce the processing time for this kind of pre-analysis rate-control technique. However, our system does increase the modeling error slightly. Hence, the model to be employed should depend on the application itself. The spline model can be employed for media with a tight storage capacity, while our model is suitable for non-real-time applications in which the time constraint is the greatest concern.

Chapter 5

Conclusion

5.1 Conclusion on the research project

Rate control is one of the most important processes in the encoder side to control the trade-off between the quality and the output bit rate. In this research project, different rate-control techniques have been investigated, including most of the typical rate-distortion modeling methods that have been adopted as the reference models for various existing video compression standards, non-real-time pre-analysis methods, and low-computation fuzzy logic rate controller, etc. In JM8.1a and VM5.0, the rate-distortion model is based on the first two terms of the Taylor expansion. The performance of the quadratic model is quite robust against different sequences and under different channel resources. As a consequence, the quantization step is obtained by solving the quadratic model for a given channel rate. Increasing the order of the model becomes very difficult since the quantization step cannot be easily solved for a rate-distortion model with an order higher than two. The determination of higher order rate-distortion models requires iterative numerical

methods. Therefore, we have developed a model that represents the quantization step as a function of the target bit rate. In other words, this model is a distortion-rate model rather than a classical rate-distortion model. However, since the developed cubic distortion-rate model itself is not derived from the expansion of the Taylor series, such a model may not be monotonically decreasing when the bit rate increases. Hence we have also developed an effective method that can determine the region in which the cubic model is not monotonically decreasing such that the system rejects the results determined in that region.

Increasing the order of the model will certainly increase the computational effort in the linear regression process of the model parameters. Our rate-control system includes a pseudo-regression method. This method allows the system to use a simpler form of regression. Classical regression is only employed when pseudo-regression fails.

Furthermore, we observe that in some cases the quantization steps always switch between two choices. This problem is caused by overestimation or underestimation to the quantization step of the previous coded units. When overestimation or underestimation has occurred, the system tends to remedy the problem by decreasing or increasing the following quantization step by two, which is the maximum possible change, as indicated in the joint model. Such a remedy cannot

always solve the problem, as it may also overestimate or underestimate again. As a result, the quantization step is oscillating within a certain period. This oscillation surely makes the quality fluctuate as well. In our rate-control system, we include a simple technique to stabilize the final quantization step and the quantization parameter. Our algorithm detects the fluctuation of the quantization steps and eliminates the range of the fluctuation. Experimental results show that our rate-control techniques outperform the classical rate-distortion model in terms of output bit rate and objective quality. Nevertheless, our model can be further extended in future development, as it is not necessary to solve the problem described above in order to determine the quantization step.

We have also studied some pre-analysis rate-control techniques and developed two new fast models for the system that originally used the spline cubic model. Limited by the original design of the spline cubic model, the number of control points is fixed to be 12. Reducing the number of control points is difficult, as the spline model requires a certain number of sampling points in order to interpolate the values between any two sampled control points. Another problem is that the original model does not consider the complexity of the rate and distortion curves. Actually, some curves can be modeled easily since the curves behave like linear lines in some parts of the curves. One of our new models can adaptively select the number

of control points based on different behaviors of the curves for intra-frames. Our experiments show that five points are sufficient to model a rate or a distortion curve. In some cases, only four points are necessary. Besides, we have also developed an inter-model which improves the performance of the system for inter-frames. This model is not adaptive. Instead, it requires a fixed number of points for different sequences. These two models together can significantly reduce the overall computational effort.

This research project has covered two of the most popular streams of rate-control techniques. We can understand the state-of-the-art of rate control clearly through the development of these two new rate-control techniques, and we have solved some of the problems in the existing rate-control techniques.

5.2 Future development of rate-control techniques

Several new rate-control techniques will be proposed every year as it is an important part of video coding and transmission. Some researchers employ the derivation from the information theory to develop a new model. Others may propose techniques that are fine-tuned versions of previous research for a new video standard. There is also some work based on new domains in which the rate and distortion characteristics behave linear so that the prediction of the curves become much easier.

As the number of handheld mobile video communication devices increases, the future development of rate control will tend to have low computational effort and low power consumption. Instances are the emergence of the third-generation mobile phones, smartphones and handheld personal computers. All these devices will support multimedia applications such as video communication in the future. The capability of video communication in handheld mobile devices is much more challenging than in those devices with continuous power supply. Therefore, the new generation of rate-control techniques should not require too many supplementary computations, as the video coding itself is already quite power-consuming. The pre-analysis rate control technique should be no longer possible in mobile devices, while the rate-distortion model and the fuzzy logic rate controller can be improved to suit those mobile applications with limited resources available.

Possible improvement can be further developed based on the work in this research project. For instance, the combination of the individual terms in our cubic model may be modified to include more terms with different shapes, which can represent the curve more effectively and efficiently with a smaller order. On the other hand, although linear regression is a very good tool to determine the model parameters based on previous coded units, it is somehow still quite power-consuming.

We believe that there exist even better methods to determine the model parameters more accurately with lower computations.

There have not been many fuzzy logic controllers in the past few years. One of the difficulties in developing such a rate-control technique is the definition of the membership function, which may include some iterative training processes that are not practical for real-time applications. Some existing fuzzy logic controllers include the neuro-network to train the system and to select the optimal parameters. All of these are quite power-consuming. In the future, it is necessary to be more concise and to obtain simpler parameters in order to continuously develop fuzzy rate control.



References

- [1] ITU-T, "Codecs for videoconferencing using primary digital group transmission (H.120 Recommendation)," 1988.
- [2] A. M. Tekalp, *Digital video processing*. Upper Saddle River, N.J.: Prentice Hall PTR, 1995.
- [3] ITU-T, "Video codec for audiovisual services at p x 64 kbit/s (H.261 Recommendation)," 1990.
- [4] D. LeGall, "MPEG: A video compression standard fro multimedia applications," *Commun. ACM*, vol. 34, pp. pp. 46-58, 1991.
- [5] ISO/IEC, "Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 2: Video, ISO/IEC 11172-2 (MPEG-1)," 1993.
- [6] B. G. Haskell, A. Puri, and A. N. Netravalli, *Digital Video: An introduction to MPEG-2*. New York, USA: Chapman and Hall, 1997.
- [7] ISO/IEC, "Information technology -- Generic coding of moving pictures and associated audio information: Video, ISO/IEC 13818-2 (MPEG-2)," 2000.
- [8] ITU-T, "Video coding for low bit rate communication (H.263 Recommendation)," 1998.
- [9] ISO/IEC, "Information technology -- Coding of audio-visual objects -- Part 2: Visual, ISO/IEC 14496-2 (MPEG-4)," 2004.
- [10] "Joint Video Team of ISO/IEC MPEG & ITU-T VCEG," Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)/JVT-G050, 2003.
- [11] G. J. Sullivan, "Joint Model Reference Encoding Methods and Decoding Concealment Methods," Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), San Diego, USA JVT-I049, 9 Sep 2003.
- [12] ISO/IEC, "Information technology -- Coding of audio-visual objects -- Part 10: Advanced Video Coding, ISO/IEC 14496-10," 2004.

-
- [13] S. J. Orfanidis, *Introduction to signal processing*. Upper Saddle River, N.J.: Prentice Hall, 1996.
- [14] A. Hallapuro and M. Karczewicz, "Low Complexity Transform and Quantization - Part II: Extensions," Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) JVT-B039, Jan 14 2002.
- [15] A. Hallapuro and M. Karczewicz, "Low Complexity Transform and Quantization - Part I: Basic Implementation," Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) JVT-B038, Jan 14 2002.
- [16] W.-K. Cham, "Development of integer cosine transforms by the principle of dyadic symmetry," *IEE Proceedings Communications, Speech and Vision*, vol. 136, pp. 276-282, 1989.
- [17] H. Song, J. kim, and C.-C. Jay Kuo, "A sliding window approach to real-time H.263+ frame rate adjustment," *Signals, Systems & Computers, Conference Record of the Thirty-Second Asilomar Conference on*, 1998.
- [18] H. Song, J. Kim, and C.-C. Jay Kuo, "Improved H.263+ rate control via variable frame rate adjustment and hybrid I-frame coding," *Image Processing, 1998. Proceedings. 1998 International Conference on*, 1998.
- [19] H. Song, J. Kim, and C.-C. Jay Kuo, "Real-time H.263+ frame rate control for low bit rate VBR video," *Circuits and Systems, Proceedings of the 1999 IEEE International Symposium on*, 1999.
- [20] E. C. Reed and F. Dufaux, "Constrained bit-rate control for very low bit-rate streaming-video applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 882-889, 2001.
- [21] E. C. Reed and J. S. Lim, "Optimal multidimensional bit-rate control for video communication," *IEEE Transactions on Image Processing*, vol. 11, pp. 873-885, 2002.
- [22] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74-90, 1998.
- [23] X. Chen, *Transporting Compressed Digital Video*. San Diego, CA: Kluwer Academic Publishers, 2002.
-

-
- [24] W. Ding, "Joint encoder and channel rate control of VBR video over ATM networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 266-278, 1997.
- [25] T. Wiegand, G. J. Sullivan, G. Bjntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560-576, 2003.
- [26] R. Schafer, T. Wiegand, and H. Schwarz, "The emerging H.264/AVC standard," Jan 2003.
- [27] G. J. Sullivan and T. Topiwala, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity range Extensions," SPIE Conference on Applications of Digital Image Processing XXVII Special Session on Advances in the Now Emerging Standard: H.264/AVC., 2004.
- [28] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 688-703, 2003.
- [29] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 620-636, 2003.
- [30] J. Ribas-Corbera and D. L. Neuhoff, "On the optimal motion vector accuracy for block-based motion-compensated video coders," Proc. SPIE Dig. Video Compr., San Jose, CA, 1996.
- [31] M. Wien, "Variable block-size transforms for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 604-613, 2003.
- [32] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 598-603, 2003.
- [33] P. List, A. Joch, J. Lainema, G. Bjntegaard, and M. Karczewicz, "Adaptive deblocking filter," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 614-619, 2003.
-

-
- [34] X. Sun, S. Li, F. Wu, J. Shen, and W. Gao, "The improved SP frame coding technique for the JVT standard," *Image Processing, 2003. Proceedings. 2003 International Conference on*, 2003.
- [35] M. Karczewicz, R. Kurceren, and F. Wu, "A Proposal for SP-frames," *Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) VCEG-L27*, 4 Jan 2001.
- [36] M. Karczewicz and R. Kurceren, "The SP- and SI-frames design for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 637-644, 2003.
- [37] R. Kurceren and M. Karczewicz, "Synchronization-predictive coding for video compression: the SP frames design for JVT/H.26L," *Proceedings, International Conference on Image Processing*, 2002.
- [38] S. Wenger, "H.264/AVC over IP," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 645-656, 2003.
- [39] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley, 1991.
- [40] N. S. Jayant and P. Noll, *Digital coding of waveforms : principles and applications to speech and video*. Englewood Cliffs, N.J.: Prentice-Hall, 1984.
- [41] R. Reininger and J. Gibson, "Distributions of the Two-Dimensional DCT Coefficients for Images," *IEEE Transactions on Communications*, vol. 31, pp. 835-839, 1983.
- [42] F. Bellifemine, A. Capellino, A. Chimienti, R. Picco, and R. Ponti, "Statistical analysis of the 2D-DCT coefficients of the differential signal for images," *Signal Processing: Image Communication*, vol. 4, pp. 477-488, 1992.
- [43] S. R. Smoot and L. A. Rowe, "Study of DCT coefficient distributions," *Proceedings of the SPIE Symposium on Electronic Imaging*, volume 2657, San Jose, CA, 1996.
- [44] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, pp. 1661-1666, 2000.
- [45] A. J. Viterbi and J. K. Omura, *Principles of digital communication and coding*. New York: McGraw-Hill, 1979.
-

-
- [46] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, pp. 12-20, 1996.
- [47] H.-M. Hang and J.-J. Chen, "Source model for transform video coder and its application. I. Fundamental theory," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 287-298, 1997.
- [48] J.-J. Chen and H.-M. Hang, "Source model for transform video coder and its application. II. Variable frame rate coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 299-311, 1997.
- [49] C. Grecos and J. Jiang, "On-line improvements of the rate-distortion performance in MPEG-2 rate control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 519-528, 2003.
- [50] G. Keesman, I. Shah, and R. Klein-Gunnewiek, "Bit-rate control for MPEG encoders," *Signal Process.: Image Commun.*, vol. 6, pp. 545-560, 1995.
- [51] W. Ding and B. Liu, "Rate-quantization modeling for rate control of MPEG video coding and recording," Proc, IS&T/SPIE Digital Video Coding and Recording, San Jose, CA, 1995.
- [52] T. M. E. Committee, "MPEG-2, Test Model 5 (TM5) Doc. ISO/IEC JTC1/SC29/WG11/93-225b," 1993.
- [53] H. Everett, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operations Research*, vol. 11, pp. 399-417, 1963.
- [54] Y. Shoham and A. Gersho, "Efficient bit allocation for an arbitrary set of quantizers," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 1445-1453, 1988.
- [55] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 23-50, 1998.
- [56] G. M. Schuster and A. K. Katsaggelos, Rate-distortion based video compression : optimal video frame compression and object boundary encoding, 1 ed. Boston: Kluwer Academic Publishers, 1997.
- [57] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Transactions on Image Processing*, vol. 3, pp. 26-40, 1994.
-

-
- [58] R. Fletcher, *Practical Methods of Optimization*, 2 ed: Hohn Wiley & Sons, 1987.
- [59] J. Ribas-Corbera, P. A. Chou, and S. L. Regunathan, "A generalized hypothetical reference decoder for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 674-687, 2003.
- [60] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 172-185, 1999.
- [61] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, pp. 246-250, 1997.
- [62] H.-J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, pp. 878-894, 2000.
- [63] S. Ma, Z. Li, and F. Wu, "Proposed Draft of Adaptive Rate Control," Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6) JVT-H017, 20-26 May 2003.
- [64] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ -domain source modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 928-940, 2001.
- [65] Z. He and S. K. Mitra, "A linear source model and a unified rate control algorithm for DCT video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 970-982, 2002.
- [66] L.-J. Lin, "Video bit-rate control with spline approximated rate-distortion characteristics," Univ. Southern California, 1997.
- [67] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 446-459, 1998.
- [68] "Joint Model 8.1a," Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), 2004.
- [69] N. Wang and Y. He, "A new bit rate control strategy for H.264," Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on
-

Multimedia. Proceedings of the Joint Conference of the Fourth International Conference on, 2003.

[70] C. W. De Silva, *Intelligent control : fuzzy logic applications*. Boca Raton: CRC Press, 1995.

[71] R. Fuller, *Introduction to neuro-fuzzy systems*. Heidelberg ; New York: Physica-Verlag, 2000.

[72] D. H. K. Tsang, B. Bensaou, and S. T. C. Lam, "Fuzzy-based rate control for real-time MPEG video," *IEEE Transactions on Fuzzy Systems*, vol. 6, pp. 504-516, 1998.

[73] Y.-S. Saw, P. M. Grant, and J. M. Hannah, "Quality-optimised MPEG2 video data rate control using fuzzy logic techniques," *IEE Proceedings-Vision, Image and Signal Processing*, vol. 145, pp. 179-186, 1998.

[74] A. Leone, A. Bellini, and R. Guerrieri, "Fuzzy-controlled perceptual coding of videophone sequences," *IEEE Transactions on Fuzzy Systems*, vol. 5, pp. 294-303, 1997.

[75] A. Leone, A. Bellini, and R. Guerrieri, "An H.261-compatible fuzzy-controlled coder for videophone sequences," *Fuzzy Systems, IEEE World Congress on Computational Intelligence., Proceedings of the Third IEEE Conference on*, 1994.

[76] M.-J. Chen, M.-C. Chi, C.-T. Hsu, and J.-W. Chen, "ROI video coding based on H.263+ with robust skin-color detection technique," *IEEE Transactions on Consumer Electronics*, vol. 49, pp. 724-730, 2003.

[77] N. Doulamis, A. Doulamis, D. Kalogeras, and S. Kollias, "Low bit-rate coding of image sequences using adaptive regions of interest," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, pp. 928-934, 1998.

[78] H. Song and C.-C. J. Kuo, "A region-based H.263+ codec and its rate control for low VBR video," *IEEE Transactions on Multimedia*, vol. 6, pp. 489-500, 2004.

[79] B. L. Bowerman and R. T. O'Connell, *Linear statistical models : an applied approach*, 2nd ed. Boston: PWS-Kent, 1990.

[80] S. Brandt, "Data analysis : statistical and computational methods for scientists and engineers," 3rd ed. New York: Springer, 1999.