# 基於位元-失真理論之 H.264/AVC 位元控制研究

研究生：許益賓　　指導教授：陳福川 博士

國立交通大學　　電控工程研究所

## 摘　　要

位元控制在訊號處理上扮演相當重要地位，對於視訊壓縮的應用更是重要。其中，位元控制有兩個取捨點作為相互影響的因子，一為資料品質，也就是失真度，另一為資料量。在視訊壓縮系統上，共定義出兩種不同的影像壓縮型別，分別是 Intra-Frame 與 Inter-Frame 形式。 Intra-Frame 扮演整個影像壓縮的起始點，除了要保持高度清晰的影像品質外，在位元的使用上更需精確應用，主因為接下來的影像需參考此一影像並使用剩下的位元進行壓縮。為了增加壓縮效能，或是減低位元需求，常會加入大量的 Inter-Frame 形式的影像。在即時的視訊壓縮系統中，藉由前一張影像的資料品質與資料量來快速預測量化係數是位元控制所扮演的腳色。然而，真實的影像資料是會隨物體以及時間產生變化，往往兩張畫面就有可能產生極大的落差。因此，如何藉由前一張影像的資料品質與資料量來預測出精確的量化係數，因而產生穩定的影像品質或是使用精確的位元數成為重要的研究課題。

本研究以視訊壓縮 H.264/AVC 為主，架構於原始的位元-失真理論，而此基準點也是目前 H.264/AVC 在位元控制所採用的方法，探討其現階段位元控制的法則，得到一個可行性的失真度預測方法，進而將之應用於 Intra-Frame 與 Inter-Frame 的兩個層面。更近一步的討論，基於這兩個層面，在不同的應用考量上，共推導出四個不同的位元控制演算法。

- 基於 Intra-Frame 的位元控制演算法。
  1. 利用所推導得到的失真度預測方法，藉由常見的標準測試影像進行參數訓練，進而得到快速且精確的 Intra-Frame 位元控制演算法。
- 基於 Inter-Frame 的位元控制演算法。
  1. 利用所推導得到的失真度預測方法與位元估測方程式，透過 Lagrange 最佳化的推導，提出一個解析性的精確位元

控制演算法，改善原本 H.264/AVC 以及現有著作在低位元率限制下表現不佳的缺點。

2. 在實務應用上，優先考量嵌入式系統的適用性。藉由 H.264/AVC 內建的量化係數預測技術，得到初始的量化係數，利用所推導得到的失真度預測方法以及影像直方統計曲線，進行量化參數修正。
3. 藉由所推導出的方程式來有效的預測影像複雜度，以達到較先前研究更精確量化參數數值。目的在於解決低位元率的視訊傳輸之應用。

由原始的位元與失真理論，推導出失真度預測方法。為驗證此一方法其效能的適用性，本文基於四種重要且常見的應用，將此一方法加入，藉由實驗結果指出，此一方法確實能有效運用位元來提升影像壓縮品質。此外亦有將嵌入式系統的硬體限制考慮，讓此方法更能貼近產業界應用。

# The Study of Rate-Distortion Theorem for H.264/AVC Rate Control

Student：Yi-Pin Hsu Advisor：Dr. Fu-Chuang Chen

Institute of Electrical Control Engineering

National Chiao Tung University

## Abstract

Rate control plays a vital role in signal processing and video compression. Two degrees of distortion and consumed counts for measurement are selected in a tradeoff consideration. The video compression system defines two distinct frame types, as follows: intra-frame and inter-frame. The start frame in encoded video sequences is assigned as an intra-frame, and the goal is to maintain a higher image quality and lower encoded bits because the intra-frame is referenced after the encoded image, and the remaining bits are used for compression. In a real-time video compression system, referred to as H.264/AVC, the determination of the quantization parameter by rate control must refer to the encoded results in the previous frame. However, sudden occurrences such as fast-moving objects or complex backgrounds in two neighboring frames are usual. Thus, because of encoded results in previous images, producing stable image quality or accurate bits consumption is a crucial topic.

This study focused on the H.264/AVC video compression system and rate-distortion theory to obtain a workable distortion prediction method for intra-frame and inter-frame rate control. Subsequently, we used this distortion prediction method to develop four rate control algorithms for various applications.

- Based on the intra-frame rate control algorithm
  1. According to the distortion prediction method and a large number of test images, we obtained a set of system parameters to develop a fast and accurate intra-frame rate control algorithm.

- Based on the inter-frame rate control algorithm
  1. We propose an analytic inter-frame rate control algorithm according to the distortion prediction method, H.264/AVC rate prediction method,

and Lagrange optimization. This algorithm can solve the problem of low bit rate limitation in H.264/AVC.

2. In real applications, we first considered the limitation of embedded systems. The initial QP value can be obtained through the built-in technology of QP calculation. We developed the intra-frame rate control algorithm according to the initial QP value, the distortion prediction method, and the information of image histogram.

3. In the proposed distortion prediction method, we can predict the measurement of image content complexity efficiently to obtain a more accurate QP value than by using the current algorithm. This algorithm was developed for low bit rate video conferencing.

We derived the distortion prediction method according to the original rate-distortion theorem. We applied the method on four differing applications to verify the results. Based on the simulation results, the method improved compressed image quality efficiently. We also considered hardware limitations for industrial application.

# 致 謝

本論文的完成首先要感謝我的指導老師—陳福川教授。在教授的研究風格中，讓學生深深體會到，研究過程中所應具備的態度與方法。先人曾經說過：「給我一個支點，我可以舉起整個地球」；而老師就是給我這個支點的重要推手。每當投稿初審回來之後，老師總是用嚴謹的態度面對每一位審查委員的意見，希望可以將意見完整修正、回覆，也因此讓學生得以成功完成好幾篇的國際論文。其中，一篇 IEEE 期刊，在超過一年的往返修改時程，學生幾乎要放棄，然而老師卻是鼓勵學生要繼續奮鬥下去，終於讓我獲得人生中的第一篇 IEEE 期刊。

在論文口試中，要感謝當天專程撥空前來的六位口試委員。分別是林銀議教授、戴顯權教授、王周珍教授、洪士程教授、蔡文錦教授以及張文鐘教授。有了各位教授的專業指點，分別指出學生在不同細節所忽略到的問題，進而增加學生在論文呈現的完整性以及指導未來可以繼續研究的方向。

二十幾年的求學生涯，包含六年多的博士研究生，背後所鼓勵與支撐的重要力量便是我的家人。包含，外公、外婆、大舅夫婦、小舅夫婦、弟弟、四個家中可愛侄兒以及最辛苦的媽媽。更要感謝總是安靜陪我思考數學分析、程式寫作以及論文閱讀的玟秀，有妳的耐心與包容，更能讓我靜心於研究論文寫作。

最後，感謝在我求學生涯中，一路上幫助、包容、鼓勵我的每一位朋友們，願這份榮耀與大家共享。

許益賓 於 新竹交通大學

# 目　　錄

# List of Tables

# List of Figures

# Chapter 1 Introduction

The rate control mechanism and its role in the encoder are described as Figure 1. The determination of the quantization parameter (QP) value is necessary in the rate control mechanism, and distinct information, such as raw data, encoded bits and distortion (such as dotted line) is arbitrarily referenced to determine the QP value. In general, when the information of encoded bits is required, the rate controller predicts a suitable QP for the encoder under a specified bit rate. Conversely, when the information of distortion is required, the rate controller predicts a suitable QP for the encoder under a specified image quality. Thus, in various applications, the system provides distinct information for the user-designed rate controller.

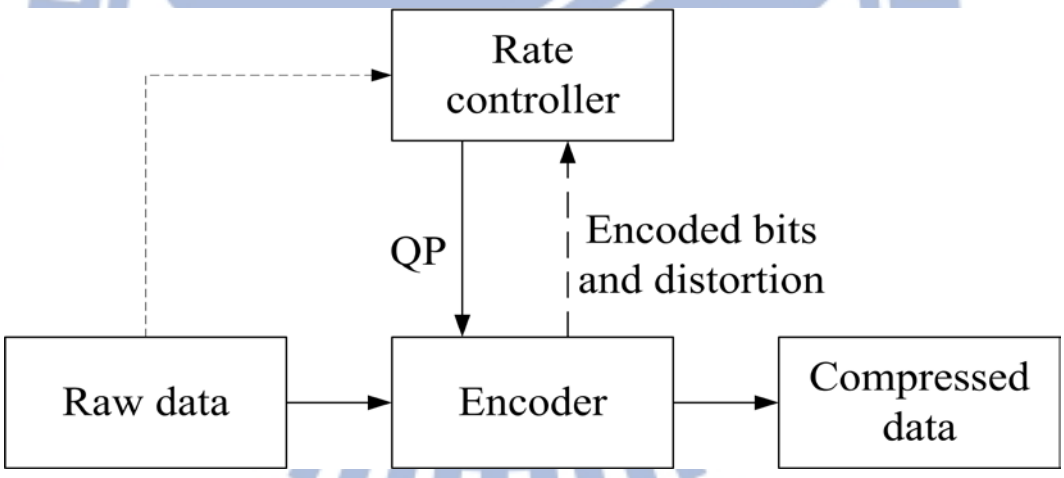



Figure 1. The relationship between rate controller, encoder and relative information.

Bandwidth is a crucial consideration for real time video applications, such as video surveillance and video conferencing. The rate control usually regulates the coded bit stream by modifying the QP value. Because the bandwidth is small and expensive, a low bit rate environment requires more accurate bit allocation and sufficient image quality. Without rate control, underflow or overflow occurs between the encode bit

rate and available channel bandwidth. In other words, an encoder without rate control is difficult to use. Consequently, the rate control mechanism was recommended and built into all prior video-coding standards. We reviewed all typical rate control mechanisms to develop an efficient rate control mechanism, including H.26x [36]-[39] series standard, which was developed by International Telecommunications Union (ITU), MPEG-x [40]-[45] series standard, which was developed by Motion Picture Expert Group (MPEG), and hybrid H.264/AVC [1], [45]-[47] standard, which was developed by Joint Video Team (JVT). We focused on the QP value obtainment equation and listed all relative mathematical expressions for an overall comparison.

## 1.1 H.26x Standard Rate Control Mechanism

### 1.1.1 H.261

H.261 [36] is a block-based codec for ISDN videophone and video conferencing applications. The requirement of constant bit rate is the main consideration because of bandwidth limitation. To satisfy this status, adjustment of the QP value is the main strategy, and the image quality is not considered. In the H.261 reference model 8 [38], the calculated equation is expressed in (1), as follows:

$$QP = 2 \times \left\lfloor \frac{BL}{200 \times p \times 64} \right\rfloor + 2 \qquad (1)$$

where the parameter $p$ is the specified bit rate and $BL$ indicates the encoded buffer level. According to $BL$, the QP can be adjusted directly to preserve buffer capacity.

### 1.1.2 H.263

In H.261, the QP value is only applied for the whole frame. Although the low bit rate requirement is achieved, the image quality is not acceptable. Thus, H.263 [37] was designed for low bit rate application. In the test model, TMN8 [39], the various QP values were applied for the frame and macro block (MB) level. The TMN8 was also considered in the distortion and encoded bits status. In the MB level, the rate model $A \cdot (k\frac{\sigma^2}{Q^2} + C)$ and the distortion model $\alpha^2 \frac{Q^2}{12}$ were used. According to Lagrange optimization, an analytic form for optimized QP estimation is presented in (2), as follows:

$$\begin{cases} Q_i^* = \sqrt{\frac{AK}{(B-ANC)}} \frac{\sigma_i}{\alpha_i} \sum_{k=1}^{N} \alpha_k \sigma_k , i = 0,..., N \\ \alpha_i = \begin{cases} 2\frac{B}{AN}(1-\sigma_i) + \sigma_i, if \ \frac{B}{AN} < 0.5 \\ 1, otherwise \end{cases} \end{cases} \quad (2)$$

where C is the overhand rate, A is the number of pixels in MB, K is a parameter, B is the target bit rate for the frame, N is the number of MB, $\sigma_i$ is the standard deviation of the prediction error of the $MB_i$, and $\alpha_i$ is the weight factor of $MB_i$. From the above result, the QP value can be obtained analytically and the MB level operation can reduce bit consumption and enhance image quality.

## 1.2 MPEG-x Standard Rate Control Mechanism

### 1.1.2 MPEG-2

The MPEG-2 [38] was established for DVD and HDTV applications. Moreover,

TMN5 [38] was used as a tool to develop and verify relative codec functions. The QP value was obtained in (3). Two variables of $R$ and $X_j$ were defined as $R = R - S_j$ and $X_j = S_j \times Q$, respectively. $R$ is the remaining number of bits after encoding; and the three indices, $j = 0$, $j = 1$ and $j = 2$ indicate the I-frame, P-frame and B-frame, respectively. The $BF_j^c$ is the buffer fullness in $c^{th}$ frame, the $BF^{\text{int}ial}$ is the initial buffer fullness, the $B$ and $F$ are the bit rate and frame rate, respectively, the $T_j$ is the target bit rate, the $N_{mb}$ is the total number of MB in the current frame, $N_p$ and $N_b$ are total number of MB in P-frame and B-frame, respectively, $K_p$ and $K_b$ are constant, and suggested as 1.0 and 1.4, and $S_j$ is generated bits of encoding various frame types.

$$\begin{cases} QP_j = \dfrac{BF_j^c \times 31}{2 \times \dfrac{B}{F}} \\ BF_j^c = BF^{\text{int}ial} + B - \dfrac{T_j \times (c-1)}{N_{mb}} \\ T_j = \max\left\{\dfrac{R}{L_j}, \dfrac{B}{8 \times F}\right\} \end{cases} \tag{3}$$

For various frame types, such as I-frame, P-frame and B-frame, $L_j$ is defined in (4), as follows:

$$L_j = \begin{cases} 1 + \dfrac{N_p X_p}{N_i X_I} + \dfrac{N_b X_b}{N_i X_I}, & j = 0 \\ N_p + \dfrac{N_b K_p X_b}{N_i X_p}, & j = 1 \\ N_b + \dfrac{N_p K_b X_p}{N_i X_b}, & j = 2 \end{cases} \tag{4}$$

### 1.2.2 MPEG-4

MPEG-4 [43], based on the video object plane concept, was designed. The main goal was to improve image quality and reduce bits consumption. In the rate control mechanism, it is referred to as Q2 and realized on the VM codec [44]. The QP value was obtained in (5).

$$\begin{cases} T_2 = x_1 \dfrac{MAD}{QP} + x_2 \dfrac{MAD}{QP^2} \\ \begin{cases} T_2 = T_1 \times \dfrac{2 \times B_s - B}{B_s + B} \\ T_1 = \max\left( \dfrac{R_s}{30}, 0.95 \times \dfrac{R_r}{N_r} + 0.05 \times S \right) \end{cases} \end{cases} \tag{5}$$

where $T_1$ and $T_2$ are the initial bits budget and target bits budget, respectively; $B$ and $B_s$ are the buffer fullness and buffer size, respectively; $R_s$ and $R_r$ are the assigned bit rate and remaining bits, respectively; $N_r$ is the remaining frame for encoding; and $S$ is the model parameter. The MAD (mean absolute difference) was obtained after motion compensation in the current frame; and $x_1$ and $x_2$ are the model parameters.

## 1.3 The Relationship between Rate-Distortion Theorem and H.264/AVC Rate Control

H.264/AVC [45] is an advanced coding standard that was jointly developed by ITU-T and ISO/IE. Although H.264 provides several useful functions, such as multiple reference frames, numerous MB types, 4x4 transforms than previous video coding standards, the main difference is the rate-optimization mechanism, which simultaneously considers motion estimation and rate control information. The QP

value estimation equation follows MPEG-4 and is realized in the joint model (JM) reference software [46] and [47]. All of the detailed equations are listed in (6).

$$\begin{cases} T_1 = x_1 \dfrac{MAD}{QP} + x_2 \dfrac{MAD}{QP^2} \\ \begin{cases} T_1 = T - T_h \\ T = \beta \times \dfrac{R_r}{N_r} + (1-\beta) \times T_b \\ T_b = \dfrac{u(n_j)}{F_r} + \gamma \times (T_{bl}(n_j) - B_c(n_j)) \end{cases} \end{cases} \tag{6}$$

where $T$, $T_1$, and $T_h$ are the target bits for the current frame, the number of bits estimated for the texture information for the current frame, and the number of bits estimated for the header and motion information of the previous frame, respectively; and $\gamma$, $\beta$, $x_1$, and $x_2$ are model parameters. MAD was obtained after motion compensation in the current frame; $R_r$ and $N_r$ are the number of remaining bits and frames, respectively; $F_r$ is the user-defined frame rate; and $u(n_j)$, $T_{bl}(n_j)$, and $B_c(n_j)$ are available channel bandwidth, target buffer level, and actual buffer level, respectively.

According to above analysis, in MPEG-4 and H.264/AVC, the quadric form for the QP value estimation is a crucial equation and is applied in the current standard. It is necessary to understand the original theorem for efficient rate control development. In the H.264/AVC video-coding standards, the QP specifies the quantization step size, which subsequently influences image quality and bit budget. Several coding parameters must be determined, such as the MB type, QP, and multiple reference frames. Consequently, it is able to achieve considerably higher rate control efficiency than conventional video coding methods. The rate-distortion optimized mode decision

and motion estimation process are also affected by the QP. The QP can effectively be used to control the bit stream to maximize the coding efficiency without the overflow or underflow channel rates. Because QP is specified in rate control and rate-distortion optimization (RDO), a problem known as “the chicken and egg dilemma” must be considered when implementing rate control. The rate control typically separates into two parts, including the frame layer and the MB layer. A typical rate control scheme consists of two fundamental operations, called bit allocation and bit allocation achievement. The optimal rate control should be achieved jointly for the two operations. The following expresses the optimal bit allocation problem:

$$\begin{aligned} &\min\ D, \\ &\textit{subject to } R < R_c \end{aligned} \tag{7}$$

where $D$ denotes the distortion between the original and the reconstructed frame, $R_c$ denotes the target bit budget, and R denotes the total bits after the encoding process. The main influence of distortion is QP determination, thus an appropriate method is necessary to achieve the target bit budget. Three relations of rate-distortion (R-D), rate-quantization (R-Q) and distortion-quantization (D-Q) have been investigated recently. The main contribution of different relations is to predict a suitable QP under distortion or bit budget.

The theorem in [1] and [48] provides the original definition of the rate-distortion form used in this study. We discuss several necessary aspects for enhanced modeling operation. The relation between rate (R) and distortion (D) can be formulated as follows:

$$R(D) = s \times D + \int_{-\infty}^{\infty} Q(u) \ln(\lambda(u)) du \tag{8}$$

where $s \le 0$, D and $\lambda(u)$ are defined as follows:

$$\begin{cases} D = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \lambda(u)Q(u)P(v)e^{sd(u,v)}d(u,v)dudv \\ \lambda(u) = \left[\int_{-\infty}^{\infty} P(v)e^{sd(u,v)}d(u,v)dv\right]^{-1} \end{cases} \quad (9)$$

$d(u,v)$ is an error metric measurement. Consider a source, $Q(u)$, that outputs an independent Gaussian random variable. The original source can be treated as $P(v)$, which is also an independent Gaussian random variable, with the assumption that a squared-error distortion measure is $d(u,v) = (u-v)^2$. Suppose the terms $\sigma$ and $\beta$ represent the standard deviation of the output $Q(u)$ and input $P(v)$ source, respectively. Taking $Q(u)$ and substituting $P(v)$ into the formula (9) yields the following relation:

$$D = \frac{\alpha^2\beta^2}{\alpha^2+\beta^2} + \left(\frac{\alpha^2}{\alpha^2+\beta^2}\right)^2\alpha^2 \quad (10)$$

where $\alpha^2 = \frac{-1}{2S}$. Therefore, $\alpha^2$ is directly related to the parameter *S*, whereas $\beta^2$ is unrestricted. Thus $\beta^2$ is chosen to satisfy $\alpha^2+\beta^2=\sigma^2$, producing the following relation on *S*:

$$D = \alpha^2 = \frac{-1}{2S} \quad (11)$$

The expression for R(D) then becomes

$$R(D) = \frac{1}{2}\ln\left(\frac{\sigma^2}{D}\right) \quad (12)$$

In [5], *R(D)* can be expanded by a Taylor series to construct a new model to evaluate the target bit rate before performing the actual encoding process. The quadratic form is modified by substituting *D* and $\sigma$ for *QP* and *MAD*, respectively. The new model is formulated in following equation:

$$R = X_1 \frac{MAD}{QP} + X_2 \frac{MAD}{QP^2} \quad (13)$$

where $X_1$ and $X_2$ are first- and second-order coefficients that can be updated by the linear regression method. To avoid the chicken-egg problem, the *MAD* is predicted from the previous frame. The resulting model is used in standard closed-form to calculate the QP directly for JVT-G012.

## 1.4 Brief Review of Past Approaches of H.264/AVC-based Rate Control

Based on H.264/AVC architecture, the studies of [2] and [3] proposed improved algorithms to optimize the subjective quality of video in off-line processing. Moreover, [3] used a video quality metric and considered human visual to enhance the previous study [2]. The optimal application in [2] and [3] is consumer products, such as DVD players. Moreover, H.264/AVC defines two distinct frame types. The first frame type without any referenced image is referred to as intra-frame and the other frame type is referred to as inter-frame in encoded video sequences. According to larger video training, obtaining an approximate equation is necessary to estimate the QP value for intra-frame rate control. Moreover, the MAD information was used in [8], and the information of image density and gradients was used in [9]. In [22], according to the combination of image gradients and histograms, it is useful to represent QP estimation under the exponent-based equation. The relationship between MSE and QP was verified in [12]. In Chapter 2, we introduce the relative articles of the intra-frame rate control algorithm in detail.

Conversely, the inter-frame can refer the previous and next frame simultaneously to encode the current frame. The rate and distortion model was discussed and

developed for QP estimation in rate-distortion optimization applications. Based on the same distortion model, [14] and [15] developed differing rate models to estimate the QP value. By contrast, [15], [18], and [20] used the same rate model that was used for H.264/AVC reference software [7] to develop various distortion models for QP value estimation. In real time applications, the use of the empirical MAD prediction in [27] and [28] was developed to determine a suitable QP, which was the main consideration. An approximated rate-distortion model and an estimated texture-complexity model was the main consideration in [29]. We also introduce relative articles of the inter-frame rate control algorithm in detail from chapters 3 to 5.

## 1.5 Motivation and Contributions

Based on the above analysis, the rate control mechanism plays a vital role in various video-coding standards. For specified applications, the QP value estimation must be modified according to the various conditions. The rate model and/or the distortion model are developed under a user-defined assumption to achieve a more accurate QP estimation. In MPEG-4 and H.264/AVC, the quadratic equation of QP value estimation is used because of the distortion, and because two model parameters assist and correct QP estimation. However, in H.264/AVC, the distortion measurement is infrequently obtained because of “the chicken and egg dilemma,” and the linear regressive method was commonly used. The linear regressive method is not suitable for real time application and is not accurate for rate control. Moreover, the distinct frame types have differing referenced information, thus, a separated rate control algorithm should be developed from the original rate distortion theorem for the distinct frame types and replace the linear regressive method in the thesis.

We applied the obtained novel distortion measurement method to four differing applications to verify the efficiency of the method, as follows:

1. To improve the intra-frame quantization parameter obtainment in the first frame in each group of pictures, we used the novel distortion equation to replace mean square error metric and subsequently constructed an efficient relation between the distortion and quantization parameter.
2. An analytic closed-form equation for quantization parameter obtainment was derived to combine this novel distortion equation and rate prediction equation, which is recommended in H.264/AVC. According to this closed-form equation, the quantization parameter can be calculated directly and applied for rate control.
3. To consider the real application, this novel distortion equation was applied and the property of the image content was also used to refine the quantization parameter when the initial quantization parameter was pre-obtained. Moreover, in hardware consideration, we used the histogram as image content index because the computation complexity was lower than other approaches, such as gradient operation. Based on both properties, the refined quantization parameter was more suitable for rate control.
4. To solve the low bit rate limitation in video conferencing, we used the novel distortion equation to replace variance-based image complexity measurement in the past approach. Our main contribution is to use continuous time-domain image and to combine the past approach in quantization parameter obtainment for more adaptive rate control.

## 1.6 Organization of dissertation

In Chapter 2, we review the intra-frame rate control in JVT-G012 [6] for QP computation to introduce the proposed intra-frame rate control algorithm with a variance-based distortion. Subsequently, we evaluate the proposed algorithm and

present the simulation results. In Chapter 3, we introduce an integration of the quadratic rate model and the proposed distortion prediction equation. The Lagrange method obtained a new QP determination method to solve R-D optimization. We subsequently evaluate the proposed algorithm and provide the simulation results. In Chapter 4, we briefly review the original quadratic rate control model for QP computation. A modified HOD was added to solve the problem of local motion before an enhanced equation was analyzed and verified with the original quadratic model. Finally, we evaluate the proposed algorithm and provide the simulation results. In Chapter 5, we review the rate control algorithm of past approaches and introduce Xie's [35] rate control model. Subsequently, we propose a novel rate control algorithm with rate-distortion theorem and time-domain image, and evaluate the proposed algorithm and present the simulation results. Finally, Chapter 6 offers conclusions.

# Chapter 2 An Efficient Variance-based Intra-Frame Rate Control Algorithm

To enhance image quality and retain a limited bit budget, intra-frame in the frame type selection plays a significant role in video coding systems because the frame is treated as a key frame for temporal domain reference. This chapter presents an algorithm to predict accurate quantization parameters by developing variance-based distortion measurements. Current algorithms includes JVT-G012 [6] and Yan's [10] algorithm based on mean square error metrics for intra-frame rate controls are compared to show that the proposed distortion metric is useful to estimate quantization parameters. Experimental results show that the proposed algorithm can significantly improve video quality up to 0.19 dB and 1.28 dB, compared with the algorithm JVT-G012 [6], which is implemented in the H.264 reference software JM 15.0 and Yan's [10] algorithm.

## 2.1 Review of Related Articles

The intra-frame rate control algorithm has been embedded in the JVT-G012 [6] and reference software JM 15.0 [7], both are based on bits per pixel (BPP), frame rate, and the bits budget to decide QP. Although the BPP can provide a reference QP value for intra-frame encoding, the image-complexity property and the wide range of QP are not utilized to enhance overall video quality.

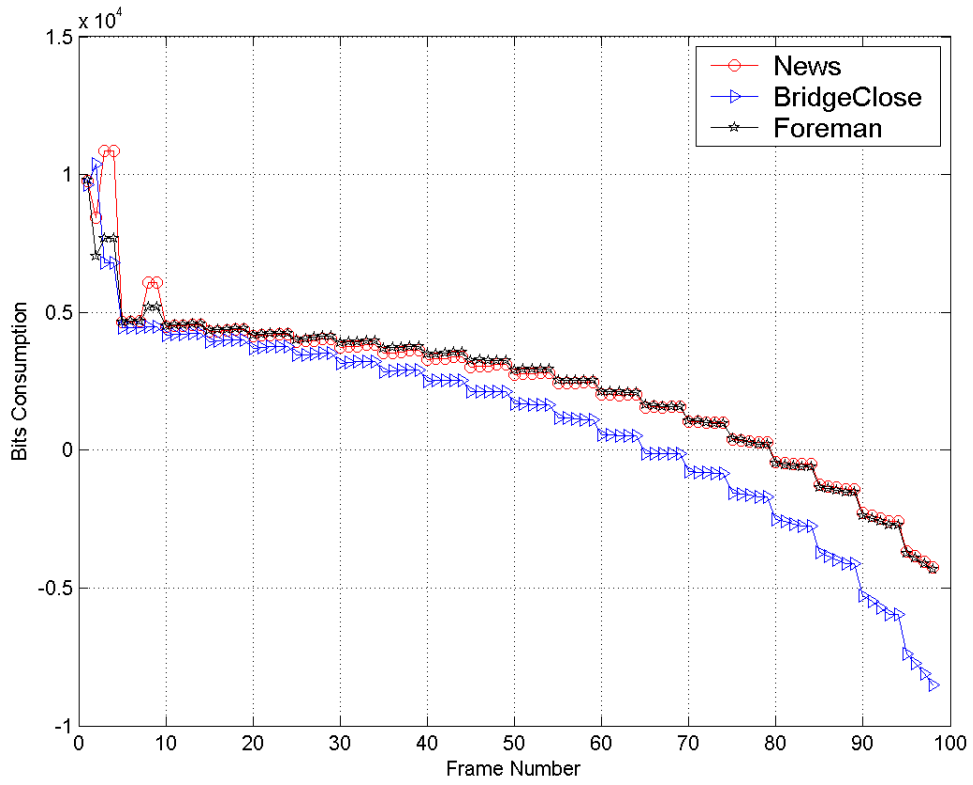
Each first frame in a picture group is a critical reference frame for post processing. In JVT-G012 [6], the algorithm of intra-frame QP selection is formulated as follows:

$$QP = \begin{cases} 35, & BPP \le L_1 \\ 25, & L_1 < BPP \le L_2 \\ 20, & L_2 < BPP \le L_3 \\ 10, & BPP > L_3 \end{cases} \quad (14)$$

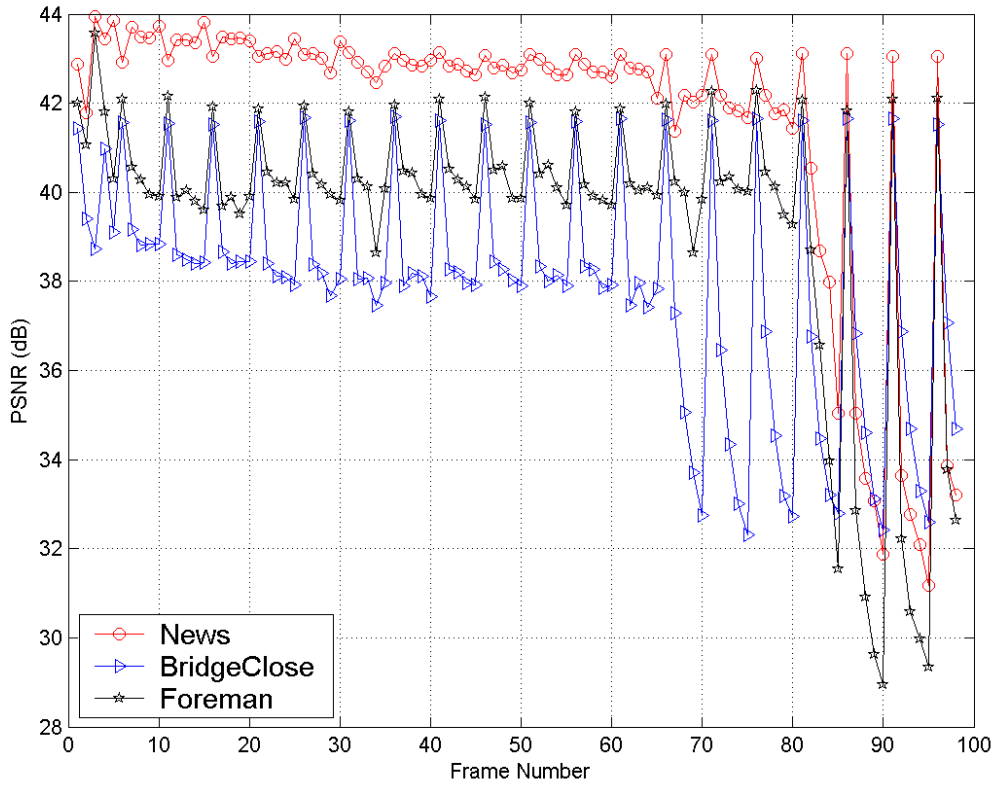
$$BPP = \frac{R}{F \times W \times H} \quad (15)$$

where R is the assigned bit rate, F is the frame rate, and W and H connote image width and height, respectively. The set of three parameters $\{L_1, L_2, L_3\}$ are predefined as {0.1, 0.3, 0.6} for QCIF video sequences and {0.2, 0.6, 1.2} for CIF video sequences, respectively. However, unsuitable BPP leads to uncontrollable bits allocation and image quality because the image content is not considered. The large QP is selected in the formula (14) when a large budget R is applied in the formula (15). However, if the image content is complex the QP should be slightly reduced to maintain complete image quality.

To prove the condition of unsuitable QP estimation in [6], the “News”, “Bridge-close” and “Foreman” video sequences in Figure 2 are selected for evaluation. For the main parameters comprising 300 Kbps for the target bit rate, 30 fps frame rate, the encoding structure is IPPPPIPPPP, and JVT-G012 decides the entire intra-frame QP is constant by [6]. The presence of an unbalanced bit allocation is worth noting. Because several frames consume more bits in the front video sequence, the status of bits underflow occurs in the hind video sequence. Though the large number of bits is used to encode frames between the 1st and 5th frame when a small QP is applied, the increase in image quality is barely noticeable. During the interval between the 61st and 99th frame, the image quality is large reduced to balance the target bits budget, resulting in unsmooth visualization. In summation, the two drawbacks comprise the uncontrollable bits allocation and noticeable decay in image quality. Thus developing an efficient rate control algorithm for intra-frame is an essential strategy.




(a)




(b)

Figure 2. Comparison of the bits consumption in (a) and corresponding PSNR in (b) are shown in "News", "Bridge-close" and "Foreman" video sequences.

In [8] and [9], according to image and larger video training, obtaining an approximate equation is necessary to estimate the QP value. For actual application, the information of image density and gradients is used in [9], and an image distortion such as MAD is also employed in [8] as the indicator, to calculate the final QP value. In [11], image edge characteristics and BPP are applied to construct an equation for QP estimation. The image edge characteristic applies a Laplacian matrix of Gaussian (LoG) operators. In [22], according to extensive experimentation results, the combination of image gradients and histograms, both including luminance and chrominance coefficients, is extremely useful to represent QP estimation under the exponent-based equation.

[12] verified the relation between MSE and quantization step ( $Q_{step}$ ). Obtaining $Q_{step}$ requires estimating MSE from the previous frame. A real application equation is expressed as follows:

$$Q_{step} = \rho \times MSE \tag{16}$$

where $\rho$ is the variable for difference video sequences. To fit $Q_{step}$ estimation more accurately, however, [8] have proposed a gradient-based equation to modify the formula (16) as follows:

$$Q_{step} = \eta \times MSE^{prev} \times \left(\frac{Grad^{curr}}{MG} \times \frac{1}{1 - \frac{BF^{prev}}{Buffer\_Size}} + \theta\right) + \varepsilon \tag{17}$$

Where $\eta$, $\theta$, and $\varepsilon$ are variables for difference video sequences. $MSE^{prev}$ is the previous frame MSE and $BF^{prev}$ is the previous buffer fullness after encoding the i$^{th}$ group of picture (GOP). $Grad^{curr}$ is the gradient value of the current frame, and MG is

the average gradient value of the previously encoded I-frame in this sequence. According to various encoding test sequences, the relationship between $Q_{step}$ and the gradient value is quasi-linear. This study also considers the buffer fullness to further improve the QP estimation and avoid the buffer overflow or underflow. The image complexity and bits buffer status are used to adaptively modify $Q_{step}$ because the MSE cannot accurately reflect QP estimation.

According to the above analysis, the distortion will primarily influence the $Q_{step}$ estimation; thus, attempting to discuss the relation between distortion and QP becomes an important topic. Distortion is performed via the quantization process in the video encoding system. According to the theoretical definition, the MSE can be expressed as follows:

$$\begin{aligned} MSE &= \frac{1}{N}\sum_{i=0}^{N}(x_i^{Curr} - x_i^{Quan})^2 \\ &= \frac{1}{N}\sum_{i=0}^{N}(x_i^{Curr} - \frac{x_i^{Curr}}{QP})^2 \\ &= \frac{1}{N}\sum_{i=0}^{N}(x_i^{Curr})^2(1-QP^{-1})^2 \end{aligned} \tag{18}$$

where N represents all the image pixels, and $x_i^{Curr}$ and $x_i^{Quan}$ are the current image and the quantized image at the $i^{th}$ pixel, respectively. In the formula (18), the MSE is influenced by image content and QP value. The increase in QP enlarges the MSE when the image content is constant, and vice versa.

## 2.2 Our Proposed Algorithm

According to the aforementioned observations in the formula (11), the distortion is evidently arbitrarily variable. The condition $\alpha^2 + \beta^2 = \sigma^2$ and the formula (11) lead to the following equation:

$$D = \sigma^2 - \beta^2 \tag{19}$$

□The analytic result (13) is an analytic and novel equation to estimate distortion. This equation is used to develop an efficient rate control algorithm.

This study performs a new variance-based distortion measurement. The formula (20) shows a detailed deduction process for evaluating the distortion:

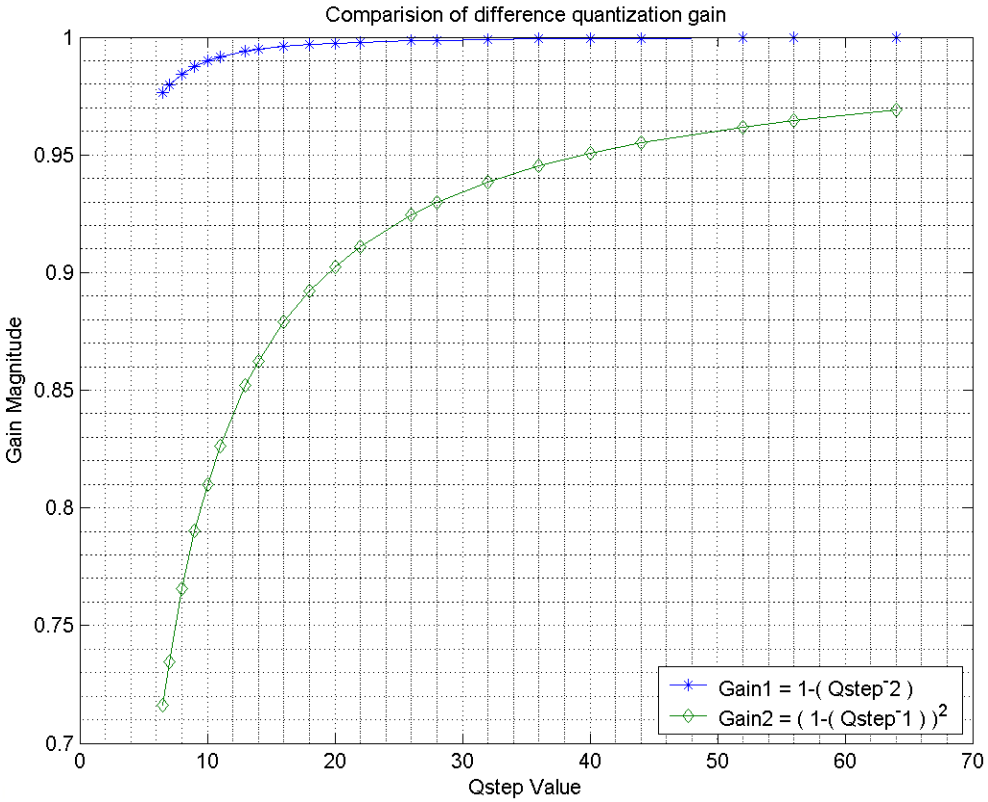
$$\begin{aligned} D &= \sigma^2 - \beta^2 \\ &= \frac{1}{N}\sum_{i=0}^{N}(x_i^{Curr} - M_\alpha)^2 - \frac{1}{N}\sum_{i=0}^{N}(x_i^{Quan} - M_\beta)^2 \\ &= \frac{1}{N}\sum_{i=0}^{N}(x_i^{Curr} - M_\alpha)^2 - \frac{1}{N}\sum_{i=0}^{N}(\frac{x_i^{Curr}}{QP} - \frac{M_\alpha}{QP})^2 \\ &= \frac{1}{N}\sum_{i=0}^{N}(x_i^{Curr} - M_\alpha)^2(1 - QP^{-2}) \end{aligned} \tag{20}$$

where N comprises all the image pixels, $x_i^{Curr}$ and $x_i^{Quan}$ are the current image and the quantized image at the i[th] pixel, respectively. The terms $M_\alpha$ and $M_\beta$ are the mean value of the current image and quantized image. This distortion, D, comprises two items to compare with the formula (18): the first item is the consideration of the image mean, and the other is the new QP equation.

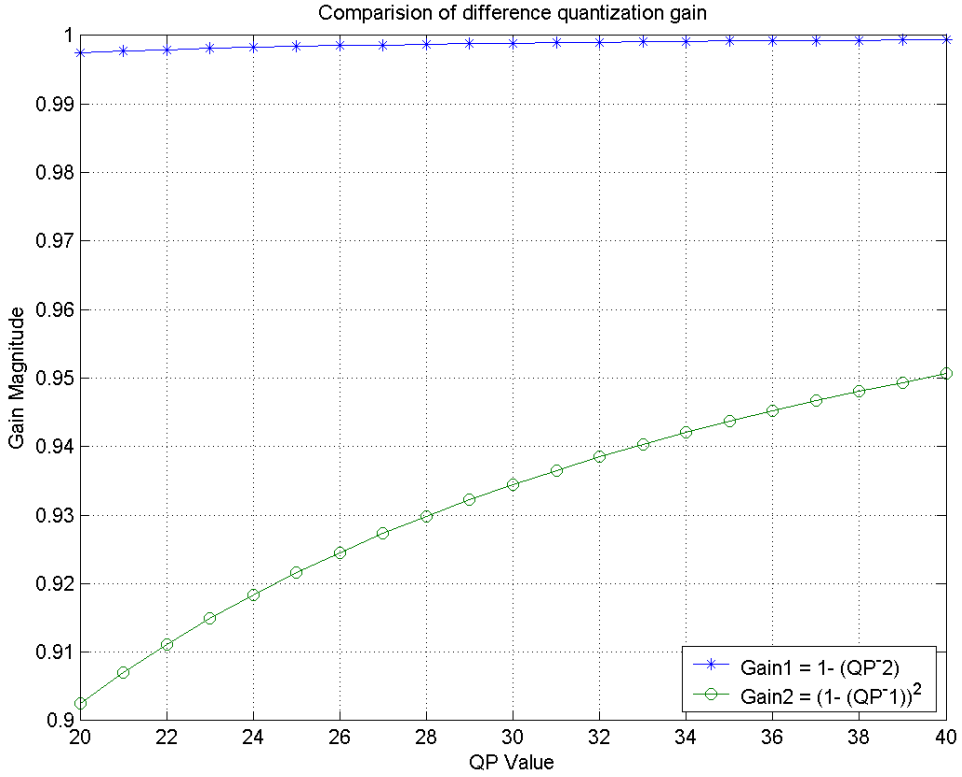
In past approaches [12], the estimated QP could be obtained via distortion, and the relative parameters are obtained via pre-training, the main reason being that the QP estimation of the first frame in the video sequence is not referential to any frame.

Although the MSE can reflect actual QP estimation, the precision is not accurate. Due to this reason, the image content is considered in [10] for adaptive QP modification. The training procedure requires the linear relationship between QP and the distortion, but the intrinsic problem is not clearly identified. To compare the formulas (18) and (20), the main effect item is QP for distortion obtainment. Explaining the QP effect necessitates setting the bit rate to 400 kbps and the range of $Q_{step}$ from 6.5 to 64 (QP is between 20 and 40) to plot the relationship.

Figure 3 shows the selection of two different variables to present the gain magnitude. Results demonstrate that the linear mapping property in Gain1 is a better fit than Gain2, and the vibration is also small because this property can perform a more accurate QP estimation in the training procedure than the nonlinear mapping property [8], [22]. Thus, selecting a variance-based distortion for QP estimation is more suitable than selecting an MSE-based distortion.




(a)




(b)

Figure 3. Comparison of the difference quantization gain is shown in two different equations: (a) using the $Q_{step}$ variable, and (b) using the QP variable.

When the variance-based distortion utilizes either of the two parameters, the linear mapping property is also sustained. The QP parameter is more suitable than $Q_{step}$ for constructing the quantization-distortion (Q-D) relation equation, as follows:

$$QP = \gamma \times D + \delta \tag{21}$$

The formula (21) also includes the image complexity through the variance information to more accurately reflect the QP estimate. Two variables of $\gamma$ and $\delta$ are system parameters.

To integrate the intra-frame QP estimation in JVT-G012, the description of the algorithm is represented as follows: if the input frame is first in the entire video sequence, QP value estimation can be achieved via the formula (14), restoring the reconstructed image variance. If the input frame is the first frame in the next GOP, the distortion could be calculated via current frame variance and the previous intra-frame variance. According to the distortion, the estimated QP value can be obtained by the formula (21). The QP estimation for the inter-frame is calculated by the JVT-G012 quadric close-form, which is also implemented in JM15.0. A whole view of the intra-frame rate control is depicted in Figure 4.

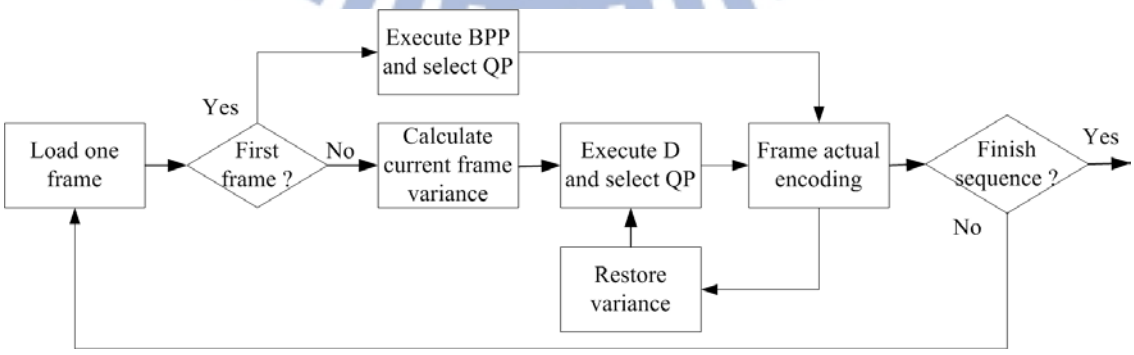



Figure 4. Integration of JVT-G012 and variance-based intra-frame rate control

## 2.3 Simulation Results

This algorithm used H.264/AVC reference software JM15.0 [7] to evaluate the proposed rate control algorithm. The conducted evaluation required using the first 100 frames of five QCIF test sequences. The test target bit rate is 300 Kbps for ” Foreman,” and ” News,” and 400 Kbps for “Bridge-close,” “Highway,” and “Grandma.” Each sequence was coded at 15 fps according to the IPPPPIPPPP structure. The reference frame was set to 5, and the search window was set to 15. CAVLC, RDO, and rate control were also enabled. The study equally selected other relative parameters for JM15.0 [7], [10], and the proposed algorithm. To obtain the model parameters in the off-line operation, five QPs were attempted, including 20, 25, 30, 35, and 40 in the first frame of the video sequence, and their respective distortions D were recorded. Via linear regression and recorded distortions D, the good Q-D performance curve was constructed. Table 1 lists the model parameters of $\gamma$ and $\delta$ on various test sequences.

Table 1. Model parameters of Q-D model with different video sequences

| Video sequence | $\gamma$ | $\delta$ |
|---|---|---|
| Foreman | 0.0465 | -67.3608 |
| News | 0.5885 | -40.2565 |
| Bridge-close | 0.4794 | -77.6881 |
| Highway | 0.8383 | -135.2993 |
| Grandma | 0.9146 | -57.6692 |

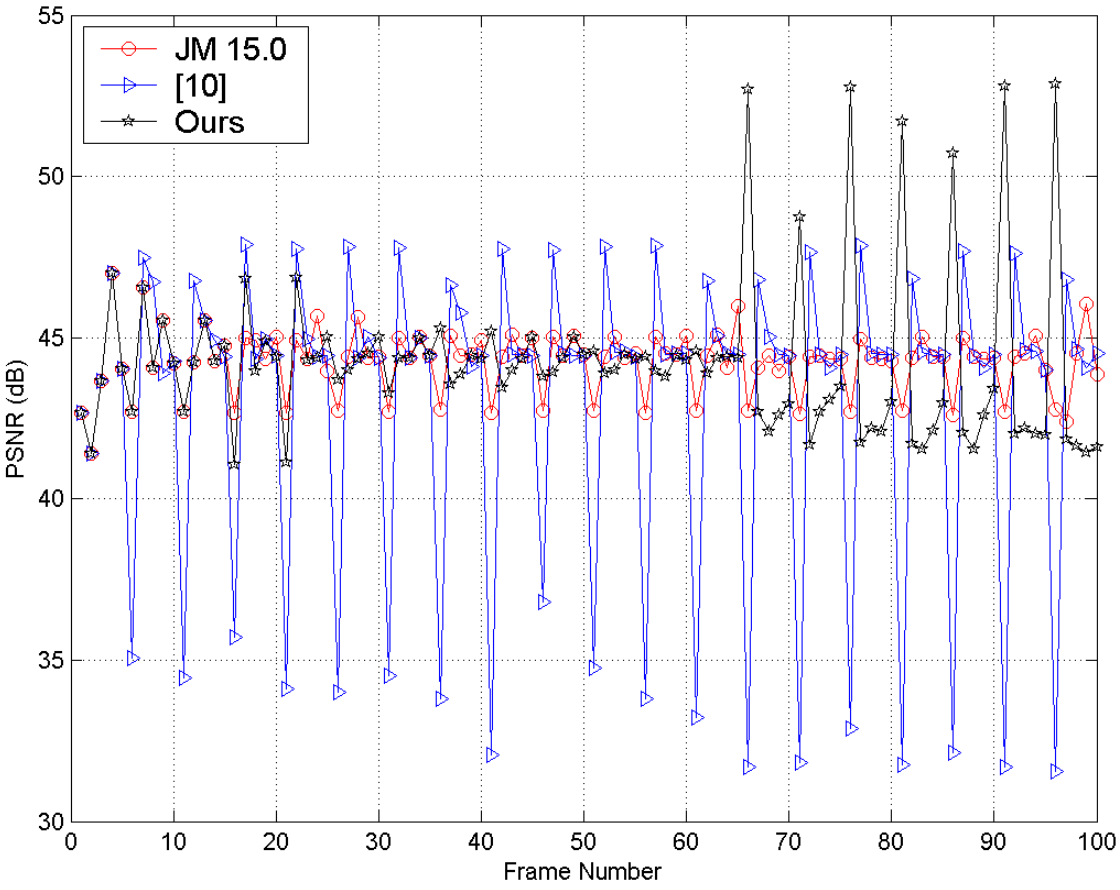



Figure 5. PSNR comparison of JVT-G012, [10] and our proposed algorithm on “Highway” video sequence.

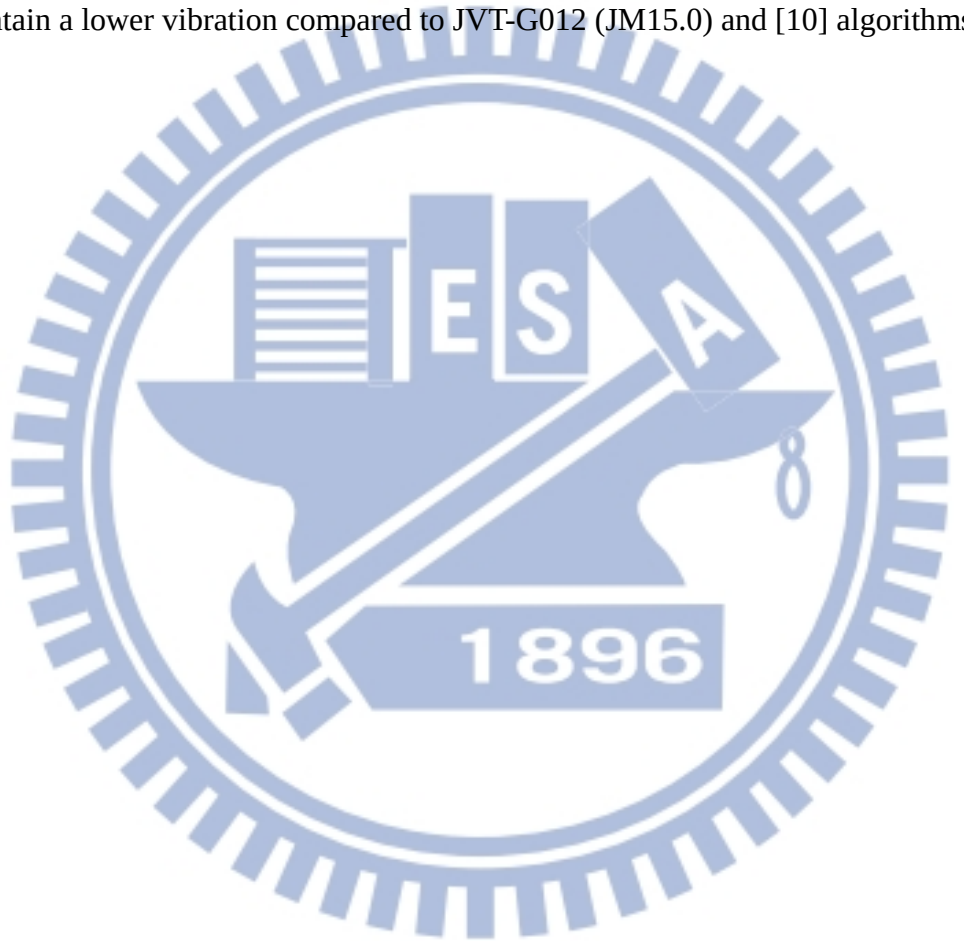
Figure 5 shows the frame-by-frame PSNR comparison of three algorithms for the "Highway" video sequences. Since the JVT-G012 algorithm in the intra-frame QP estimation is not adaptive, all the intra-frames utilize the same QP. [10] predicted that QP can sometimes be unsuitable; for example, a disadvantage is that the inter-frame only uses the large QP to preserve the target budget. Our proposed algorithm leads to image quality enhancement in the 65$^{th}$ frame, 71$^{st}$ frame, 75$^{th}$ frame, 81$^{st}$ frame, 85$^{th}$ frame, 91$^{st}$ frame and 96$^{th}$ frame in Figure 5 especially. The proposed algorithm significantly enhances the PSNR in intra-frame and also holds suitable quality in inter-frame when compared to JVT-G012 and [10].

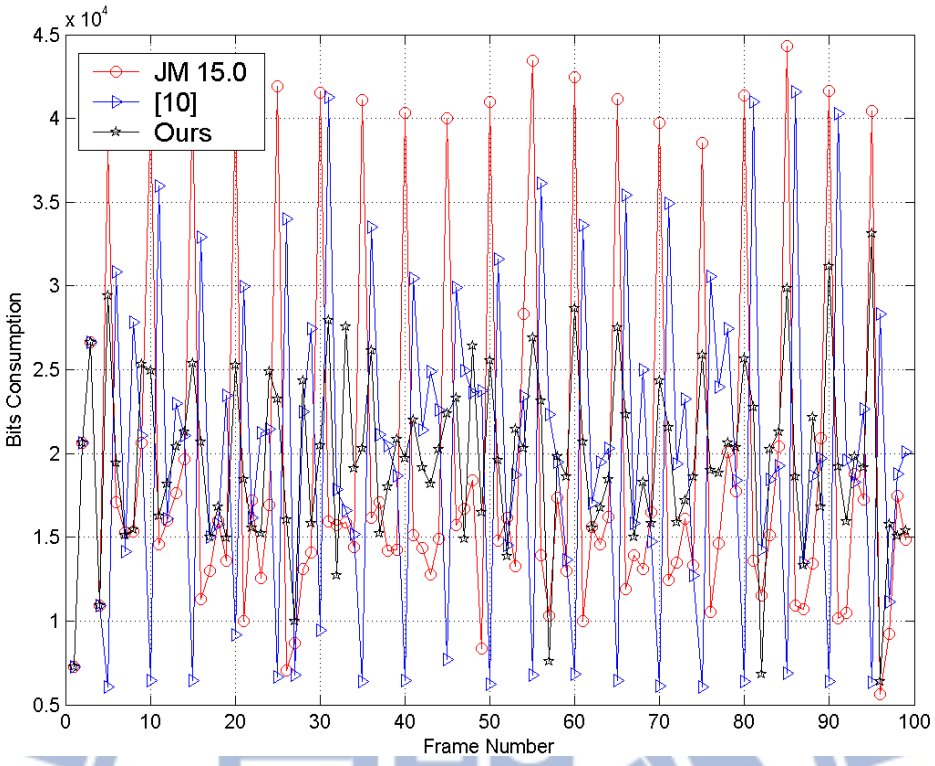
Figure 6 displays the bits consumption for "Foreman" and "News" test video sequences. The figure demonstrates that the proposed algorithm can maintain smoother and more suitable bits operation than both JVT-G012 and [10], because both use QP values which are too small for previous frames, consuming large numbers of bits. In our algorithm, bit allocation is smoother, not only for intra-frames, but also for inter-frames because the intra-frame receives suitable QP and saves bit count for post inter-frame encoding. Our result sustains smooth image quality and is highly suitable for limited bandwidth networks.

For a more detailed comparison and performance analysis, Figure 7 shows the number of encoded bits in the buffer at each frame. Moreover, the positive and negative value in the buffer fullness, called "overflow" and "underflow," respectively, indicate that the encoded frame consumes more bits than the pre-allocated budget or the wasted allocated budget.

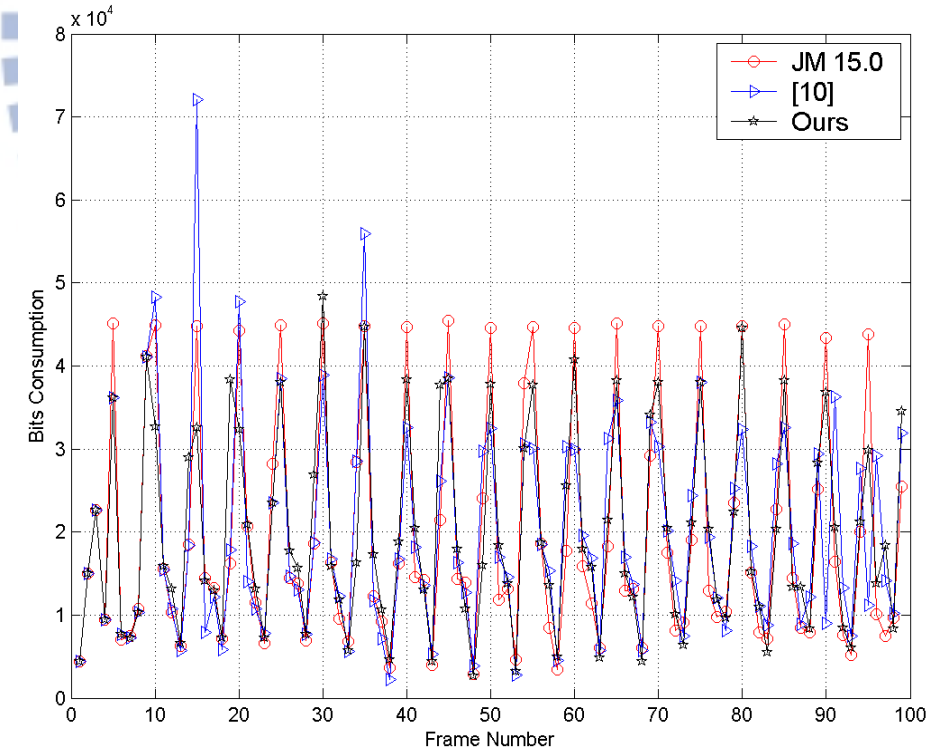
In [27]-[28], the authors detail a fair evaluation for buffer fullness, called the buffer size. A recommended buffer size is set to two-thirds (2/3) of the channel bandwidth (one second). The goal of the buffer size is for the encoded bits in every frame to be limited in this boundary to meet real-time requirements. According to the

buffer size limitation and 15 fps condition, maximum buffer fullness is approximately (400 k/15 fps)×(2/3) = 17.7 k bits per frame. Figure 5(a) displays a good result and shows that the proposed algorithm achieves steadier buffer fullness when compared to JVT-G012 (JM15.0) and [10] algorithms. In total, five encoded frames in Figure 5(b) over the pre-allocated bit budget in the 1st, 10th, 15th, 21st and 37th frames. Although five over-consumed bit frames are shown in Figure 5(b), overall video sequences still maintain a lower vibration compared to JVT-G012 (JM15.0) and [10] algorithms.



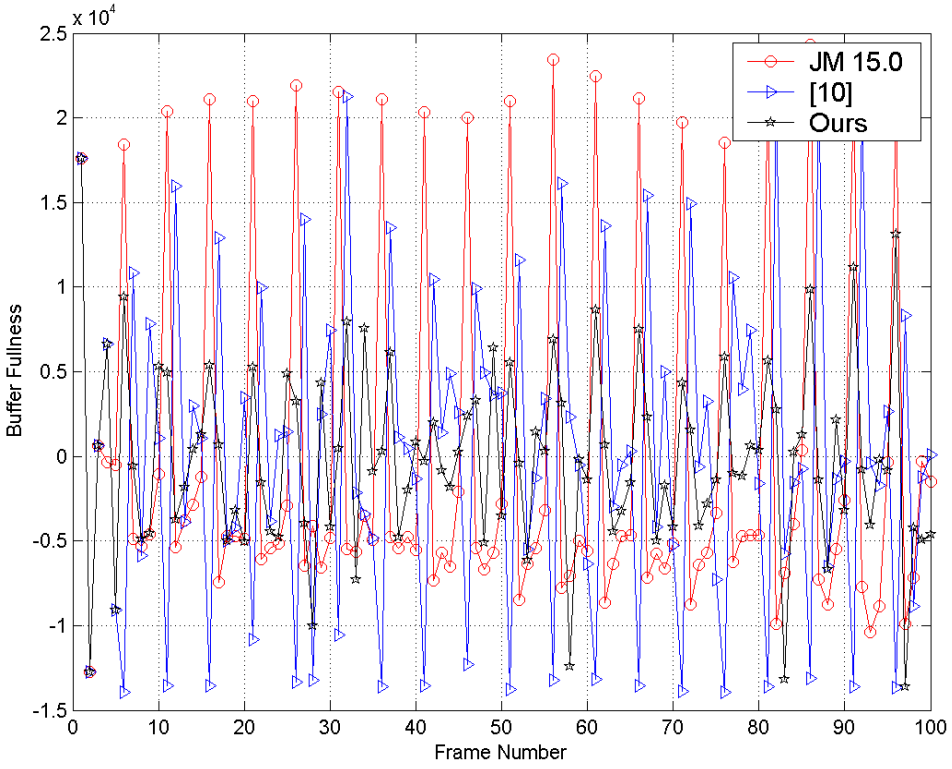



(a)

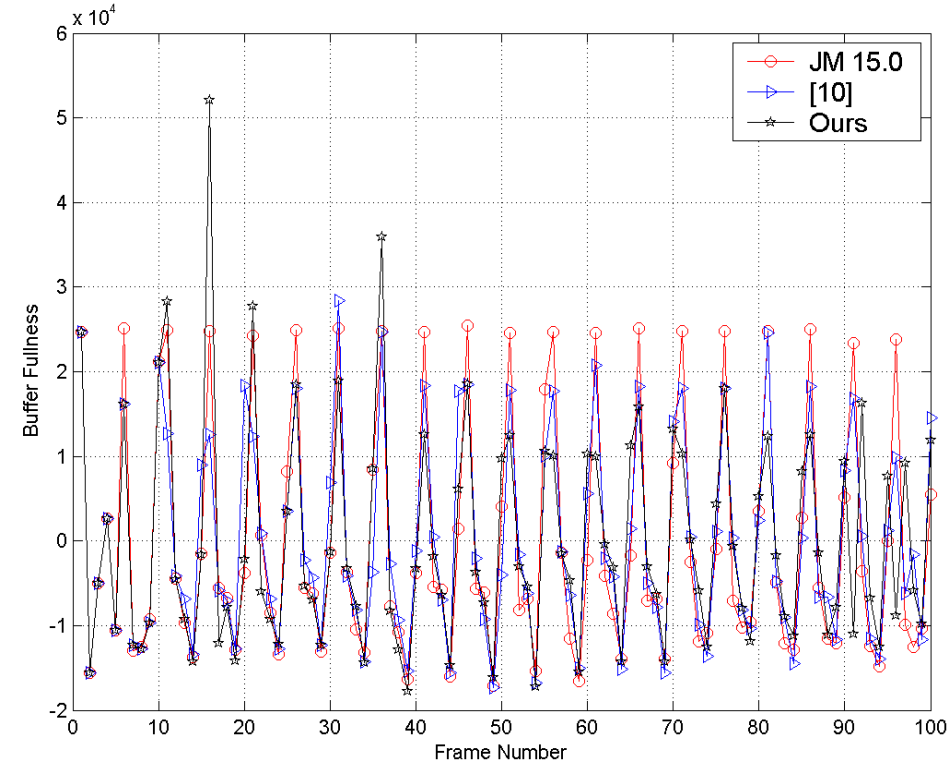



(b)

Figure 6. Bits consumption comparison of JVT-G012 (JM15.0), [10] and the proposed algorithm, on two different video sequences: (a)“Foreman” and (b) “News”.




(a)




(b)

Figure 7. Buffer fullness comparison of JVT-G012 (JM15.0), [10] and the proposed algorithm, on two different video sequences: (a)“Foreman” and (b) “News”.

Table 2. Performance of three algorithms in term of average PSNR, PSNR std. deviation, bit rate and ΔR

| Sequence | Method | Average PSNR (dB) | PSNR Std. Deviation | Bit Rate (kbps) | ΔR |
|---|---|---|---|---|---|
| Foreman (300 kbps) | JVT-G012 | 43.25 | 17.18 | 298.63 | 0.46 |
| | [10] | 42.87 | 18.17 | 298.97 | 0.34 |
| | Ours | 43.39 | 17.61 | 298.34 | 0.55 |
| | Gain 1 | +0.14 | — | — | — |
| | Gain 2 | +0.52 | — | — | — |
| News (300 kbps) | JVT-G012 | 47.26 | 19.09 | 299.50 | 0.17 |
| | [10] | 47.25 | 19.34 | 299.67 | 0.11 |
| | Ours | 47.42 | 19.34 | 300.38 | 0.13 |
| | Gain 1 | +0.16 | — | — | — |
| | Gain 2 | +0.17 | — | — | — |
| Bridge-close (400 kbps) | JVT-G012 | 41.76 | 16.48 | 399.81 | 0.05 |
| | [10] | 41.63 | 16.74 | 400.21 | 0.05 |
| | Ours | 41.95 | 16.21 | 399.60 | 0.10 |
| | Gain 1 | +0.19 | — | — | — |
| | Gain 2 | +0.32 | — | — | — |
| Highway (400 kbps) | JVT-G012 | 44.23 | 17.58 | 400.08 | 0.02 |
| | [10] | 42.96 | 17.86 | 400.13 | 0.03 |
| | Ours | 44.24 | 17.25 | 400.76 | 0.19 |
| | Gain 1 | +0.01 | — | — | — |
| | Gain 2 | +1.28 | — | — | — |
| Grandma (400 kbps) | JVT-G012 | 47.93 | 19.45 | 400.69 | 0.17 |
| | [10] | 47.86 | 19.44 | 398.49 | 0.38 |
| | Ours | 48.12 | 18.96 | 398.62 | 0.25 |
| | Gain 1 | +0.19 | — | — | — |
| | Gain 2 | +0.26 | — | — | — |

The proposed algorithm is more adaptive, producing smooth frame type changes, because it can adjust the QP by using complexity measurements. This study adopts a formula in [12] to further evaluate the bit rate mismatch quantification, as follows:

$$\Delta R = \frac{|R_t - R_b|}{R_b} \times 100\% \quad (22)$$

where $R_t$ is the bit rate of the test algorithm and $R_b$ is the target bit rate. ΔR represents the degree of mismatch in the produced bit rate, and the small ΔR indicates that the QP-produced bit rate is closer to the budget, and vice versa.

Table 2 presents detailed numerical simulation results, showing that the proposed algorithm can provide excellent performance, up to 0.19 dB and 1.28 dB better PSNR than JVT-G012 and Yan and Wang [10], respectively, while the output bit rate is also close to the target budget. This table also displays two gains, Gain 1 and Gain 2, to depict the performance improvement over JVT-G012 and [10], respectively. Compared with JVT-G012 and [10], the proposed algorithm improves results in all test sequences. These results show that the proposed algorithm produces lower ΔR results than JVT-G012 for the News sequence. Table 2 also shows that the proposed algorithm is capable of controlling precision for the target bit rate. Moreover the proposed algorithm presents better image quality than both JVT-G012 and [10]. Furthermore, the bit rate variation is near equal to JVT-G012 and [10]. Thus, the proposed algorithm can be applied to real-time multimedia data streaming and can produce excellent image quality.


(a)

(b)

(c)

Figure 8. Visual comparison of Yan and Wang [10] (left) and the proposed algorithm (right) in three different video sequences: (a) “Foreman,” (b) “News” and (c) “Highway.” The frame is in the 56$^{th}$, 91$^{st}$ and 96$^{th}$ location, from the top.

Worth noting is that since all the inter-frame QP estimates in [10] are unsuitable, the inter-frame should increase QP to balance the overall bits target. To further improve fitting the algorithm to real visualization application, Figure 8 shows the results of decoding frames with the [10] versus the proposed algorithm in three test video sequences, the 56$^{th}$ frame in “Foreman”, the 91$^{st}$ frame in “News,” and the 96$^{th}$ frame in “Highway” respectively. The moving objects include the face in Figure 8(a), the dancer and reporter in Figure 8(b), and the road sign in Figure 8(c), though they are blurry, and the block effect is especially clear. The proposed algorithm can evidently improve the image quality compared to the currently used algorithm.

## 2.4 Summary

This algorithm presents a variance-based intra-frame rate control for H.264/AVC. The purpose of this algorithm is to make use of the advantage in the relationship between distortion and QP, in order to develop a more efficient method for intra-frame rate control. Experimental results show that the proposed algorithm can simultaneously improve both the PSNR and the bit rate. This algorithm is also constructed on a one-pass scheme. Numerous simulation results show that the proposed algorithm can also achieve a higher average PSNR and lower bit rate mismatch. The balanced bit allocation and control are suitable for strict network environment. Thus, this algorithm is better suited to H.264/AVC rate control than existing algorithms.

# Chapter 3 A Rate-Distortion Optimization Inter-Frame Rate Control Algorithm

This study proposes a joint rate-distortion optimization for the H.264/AVC rate control algorithm with a novel distortion prediction equation. First, based on the original distortion theorem, we produced a novel, analytically derived equation for the prediction of distortion. This novel distortion predictor avoids linear regression employed in other distortion predictors, and therefore can considerably speed up rate estimation. Then, this work constructed an optimal quantization parameter selection for encoder processing on the new distortion prediction equation and the JVT-G012 rate estimation equation at block level. Simulations demonstrated that the proposed algorithm had achieved an improvement in image quality and a computational saving of at least 34% over two existing algorithms, including the JVT-G012, which was the recommended rate control algorithm in the H.264/AVC reference software JM15.0

## 3.1 Review of Related Articles

In real time application, [13] proposed a single-pass distributed rate control algorithm to reduce the distortion variation for H.264/AVC high definition video coding at different frame types. A low-complexity decision rule in current frame QP is the main contribution by its relationship with all previous QPs of the frames. Even if the distortion is given exact accuracy in established relation to predict a suitable QP, the bit allocation may not be able to be achieved under a specified target and vice versa [14], [15]. References [15], [18] and [20] have the same rate model because in [15] the rate model is verified, adopted for H.264/AVC reference software [7], and

only different in distortion model. In [15], [18] and [20], the distortion model is expressed in (23), (24) and (25) respectively.

$$D = \frac{Qstep^2}{12} \tag{23}$$

$$D = \frac{D^{prev}}{Qstep^{prev}} \times Qstep \tag{24}$$

$$D = \rho \times Qstep \tag{25}$$

where the $D$ and $D^{prev}$ denote current and previous distortion respectively, $Qstep$ and $Qstep^{prev}$ denote the current and previous QP step respectively, and the term $\rho$ is the model parameter. Although previous works have attempted to perform a suitable distortion model, these were established under large databases experiment or depended on previous encoding information and image content in the current MB was not used.

## 3.2 Our Proposed Algorithm

### 3.2.1 Dedication of a Novel Rate Prediction Equation

The well-known quadratic R-Q model used in H.264/AVC reference software models is a second-order polynomial and the MAD is selected for distortion estimation. The detailed mathematical description is as follows:

$$\begin{cases} R_i = MAD \times \left( \frac{\beta}{Qstep} + \frac{\alpha}{Qstep^2} \right) + H^p \\ MAD = \phi \times MAD_{previous} + \varphi \end{cases} \tag{26}$$

where $\alpha$, $\beta$, $\phi$ and $\varphi$ are four model parameters, and MAD and $MAD_{previous}$ indicate current and previous MB distortions for this model to solve the

chicken-and-egg dilemma. The term $R_i$ is the number of bits used for texture encoding of the $i^{th}$ block. These four parameters will be updated by the linear regression method. The term $H^p$ is the predicted header bits budget of the current MB.

In detail, the model parameters and MAD are estimated dynamically. Because the image content is varied and the model is based on statistical modeling, the estimated results are often inaccurate. Inaccurate estimated rate model parameters will lead to large oscillations of future estimated rate model parameters [16]. To improve rate model parameter estimation accuracy, a content-adaptive prediction scheme is proposed [17]. Two main contributions are proposed, including a one-order linear R-Q model, and the modified sliding window size and adaptive MB location selection for MAD prediction. Abundant testing shows that an efficient estimation for model parameters is very important when selecting linear regression. The results in [17] indicate that the small magnitudes in model parameters perform more accurate rate estimation. Even though a new scheme [17] is provided for model parameter estimation, the large computation in linear regression is still required. To reduce computation, we propose a simple distortion prediction equation to replace the original linear regression. From the above analysis, this study deduces a well-defined equation that only includes two variables of variance and QP. Not only the linear regression is avoided in distortion prediction, but also the term of variance is calculated in the current MB. In real-world applications, the variance can be calculated directly without the chicken-and-egg problem and the QP of previous MB, $QP_{previous}$, is used. Finally, the new rate prediction equation is presented in (27) and only the quadratic form for rate prediction needs to execute the linear regression method.

$$\begin{cases} R_i = D \times \left( \frac{\beta}{Qstep} + \frac{\alpha}{Qstep^2} \right) + H^p \\ D = \sigma^2 (1 - QP_{previous}^{-2}) \end{cases} \quad (27)$$

This equation offers two important advantages; the inter-frame domain transforms to the intra-frame domain and the large linear regression computation can be canceled in distortion prediction. Because intra-frame information is used, a more sensitive response to image content can help with R-D optimization in obtaining a suitable QP.

### 3.2.2 Calculation of QP determination based on R-D optimization

A well-known factor of high efficiency compression is that H.264/AVC supports 52 QP values, which is a real-world application for encoding processing. In other words, the quadratic equation in (27) uses *Qstep* to obtain the predicted rate. The definition of the relationship between QP and *Qstep* is shown as follows,

$$Qstep = Qstep(QP \bmod 6) \times 2^{floor(QP/6)} \quad (28)$$

Because *Qstep* sizes are not defined in H.264/AVC, the first six values are pre-assigned in the real-world application [19]. The six initial values are 0.625, 0.6875, 0.8125, 0.875, 1, and 1.125. After, *Qstep* doubles in size for every six-unit increment in QP. Although the QP and *Qstep* are not linear mapping, the one-to-one property remains true [21]. Thus, we assume that the QP is equal to *Qstep* in the deduction of R-D optimization.

$$D = \sigma^2 (1 - Qstep^{-2}) \quad (29)$$

Assuming the target bits T for residual MB of a frame, a set of optimal quantization parameters is calculated by minimizing the distortion in (29). The minimization problem can be expressed as

$$Qstep_1^*,...,Qstep_N^* = \underset{\substack{Qstep_1,...,Qstep_N \\ \sum_{i=1}^{N} R_i = T}}{\arg\min} \frac{1}{N}\sum_{j=i}^{N}\sigma_j^2(1-Qstep_j^{-2}) \tag{30}$$

where N is the total number of MBs in a frame and the i=1 is defined initially. Through the Lagrange multiplier method and the R-Q model in (27), we can rewrite the optimization problem in (30) in the same expression form as (31)

$$\begin{aligned} & Qstep_1^*,...,Qstep_N^*,\lambda^* \\ &= \underset{Qstep_1,...,Qstep_N,\lambda}{\arg\min} \left\{ \frac{1}{N}\sum_{j=i}^{N}\sigma_j^2(1-Qstep_j^{-2}) + \lambda(\sum_{j=i}^{N} R_j - T) \right\} \\ &= \underset{Qstep_1,...,Qstep_N,\lambda}{\arg\min} \left\{ \frac{1}{N}\sum_{j=i}^{N}\sigma_j^2(1-Qstep_j^{-2}) \right. \\ &\left. + \lambda\left[ \sum_{j=i}^{N}(H_j^p + \alpha_j D_j Qstep_j^{-2} + \beta_j D_j Qstep_j^{-1}) - T \right] \right\} \end{aligned} \tag{31}$$

where $H_j^p$ are the predicted header bits budget of the j[th] MB and its values are equal to the previous MB at the same MB location and frame type. Because the rate model only predicts text bit counts, we need to estimate the header prediction, which is proposed in [18]. By taking partial derivatives of (31) for all parameters and setting them to be zero, we can obtain the quantization parameter step. After mathematical manipulations, a closed-form optimal quantization parameter step size is obtained as follows.

$$Qstep_i^* = \left\{ -\frac{\beta_i}{2\alpha_i} + \frac{\sigma_i^2}{\alpha_i D_i} \sqrt{\frac{T + \sum_{j=i}^{N} \left( \frac{\beta_j{}^2 D_j}{4\alpha_j} - H_j^p \right)}{\sum_{j=i}^{N} \alpha_j{}^{-1} D_j{}^{-1} \sigma_j^{-4}}} \right\}^{-1} \tag{32}$$

where $1 \le i \le N$. It is important that the encoding the i[th] MB, two parameters $\alpha_j$ and $\beta_j$ for $i \le j \le N$ are non-variable. To solve this problem, we approximate the corresponding parameters $\alpha_i$ and $\beta_i$. We can pre-calculate $\sigma_k$ for $1 \le k \le N$ after all MBs. Finally, an approximate formula (33) is simplified as

$$Qstep_i^* \cong \left\{ -\frac{\beta_i}{2\alpha_i} + \frac{\sigma_i^2}{D_i} \sqrt{\frac{T + \frac{\beta_i{}^2}{4\alpha_i} \sum_{j=i}^{N} (D_j - H_j^p)}{\alpha_i \sum_{j=i}^{N} D_j{}^{-1} \sigma_j^{-4}}} \right\}^{-1} \tag{33}$$

### 3.2.3 Proposed Rate control algorithm at MB-layer

In H.264/AVC reference software, each first frame is given an initial reference QP value that denotes the $QP_{ave}$ as the average QP value of the previous frame with the same frame type. For all P-frames, the QP is optimally evaluated for all MBs to fit the target bit budget T. When the target T is negative or $\alpha_i$ is zero, the initial QP value is referenced to increase QP value as the predicted QP for current MB encoding. Moreover, when $T - \sum_{j=i}^{N} H_j^p < 0$ is true then decrease QP because the residual bits are not enough after encoding and if $\alpha_i > 0$ and $T + \frac{\beta_i^2}{4\alpha_i} \sum_{j=i}^{N} (D_j - H_j^p) > 0$ are true then we calculate the optimal $Q_{step}^*$ by (27) and transfer it to QP. To maintain smooth visual quality, the QP is modified as QP = max{ $QP_{ave}$ -2, min{ $QP_{ave}$ +2, QP } }. To fit the H.264/AVC specification for the distribution range of quantization parameters, it is bounded by QP = max{ 1, min{ 51, QP } }. A detailed flowchart of the proposed rate control algorithm on the MB layer is illustrated in Figure 9.

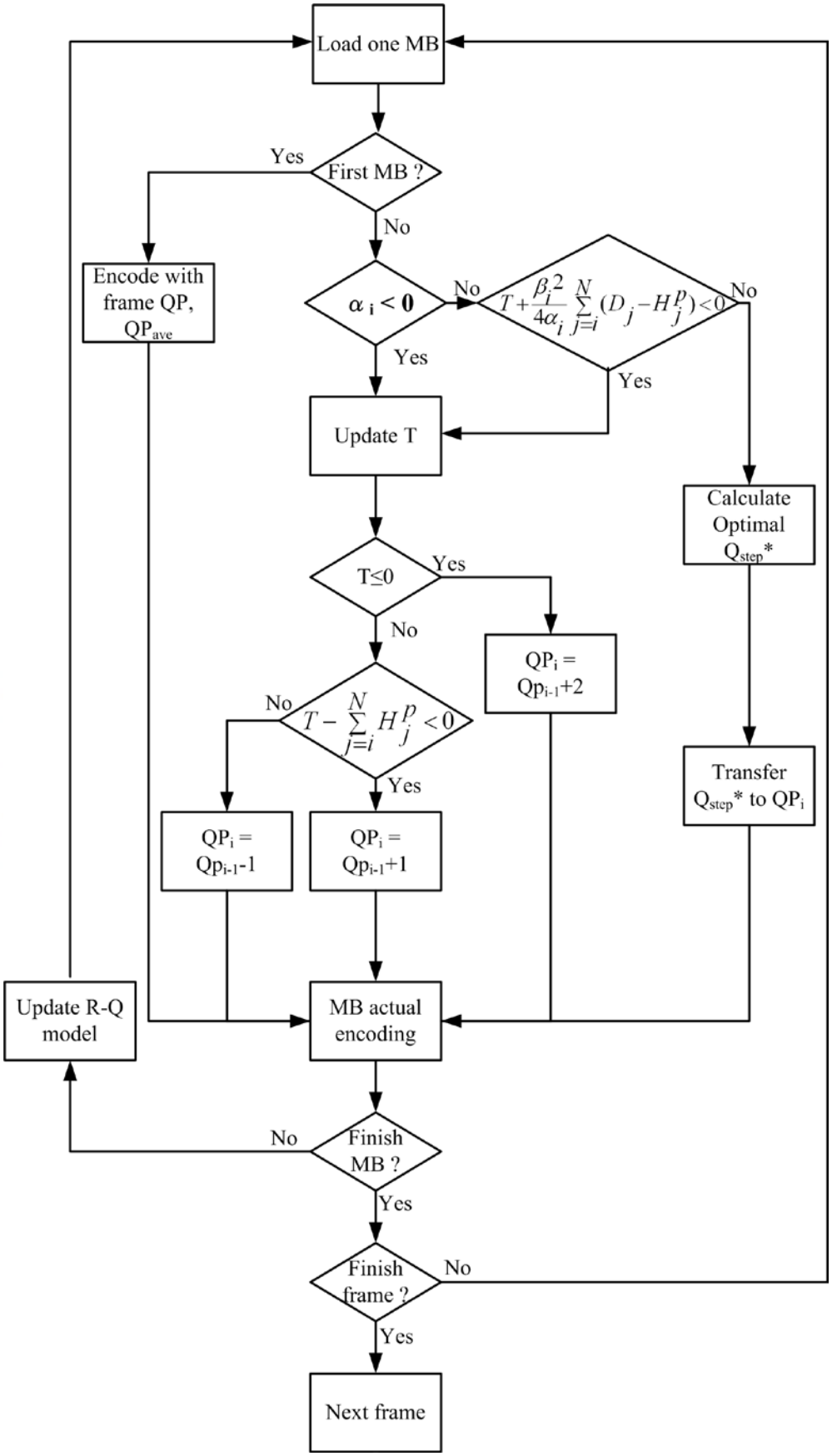
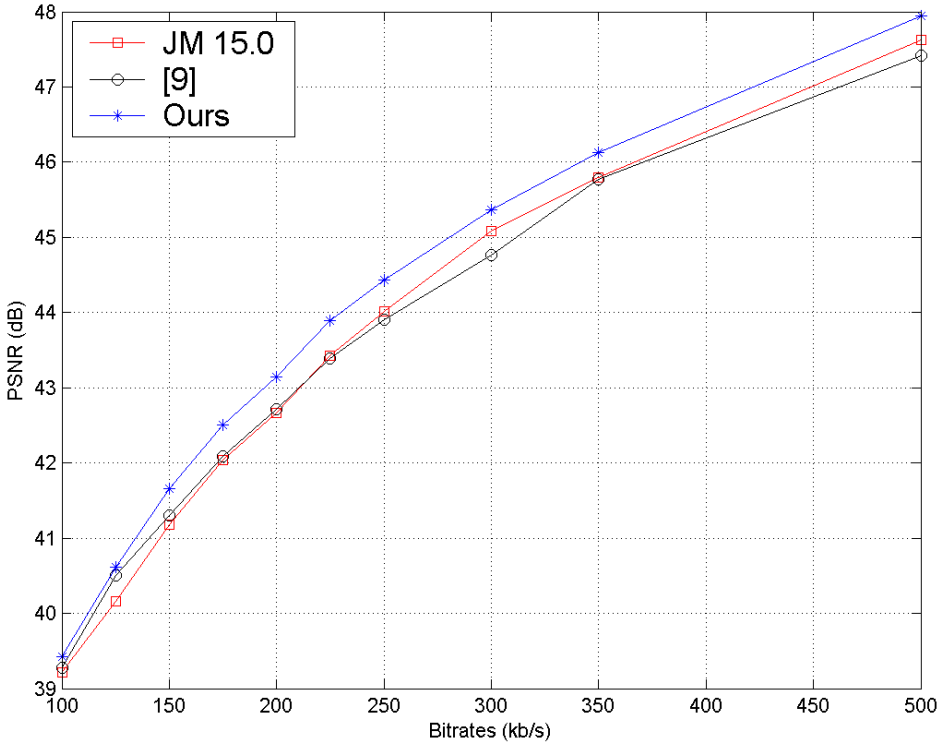



Figure 9. Flowchart of the proposed rate control algorithm at the MB layer.

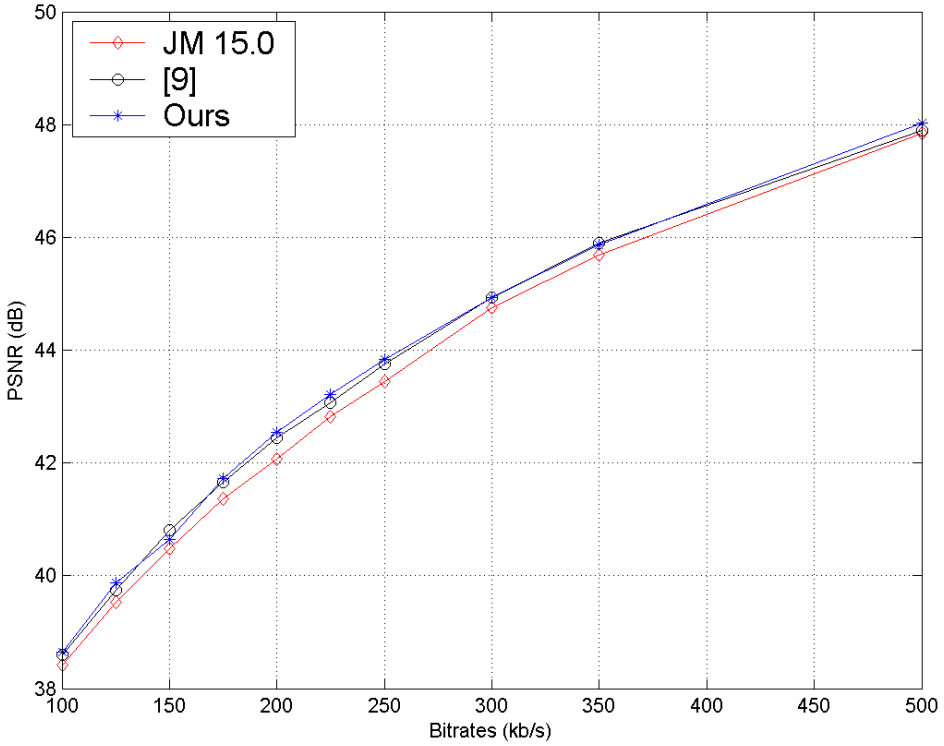
## 3.3 Simulation Results

The current investigation implemented the proposed rate control algorithm and used the existing rate control algorithm in JM15.0 [7] on a 2.33 GHz PC, with 4 GB memory as a benchmark for comparison. The JM version provided four modes for rate control and we set mode 0. After rate distortion optimization was enabled, we selected the structure of 1-I and 99-P frames, as a fair comparison. The configuration file, encoder.cfg, in JM15.0 was used for residual parameter settings. Five different testing sequences in the same image resolution quarter common intermediate format (QCIF), including “Carphone”, “Silent”, “Salesman”, “Bridge-close” and “News” were selected, and ten different bit rate including 100, 125, 150, 175, 200, 225, 250, 300, 350, 500 kbps and four different initial QPs including 22, 24, 28, 32 were evaluated for respectively wider comparison.

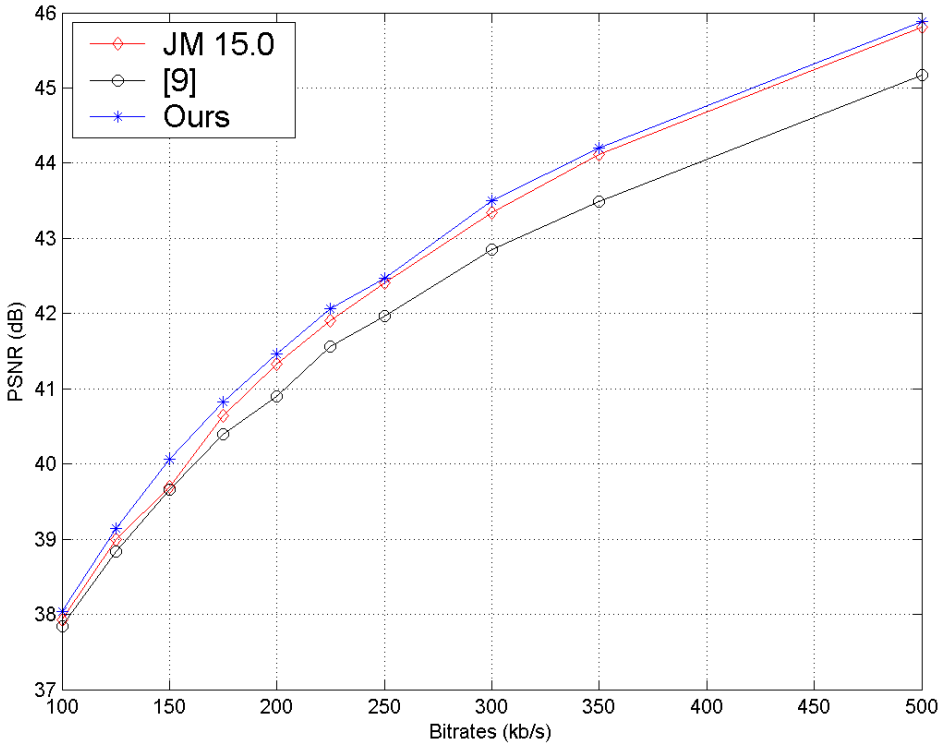
In Figure 10, three testing video sequences for “Salesman”, “Silent”, and “Carphone” were selected. We plotted the overall rate-distortion performance of the proposed rate control algorithm with the JM15.0 in different bit rate constraints. For a fairer comparison, another optimized method in [9] was also implemented. Findings show that the curve of our proposed rate control algorithm lay above JM15.0 and [9] at all different bit rates, meaning that the algorithm was suitable for a application with wide bit rate range. In Figure 10(a) and (c), the peak signal-to-noise ratio (PSNR) of [9] was lower than that of JM15.0 when the high bit rate condition occurs, therefore only our proposed algorithm was suitable for different conditions.


JM 15.0
[9]
Ours
PSNR (dB)
Bitrates (kb/s)
48
47
46
45
44
43
42
41
40
39
100
150
200
250
300
350
400
450
500

(a)


JM 15.0
[9]
Ours
PSNR (dB)
Bitrates (kb/s)
50
48
46
44
42
40
38
100
150
200
250
300
350
400
450
500

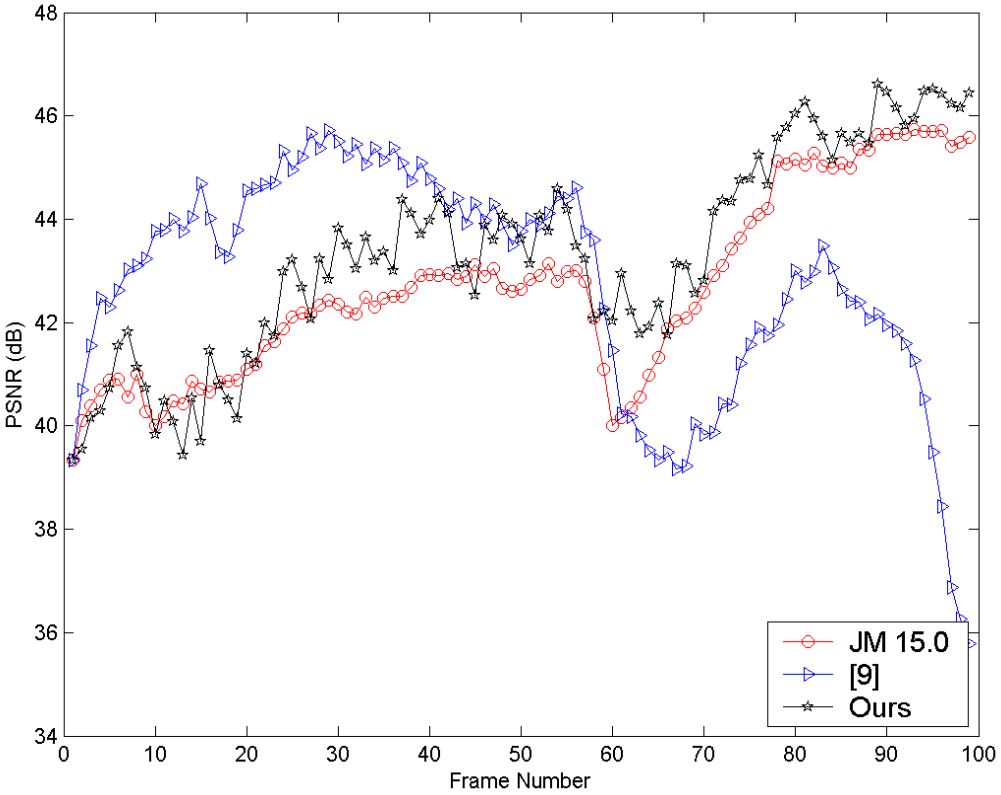
(b)



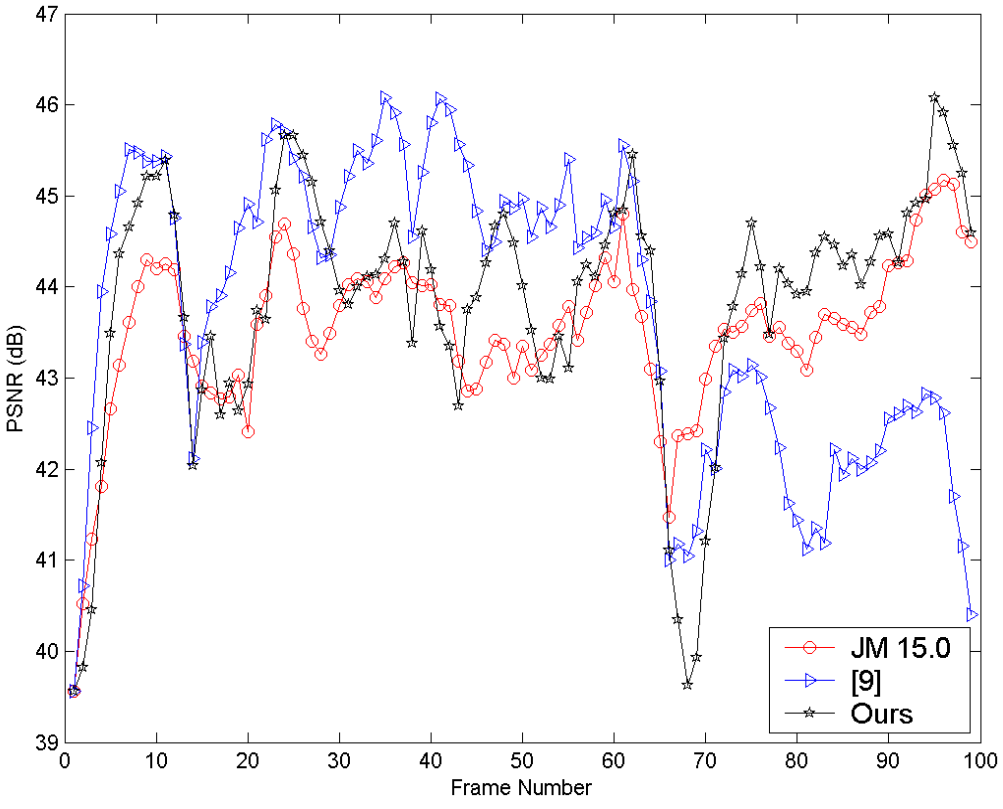

(c)

Figure 10. PSNR comparisons of three algorithms over a range of bitrates. (a) “Salesman,” with initial QP 24. (b) “Silent,” with initial QP 24. (c) “Carphone,” with initial QP 22.

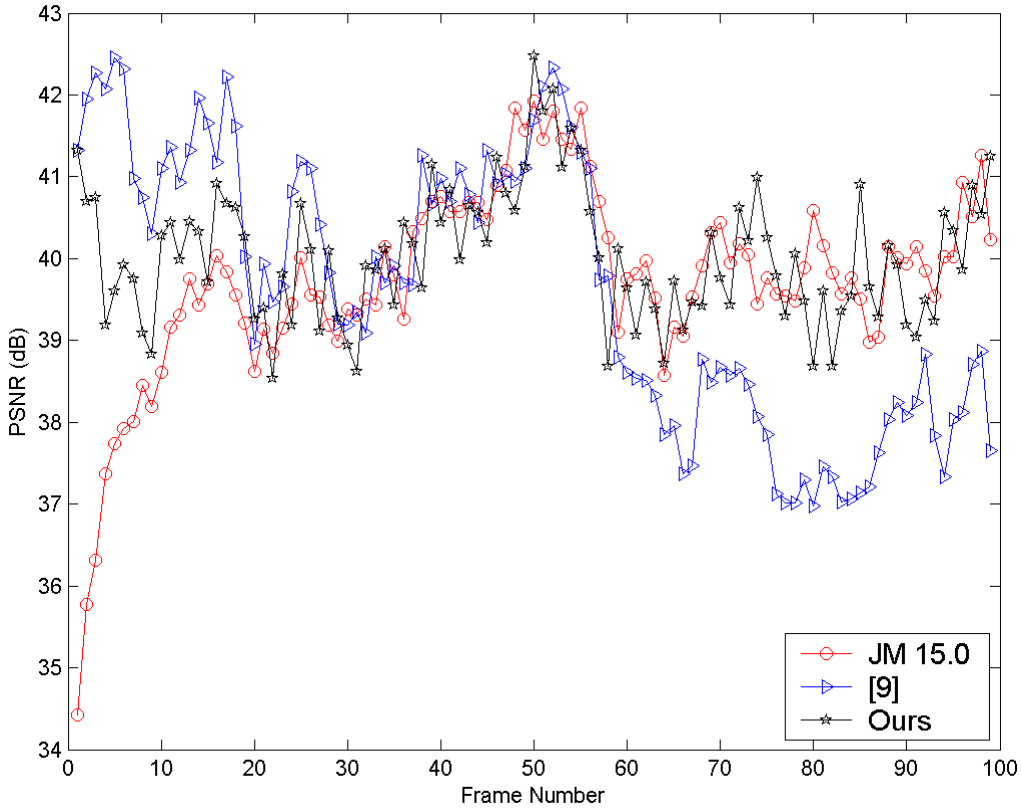
For a specified bit rate condition such as 150, 200 and 250 kbps, Figure 11 depicts the frame-by-frame PSNR comparison. The PSNR of our algorithm was maintained in smooth status and holds on to excellent performance in Figure 11 because the distortion prediction equation was more adaptive and balanced for a bit budget allocation. Although [9] provided good performance in the first fifty frames presented, a rapid falloff was exhibited after the fifty frames appeared. The main reason was that variations in image content were not considered in [9]. Furthermore, in Figure 11 (b), our proposed algorithm improved image quality from the 7th to 15th frame and the 45th to 60th frame, with enhanced smoothing from the 65th to 75th frame. This was because the “silent” video sequence had large local motion in the foreground object and accurate QP selection improved the whole image quality.


PSNR (dB)
Frame Number
JM 15.0
[9]
Ours

(a)


PSNR (dB)
Frame Number
JM 15.0
[9]
Ours

(b)




(c)

Figure 11. Frame by frame PSNR comparison. (a) “Salesman,” 200 kbps, initial QP is 24. (b) “Silent,” 250 kbps, initial QP is 24. (c) “Carphone,” 150kbps, initial QP is 22.

Table 3 tabulates the PSNR and bit consumption for general comparison. The table shows that the proposed rate control algorithm achieved acceptable performance in comparison with JM15.0 and the gain was up to 0.63 dB. The average PSNR gain of the proposed algorithm was 0.37 dB for “Carphone”, 0.43 dB for “Silent”, 0.59 dB for “Salesman”, 0.20 dB for “Bridge-close” and 0.32 dB for “News” compared with JM15.0. Our performance is also better than [9] in almost all test sequences. The property of smoothing image is another issue, though, since the lower vibration provides better human vision. To measure the vibration of the image, this study compared the PSNR standard deviation (PSNR Std.) of the three algorithms, with average values of approximately 1.16, 1.76 and 1.39 for JM15.0, [9] and the proposed algorithm, respectively. The PSNR Std. of the proposed algorithm was in-between

those of the other two, and closer to the better one. Although our PSNR Std. is slightly larger than JM15.0, the degree is very small and can be ignored in human vision system.

Table 3. Comparison of different bit rate limitation and QP between JM 15.0 and proposed method in psnr, psnr std. and encoded bits

| **Average PSNR (dB)** | | | | | |
|---|---|---|---|---|---|
| **Video sequences** | **Initial QP** | JM 15.0 | [9] | Ours | Gain |
| Carphone(150kbps) | 22 | 39.69 | 39.66 | 40.06 | +0.37 |
| Silent(200kbps) | 24 | 42.07 | 42.45 | 42.54 | +0.47 |
| Salesman(200kbps) | 24 | 42.66 | 42.71 | 43.29 | +0.63 |
| Bridge-close(300kbps) | 24 | 38.93 | 39.36 | 39.13 | +0.20 |
| Silent(250kbps) | 24 | 43.45 | 43.75 | 43.84 | +0.39 |
| Salesman(200kbps) | 28 | 42.45 | 42.73 | 42.99 | +0.54 |
| News(100kbps) | 28 | 38.95 | 38.79 | 39.27 | +0.32 |

| **PSNR Std.** | | | | |
|---|---|---|---|---|
| **Video sequences** | **Initial QP** | JM 15.0 | [9] | Ours |
| Carphone(150kbps) | 22 | 1.18 | 1.65 | 0.81 |
| Silent(200kbps) | 24 | 0.84 | 0.93 | 1.19 |
| Salesman(200kbps) | 24 | 1.75 | 2.19 | 1.96 |
| Bridge-close(300kbps) | 24 | 0.28 | 0.82 | 0.42 |
| Silent(250kbps) | 24 | 0.88 | 1.61 | 1.35 |
| Salesman(200kbps) | 28 | 2.21 | 2.18 | 2.73 |
| News(100kbps) | 28 | 1.00 | 2.99 | 1.29 |

| **Encoded bits (kbps)** | | | | |
|---|---|---|---|---|
| **Video sequences** | **Initial QP** | JM 15.0 | [9] | Ours |
| Carphone(150kbps) | 22 | 150.32 | 150.89 | 151.30 |
| Silent(200kbps) | 24 | 200.18 | 201.68 | 201.11 |
| Salesman(200kbps) | 24 | 199.83 | 200.24 | 199.83 |
| Bridge-close(300kbps) | 24 | 301.44 | 302.66 | 301.73 |
| Silent(250kbps) | 24 | 249.98 | 251.83 | 251.10 |
| Salesman(200kbps) | 28 | 199.84 | 200.69 | 199.69 |
| News(100kbps) | 28 | 100.32 | 100.60 | 100.13 |

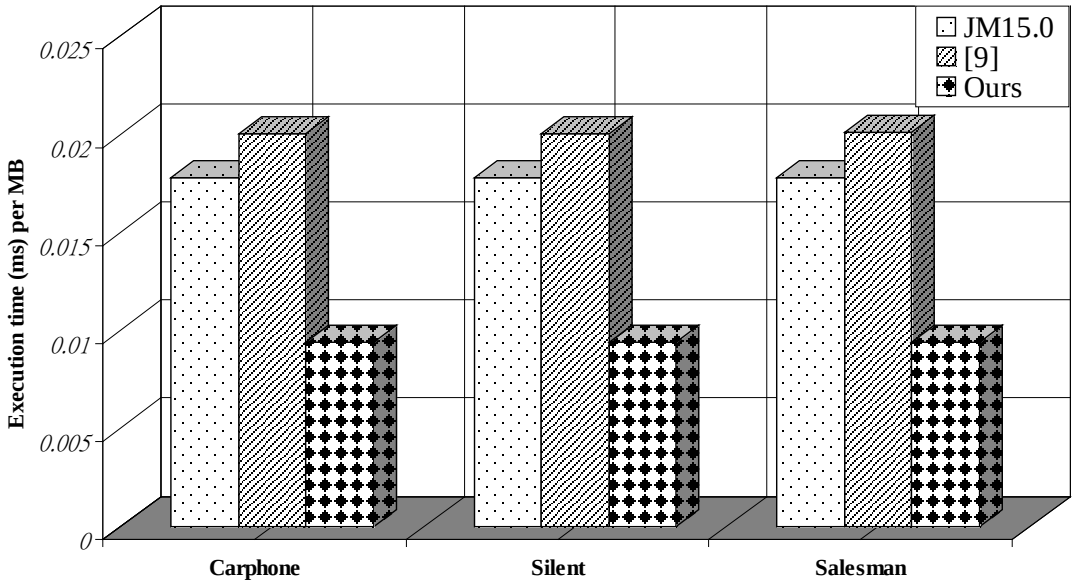


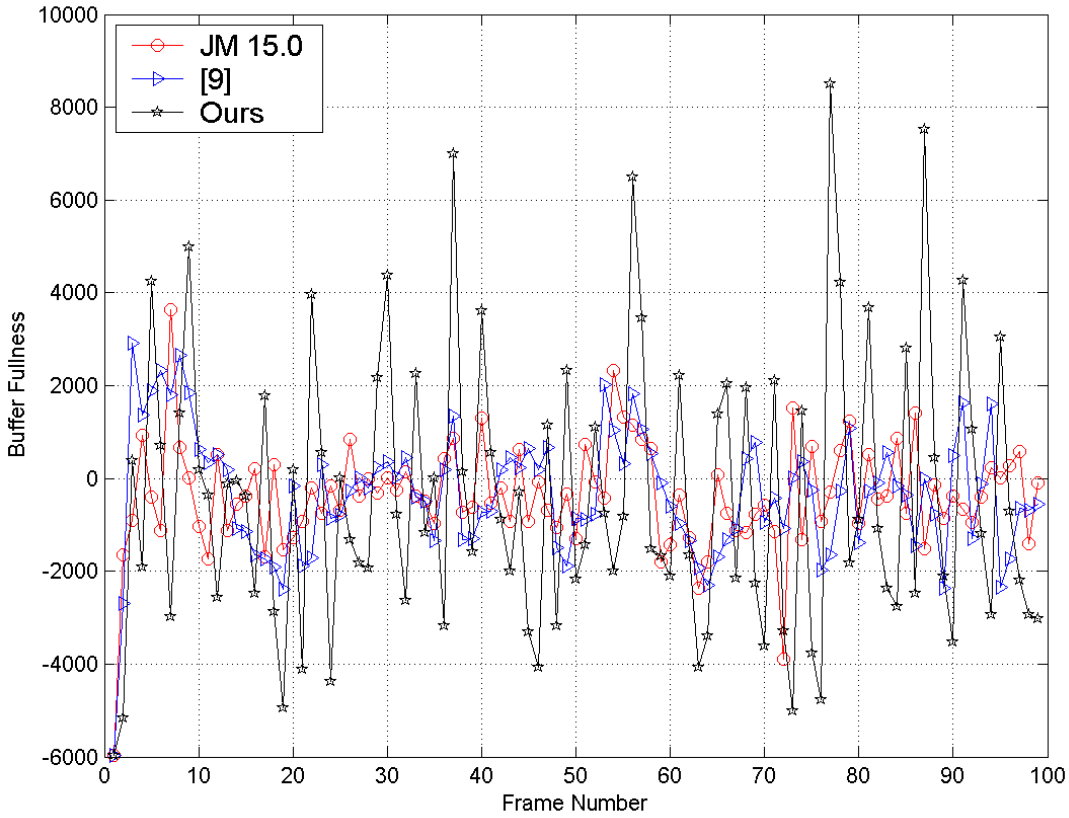

Figure 12. Execution time comparison per MB in three different video sequences.

Figure 12 displays the computational complexity of the three different rate-control algorithm in terms of average CPU execution time (unit: ms) per MB under three different video sequences. The results show that our proposed rate control algorithm saved approximately 34% and 53% of computation time compared with JM15.0 and [9] respectively. And this saving holds for all three video sequences tested, indicating an assured computational advantage over other algorithm when applied to real-world applications. The reason that JM15.0 and [9] require more computation time is due to the fact that they need to generate distortion parameter estimations, i.e. $\phi$ and $\varphi$ in (26), through linear regression.

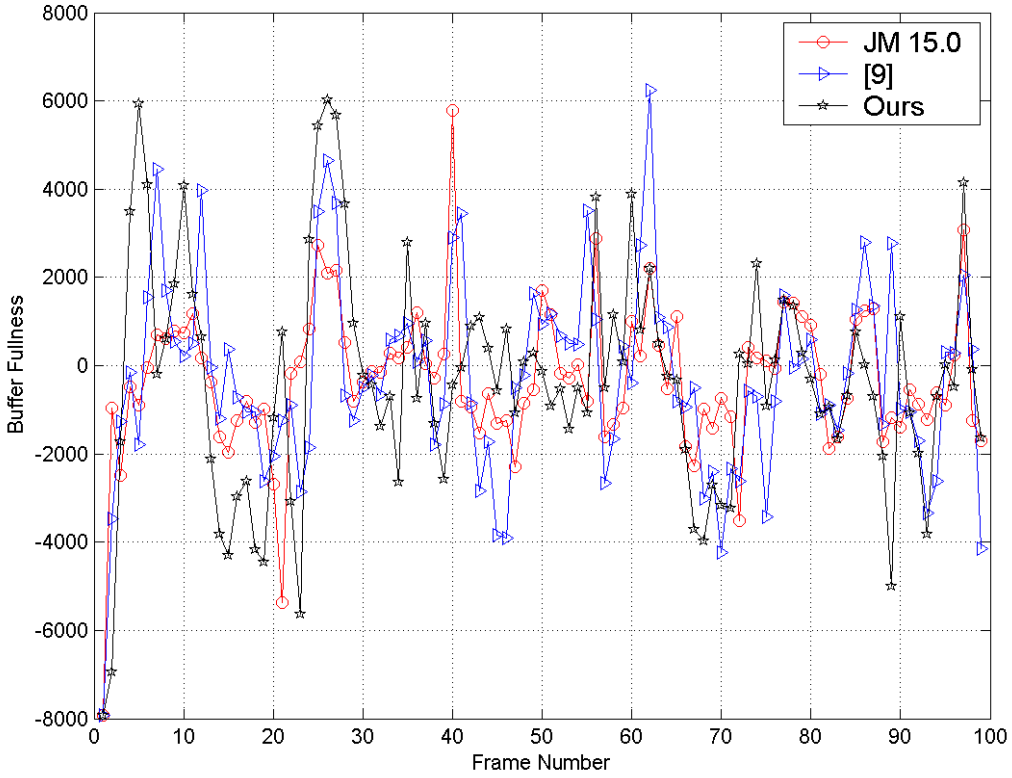
For a more detailed comparison and performance analysis, Figure 13 shows the number of bits in the buffer fullness at each frame. In general, the proposed algorithm has a large vibration of buffer fullness. For a fair evaluation of buffer fullness, we first defined a reasonable buffer size. According to [27]-[28], a recommended buffer size is set to two-thirds (2/3) of the channel bandwidth (one second). With bit budget

constraints of 250, 200, and 150 kbps and a 30-fps condition, the recommended buffer size is set to (250 k/30 fps)×(2/3) = 5.6 k, (200 k/30 fps)×(2/3) = 4.4 k, and (150 k/30 fps)×(2/3) = 3.3 k bits per frame for "Salesman," "Silent," and "Car phone" video sequences, respectively.

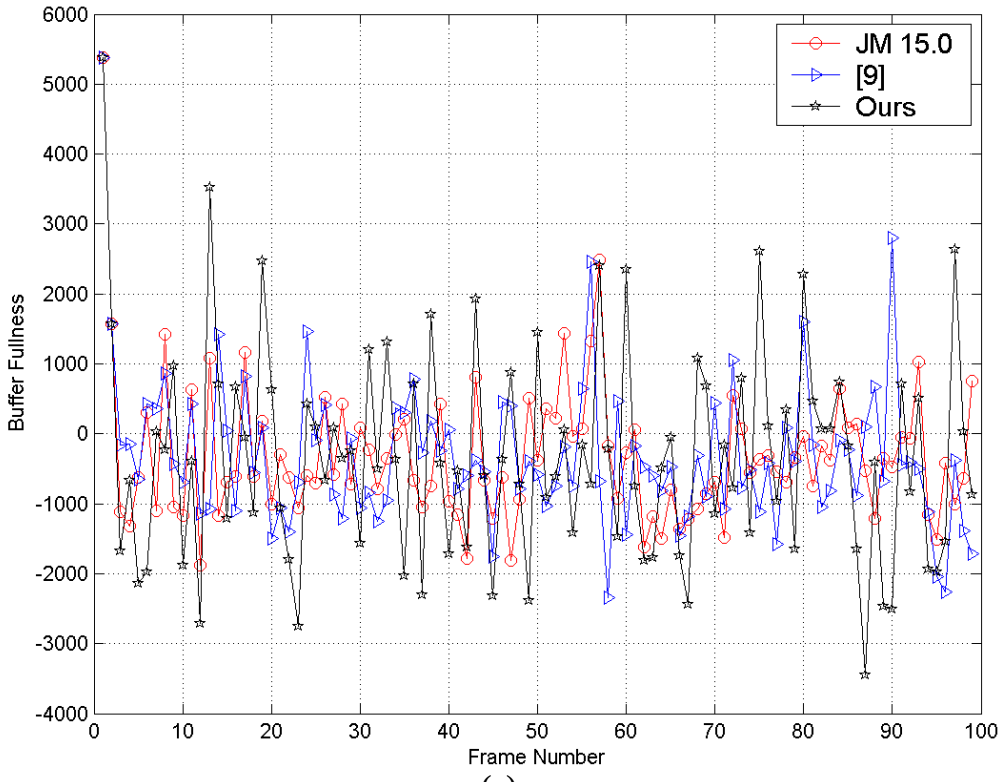
Specifically, the numbers of over-buffer fullness constraint are the 38$^{th}$ 58$^{th}$, 78$^{th}$, and 88$^{th}$ frames in Figure 13(a), the 5$^{th}$, 25$^{th}$, 26$^{th}$, and 27$^{th}$ frames in Figure 13(b), and the 1$^{st}$ and 12$^{th}$ frames in Figure 13(c), respectively. Even the drawback occasionally results in frame delay in real-time transmissions; the result is acceptable in low delay from the overall encoded frames.




(a)




(b)




(c)

Figure 13. Buffer fullness comparison of JVT-G012, [10] and the proposed algorithm, on three different video sequences: (a) “Salesman,” 200 kbps, initial QP is 24. (b) “Silent,” 250 kbps, initial QP is 24. (c) “Carphone,” 150kbps, initial QP is 22.

## 3.4 Summary

This chapter proposes a rate-distortion optimization of rate control with a novel distortion prediction equation. The two main contributions include the novel distortion prediction and the optimization of QP selection. Our distortion prediction method ignored the traditional linear regression, and was able to reduce prediction error and decrease computation. Based on this distortion prediction model, the optimized QP selection is proposed with joint rate-distortion optimization. Simulation results showed that the proposed rate control algorithm had an improved PSNR gain over JM15.0 and [9], under the condition that bit consumption is limited under a specified budget. Finally, we propose that algorithm could be implemented for real time applications requiring the optimization of rate-distortion thanks to its avoiding linear regression in distortion prediction.

# Chapter 4 An Adaptive Content-based Inter-Frame Rate Control Algorithm

In this chapter, we analyze a rate control formula, which is Laplacian based model, then propose a content based equation for an enhanced JVT-G012 algorithm. Based on the simple and fast characteristics of the JVT-G012 algorithm, which is very suitable for real time embedded systems, we integrate the equation into the JVT-G012 algorithm to create an efficient quantizer. However, active local motion will produce error distortion and influence the determination of the quantization parameter. To solve this problem, a modified histogram based detector is proposed. Combining this detector with the equation refines the predicted quantization parameter more precisely. It is shown both theoretically and experimentally that a well designed quantizer not only improves video quality, but also maintains the designed bit rate budget. In addition, our algorithm can achieve superior results to existing algorithms—up to 0.72 dB without large computation loading.

## 4.1 Review of Related Articles

The Lagrange method is used to construct a new R-D model in [23] and [8], which propose an R-D optimized rate control algorithm with an adaptive initial QP determination scheme that differs from the JVT-G012 algorithm quadratic form [6]. Although these models can possible obtain the optimum QP, the correlation between two frames is calculated to modify the initial QP and the computational complexity is enormous. In addition, a near optimization searching method is proposed in [24] that is also based on the Lagrange method to pick up a suitable QP for achieving a more

accurate bit rate budget. However, the complexity of the encoder increases proportionally with the number of iterations, and it depends on the initial QP and relative parameters, such as the Lagrange multiplier, block mode, and reference frames. Because the Lagrange method has a complex computation and cannot suit real time applications, the histogram of difference (HOD) is used to detect which frames are skipped to save encoding time; moreover, the HOD value is taken to modify the bit allocation and update the Lagrange multiplier in [25], [26]. Secondly, the quadratic form, which is assumed as statistical distribution, is used as the main skeleton for more accurately modifying the mean absolute difference (MAD) prediction in [27]-[29] and to find a suitable QP. Both [27] and [28] propose an adaptive MAD prediction method. The frame complexity is the main consideration, and using the empirical MAD ratio to change the QP is their main contribution. Experiments confirm that the estimation error of the quadratic model can be significantly reduced to perform a simplified and efficient form in [29]. The second order term is dropped in the original quadratic model without sacrificing much performance, and the first order term is modified by a model parameter to approximate the original quadratic model and the order of the power can be changed in different frame types, such as I-, P-, and B-frame.

Here, methods with large computational complexity are not used because real time embedded systems have limited resources. Fortunately, there is a simple and fast quadratic form that is an efficient closed form, which was recommended and implemented in the H.264/AVC reference software JM 15.0[7]. Although the QP determination technique used in the JVT-G012 algorithm is state of the art, it is not sufficiently accurate for low bit rate video coding, and does not address the problem of local motion. Even if the JVT-G102 algorithm is selected as the main QP prediction method, the input frame must be assumed to have a specified statistical

distribution. Determining this distribution function becomes an important issue because the frame content varies. To use the simple and fast model, we propose an equation constructed based on an analytic rate distortion function and fed into the JVT-G102 algorithm to refine the initial QP as a middle result. To avoid high computational complexity, the middle QP will be refined again as a final QP by a modified HOD.

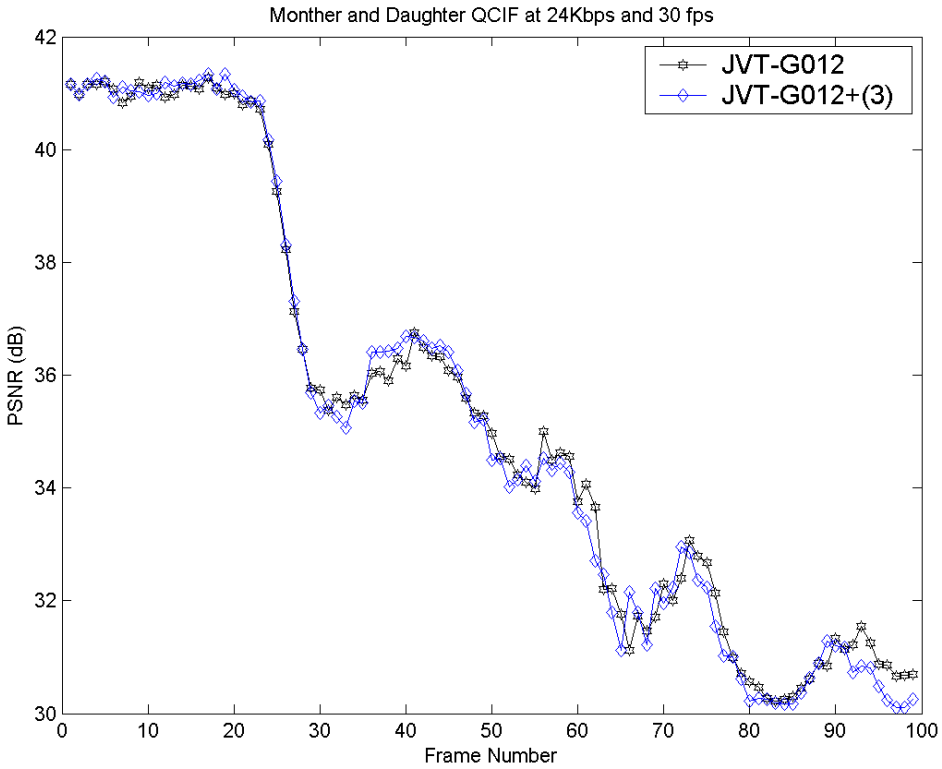
In [1], a mathematical rate distortion was derived and proved in detail. To fit this theorem, however, the selection of a suitable distribution model is important. In previous approaches, the Laplacian model was derived as the rate distortion equation—this has also been used in recent research [6], [27]-[29], thus, this model is worth consideration. From [1] and [48], the relation between rate and distortion can be formulated as

$$R = \ln \frac{1}{\alpha \times D} \tag{34}$$

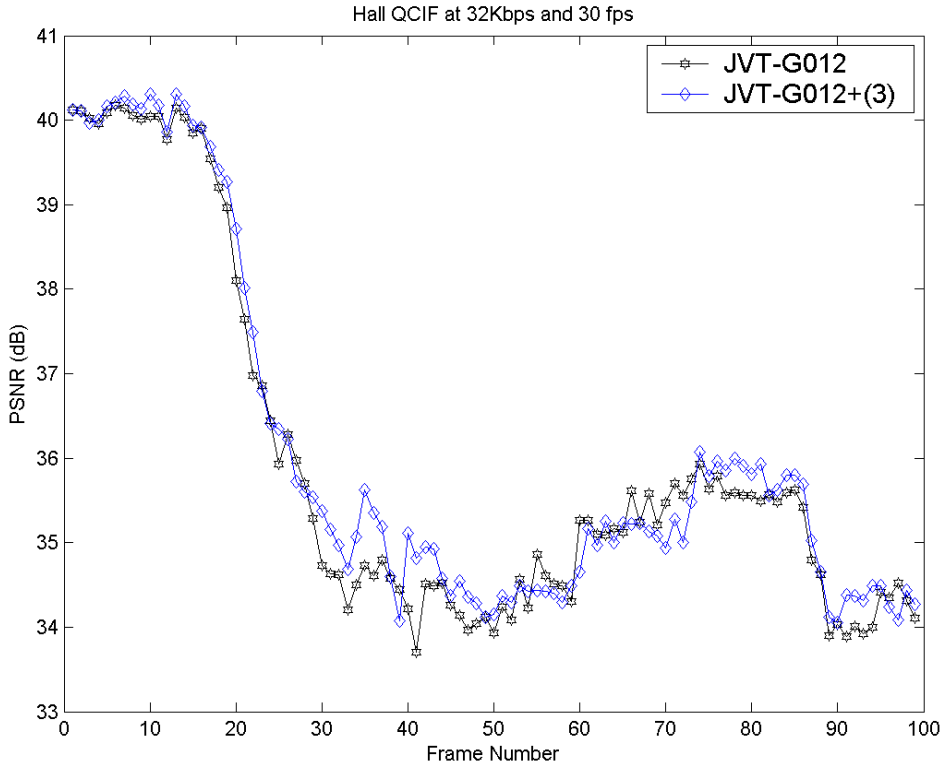
where R, $\alpha$ and D represents the bit rate budget, scale parameter and distortion respectively. Equation (7) can be expanded by a Taylor series to build the new model. It is modified by substituting QP and MAD for D and $\frac{1}{\alpha}$, respectively. The new model, following [5], is formulated as

$$R = X_1 \frac{MAD}{QP} + X_2 \frac{MAD}{QP^2} \tag{35}$$

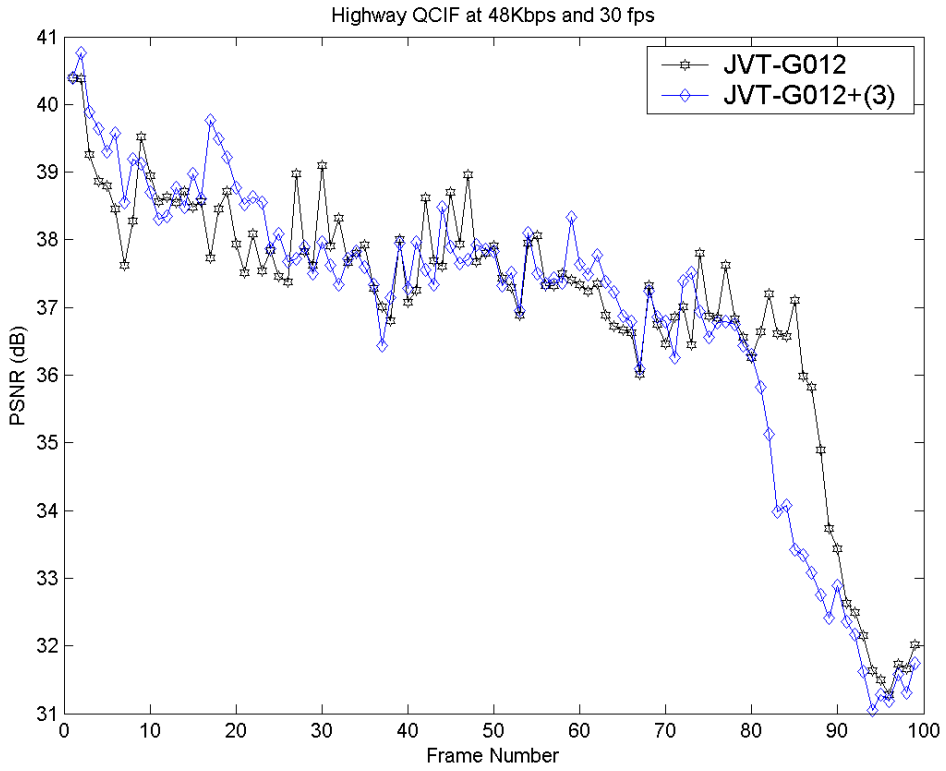
Thus, the QP can be obtained when the coefficients and target bit rate are given; this formula is adopted by the JVT-G102 algorithm. Because it is simple and fast, it is suitable for real time implementation. However, past approaches, such as [25]-[26], [29], show that JVT-G102 cannot provide a suitable solution to accurately estimate the QP of a P-frame at a low bit rate. Even if these approaches solve this problem and provide acceptable results, the computation loading is too large.


Monther and Daughter QCIF at 24Kbps and 30 fps
JVT-G012
JVT-G012+(3)
PSNR (dB)
Frame Number
42
40
38
36
34
32
30
0
10
20
30
40
50
60
70
80
90
100

(a)


Hall QCIF at 32Kbps and 30 fps
JVT-G012
JVT-G012+(3)
PSNR (dB)
Frame Number
41
40
39
38
37
36
35
34
33
0
10
20
30
40
50
60
70
80
90
100

(b)




(c)

Figure 14. PSNR comparison between JVT-G012, and JVT-G012 with Eq. (36) for three standard test sequences: (a) “Mother and Daughter,” (b) “Hall,” and (c) “Highway.”

## 4.2 Our Proposed Algorithm

### 4.2.1 Analyze and verify a Laplacian based model on the JVT-G012 algorithm

Although the JVT-G102 algorithm cannot accurately obtain a QP in low bit rate video immediately, the result is quite close to the real requirements. Even if a two pass algorithm in [29] is proposed to enhance the JVT-G102 algorithm, the computation is still large. Consequently, we propose an efficient equation for the JVT-G102 algorithm that will improve performance and provide a low computational

burden. Equation (34) shows that the relation between rate and distortion can be demonstrated, and the relation between rate and QP is formulated in (35). We can modify (34) to create a new equation as

$$D = \frac{1}{\alpha \times e^{R}} \tag{36}$$

In (3), we assume that the distortion D is variable when bit rate R and let $\alpha = \frac{1}{\sigma}$, variance $\sigma$ are given by current MB, the initial distortion D is predicted by the JVT-G102 algorithm. Note that D can be replaced by the mean square error or mean absolute difference for a fast and simple implementation. A well defined relation [31] is constructed as follows:

$$D \propto QP \tag{37}$$

Thus, we can adjust QP under the predicted term of the target budget to satisfy the equation. But, multiple iterations are not acceptable for real time application in resource limited systems. Nevertheless, the first result in (36) can be used as guidance to refine the initial QP; then we can combine (36) with the JVT-G102 algorithm to create a more efficient rate control algorithm. The JVT-G102 algorithm calculation produces the initial QP, $QP_{init}$, (36) converts it to an inequality to increase or decrease the $QP_{init}$. In general, if $D > \frac{1}{\alpha \times e^{R}}$ then $QP = QP_{init} - 1$; otherwise $QP = QP_{init} + 1$. The initial QP determination for an I-frame is not considered; this only focuses on the P-frame level.

The beginning experimental results, including the average peak signal to noise ratio (PSNR) and bit rates, are given in Table 4. The simulated conditions are described in Section 5.3. It focuses on three parts: the bit rates, the PSNR, and enhanced PSNR. Although this method can improve image quality, only partial test video sequences are efficient.

Table 4. Performance of two methods in term of average PSNR, and bit rate

| Sequence | Method | Average PSNR(dB) | Bit Rate (kbps) |
|---|---|---|---|
| Mother-Daughter (24 kbps) | JVT-G012 | 35.33 | 24.54 |
| | JVT-G012+(3) | 35.28 | 24.60 |
| | Gain | -0.05 | — |
| Hall (32 kbps) | JVT-G012 | 35.99 | 32.18 |
| | JVT-G012+(3) | 36.13 | 32.14 |
| | Gain | +0.14 | — |
| Highway (48 kbps) | JVT-G012 | 37.08 | 48.70 |
| | JVT-G012+(3) | 36.93 | 49.17 |
| | Gain | -0.15 | — |

Figure 14 shows a frame by frame PSNR comparison; in addition, we analyze the drop of PSNR in the visual domain. From the 80$^{th}$ frame to the 99$^{th}$ in Figure 14(c), the PSNR is influenced by the add adjustment equation, (36). In the original video sequence, the object in the frame has active local motion from back to front (Figure 14 (c)). For this reason, a noticeable result that can be addressed is the local motion problem and (30), which can provide smooth frames but does not apply to active local motion. Unfortunately, (36) refines the initial QP from the JVT-G012 algorithm in the wrong direction.

### 4.2.2 Proposed rate control algorithm

Even if multiple iterations are avoided, a single QP determination may be not accurate enough to compare with the real QP and active local motion can reduce the PSNR when (36) is applied. The QP must be decreased to improve the PSNR due to active local motion; the QP should be also increased to improve the bit rate budget when the local image is inactive and motionless. An efficient and fast detection method becomes important. The histogram based methods are considered first, as they are useful and directly detect image content variation [18]. To maintain low complexity

computation and local motion, block level processing and HOD which is defined at frame level operation are considered. A detailed description of HOD is given in [31]. The modified calculation equation, HODB, is given as

$$HODB(b_n, b_{n-1}) = \frac{\sum_{i>|TH|} hod\_block(i)}{N_{pixel}} \tag{38}$$

where i is the index of the histogram bin, the histogram of difference block is from hod_block(i), $N_{pixel}$ is the number of pixels and TH is a threshold for determining the proximity of the position to zero. $b_n$ and $b_{n-1}$ indicate the current block and previous block, respectively. Some excellent results are demonstrated and proven in [29]. From a physical viewpoint, the HODB value is generally increasing monotonically. Generally speaking, the block contains more information if the HODB is nonzero. Although the descriptive algorithm in section A decreases the PSNR during active local motion, it can enhance the PSNR in smooth areas. Thus, we try to keep the property and use the HDOB to avoid the influence of active local motion.

In JVT-G012, the bit allocation only focuses on predefined target bits and buffer status to compensate for bit cost. To meet the real time embedded system, two concepts of data-reuse and lower computation should be considered for bit allocation. Thus, we propose a modified equation in (39) that uses the HODB value for target bit adjustment without large computation. The target bit prediction, $T_{all}$, is defined in [6], [7] and [32].

$$R = (0.9 + HODB \times \varphi) \times T_{all} \tag{39}$$

The parameter φ, obtained from experiments, is set to be

$$\varphi = \begin{cases} 5 & , 0.9 < \text{HODB} \le 1 \\ 0.9 & , \text{HODB} \le 0.9 \end{cases} \tag{40}$$

After, Laplacian based model and reused HODB information to provide efficient modification to improve the initial QP from JVT-G012. The full algorithm can be

described as follows.

*Step 1:* Let R be the target bit budget for current frame, $R_i$ for $i^{th}$ MB and $i$ = 1. N is the total number of MBs in the frame after assigning the Laplacian model.

*Step 2:* Obtain the HODB value according to the previous reconstructed and current frame in the co-located MB. Adjust the $R_i$ adaptively by (39).

*Step 3:* With the JVT-G102 algorithm, use $R_i$ and the predicted MAD to get the initial QP, $QP_{initial}$, by (35).

*Step 4:* Select (36) as an adjustment equation and transfer to an inequality. Calculating variances $\sigma$. If $D > \frac{1}{\alpha \times e^{R_i}}$ then $QP_{tmp} = QP_{initial} - 1$; otherwise $QP_{tmp} = QP_{initial} + 1$.

*Step 5:* Select (38) as an adjustment equation. Calculate the histogram hod_block(i) and set TH to 32. If HODB > 0 then $QP_{final} = QP_{tmp} - 3$; otherwise $QP_{final} = QP_{tmp} + 3$.

*Step 6:* Use $QP_{final}$ to encode the current MB $i$ and set $i = i+1$. If $i \leqq$ N, go to step 2 to encode next MB. Otherwise stop the encoding procedure.

Overall, an efficient formula can be summarized as in (41). From the viewpoint of a real time embedded system, (41) can be implemented in parallel because the input data is independent and has a low computational complexity.

$$QP_{final} = \begin{cases} QP_{initial} + 4, D < \dfrac{1}{\alpha \times e^{R}} & and \quad HODB \le 0 \\ QP_{initial} + 2, D > \dfrac{1}{\alpha \times e^{R}} & and \quad HODB \le 0 \\ QP_{initial} - 2, D < \dfrac{1}{\alpha \times e^{R}} & and \quad HODB > 0 \\ QP_{initial} - 4, D > \dfrac{1}{\alpha \times e^{R}} & and \quad HODB > 0 \end{cases} \tag{41}$$

For a clearer description, Figure 15 depicts integrating our algorithm into the JVT-G012 block diagram. In addition to the shadow blocks in rate control in the proposed functions, all other functions are built in the original H.264/AVC encoder system. The main goal of rate control is to provide a more accurate QP for quantization. The computation complexity of QP selection should be limited to meet real time applications. The implementation also considers the memory requirement in the embedded system due to limited hardware resource. Figure 15 shows directly calculating the HODB without extra buffer allocation and sending the HODB value to the bit allocation adjustment and QP adjustment, after deciding the final target bit rate and final QP. Thus, the proposed algorithm is very suitable to integrate on the current H.264 encoder embedded system.

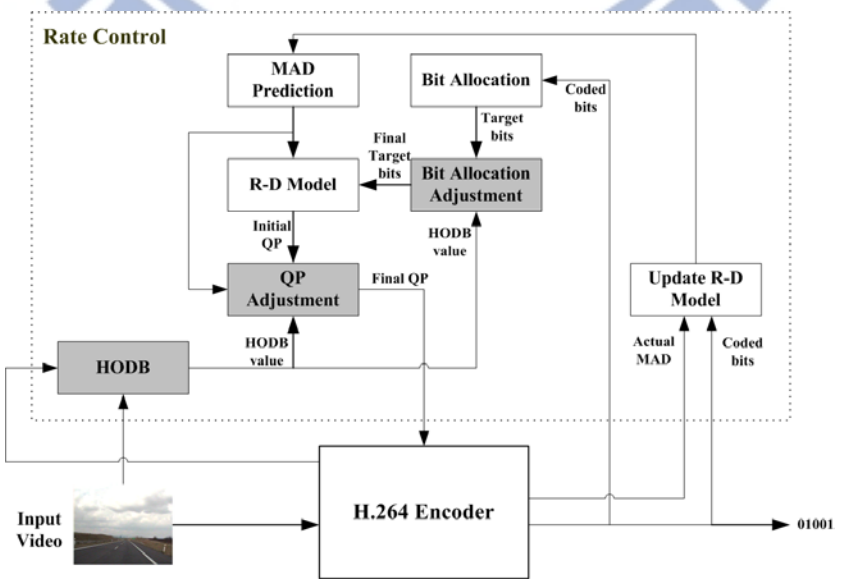



Figure 15. Integration of JVT-G012 and our proposed algorithm block diagram in H.264/AVC.

## 4.3 Simulation Results

The adaptive context based rate control algorithm with a Laplacian based model and a HOD method was incorporated into the JM H.264/AVC reference software [7]. The evaluation was conducted with the first 100 frames of six QCIF test sequences. For the low bit requirement, 24, 32, and 48 Kbps were selected. The test target bit rate is 24 Kbps for “Grandmother,” “Mother and Daughter,” and “Miss-America,” 32 Kbps for “Akiyo,” “ Bridge-far,” and “Hall,” and 48 Kbps for “Carphone,” and “Highway.” Each sequence is coded at 30 fps with an IPPP structure, which indicates only one I-frame in first frame and P-frame in residual frames, thus the influence of I-frames can be reduced and ignored. There are five reference frames and the search window is set to 15. Context adaptive variable length coding, rate distortion optimization, and rate control are enabled. The residual parameters are fixed equally. As a reference, the performance of the JVT-G102 algorithm and [28] were compared with our proposed algorithm. Although three layer rate control algorithm is presented in [28] only using the block complexity ratio to modify the Lagrange multiplier layer provides the best mode decision and QP selection. Our algorithm also uses block complexity to select a suitable QP, thus, [28] becomes a fair comparison.

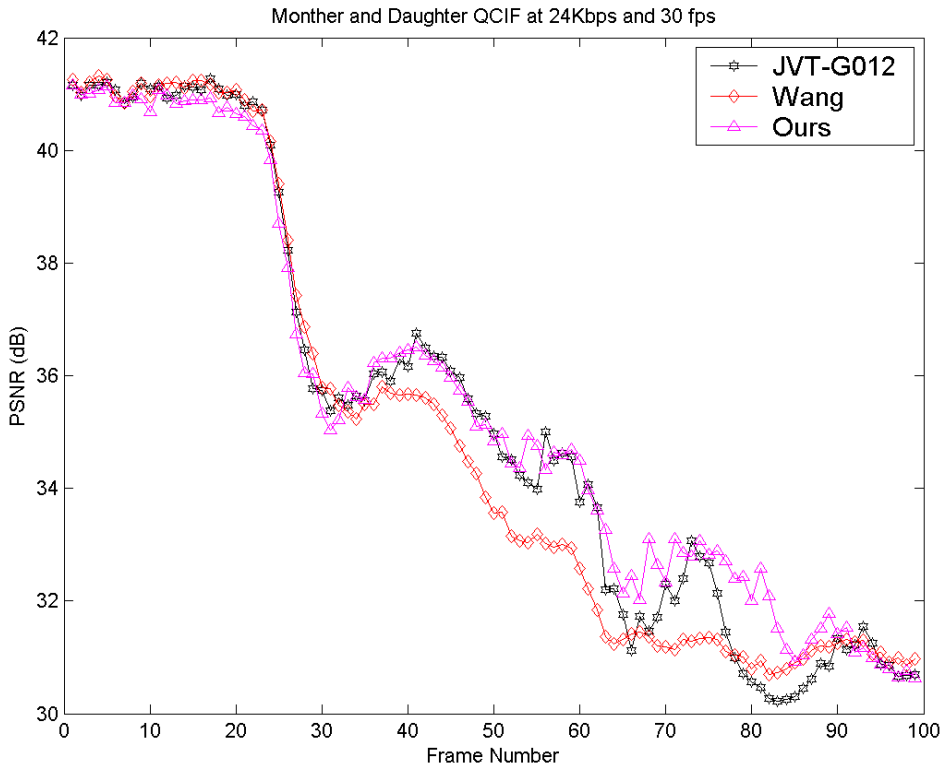
Figure 16 (a)-(c) show a frame by frame comparison in PSNR terms for three standard test sequences: “Mother and Daughter,” “Hall,” and “Highway.” The effectiveness of our algorithm is further indicated by the fact that our algorithm improved the video quality more than JVT-G102 algorithm and [28].

For a more detailed comparison and performance analysis, Figure 17 shows the number of bits in the buffer at each frame.

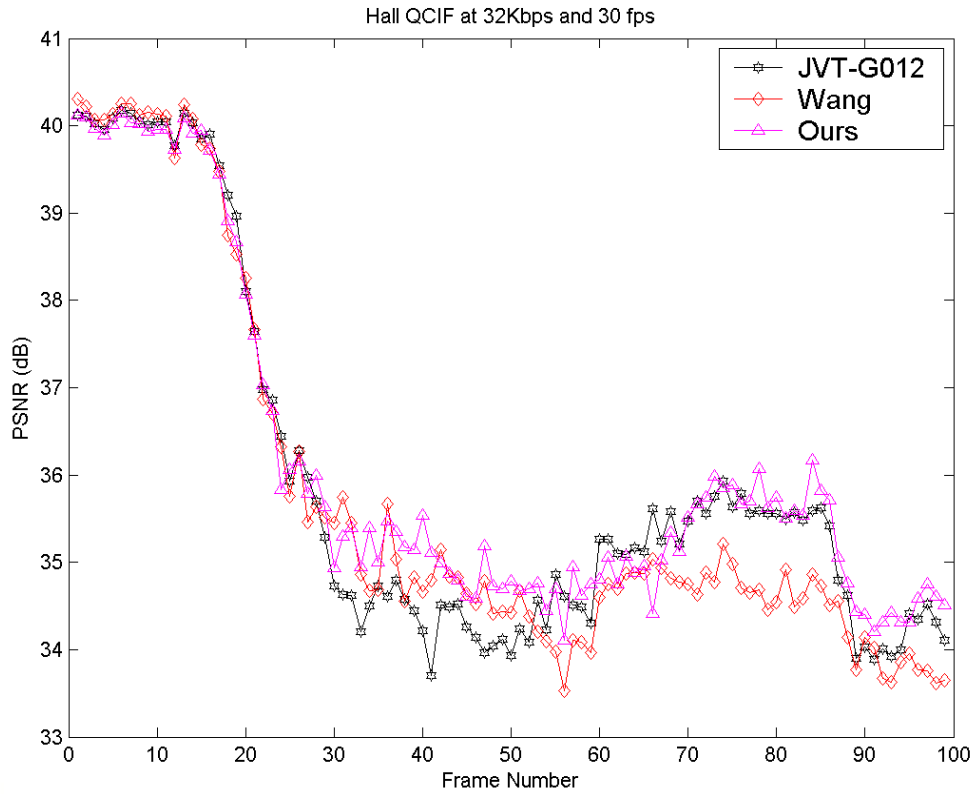
For a fair evaluation of buffer fullness, we first define a reasonable buffer size. According to [27]-[28], a recommended buffer size is set to two-thirds (2/3) of the

channel bandwidth (one second). With bit budget constraints of 24, 32, and 48 kbps and a 30-fps condition, the recommended buffer size is set to (24 k/30 fps)×(2/3) = 533, (32 k/30 fps)×(2/3) = 711, and (48 k/30 fps)×(2/3) = 1066 bits per frame, respectively.

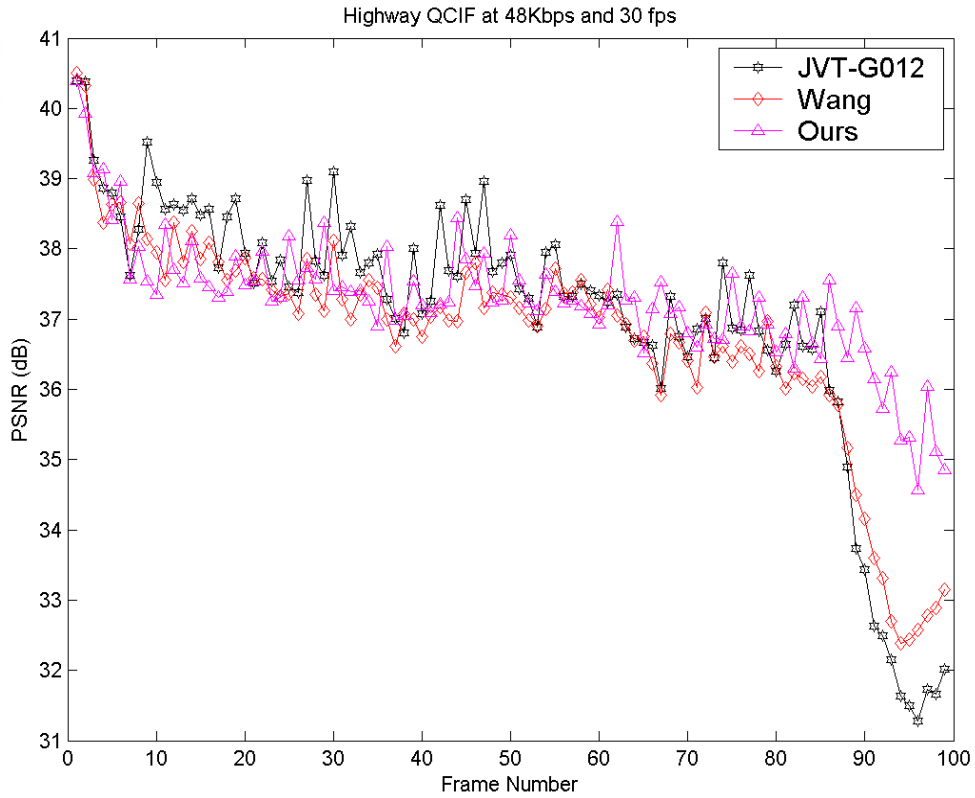
Figure 17(a) and 17(b) display a good result, and the proposed algorithm achieves steadier buffer fullness when compared to JVT-G012 and the Wang algorithms. The encoded bits in all the encoded frames are limited under the recommended buffer size. For real-time and no-delay application, a possible drawback of overflow in Figure 17(c) occurs because approximately 10 encoded frames exceed the limited buffer size. To avoid this problem, a possible solution is to enlarge the buffer size if the bandwidth is available.




(a)

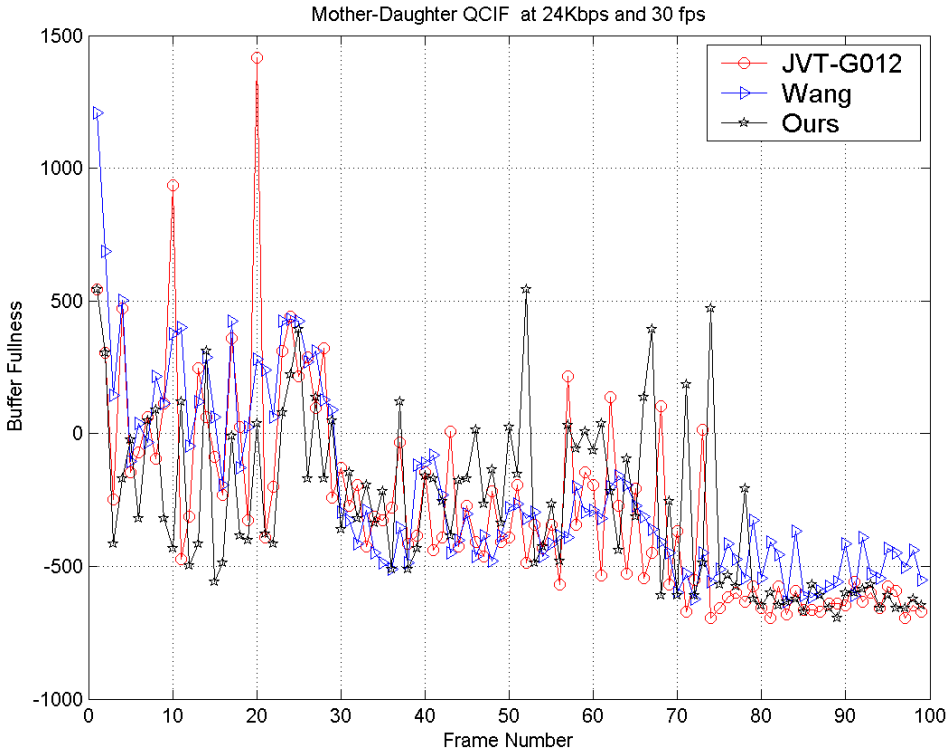



(b)

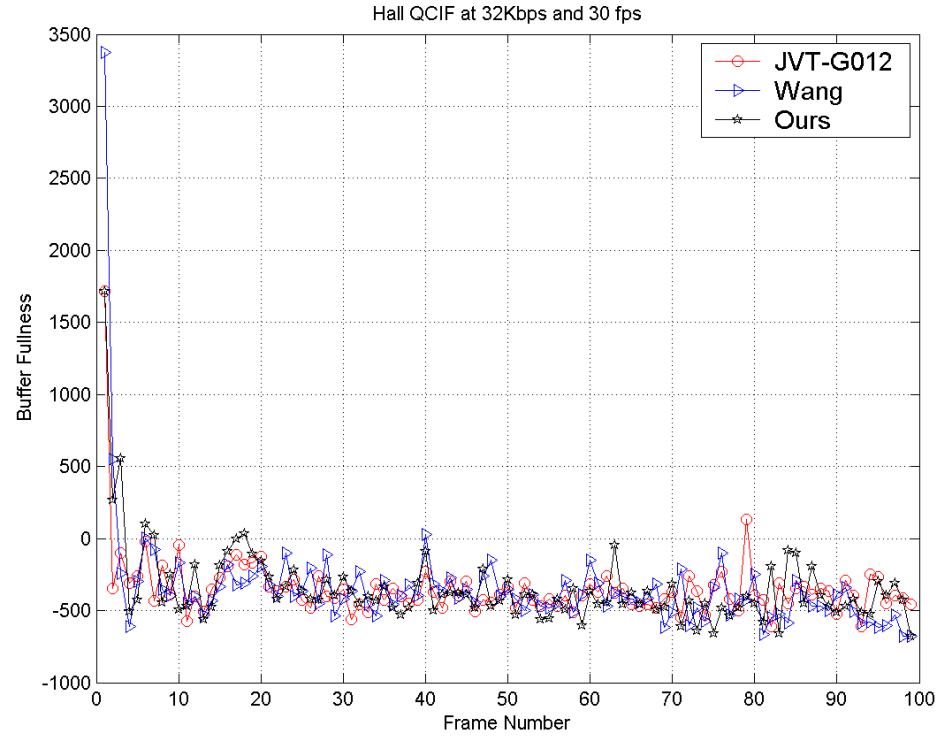



(c)

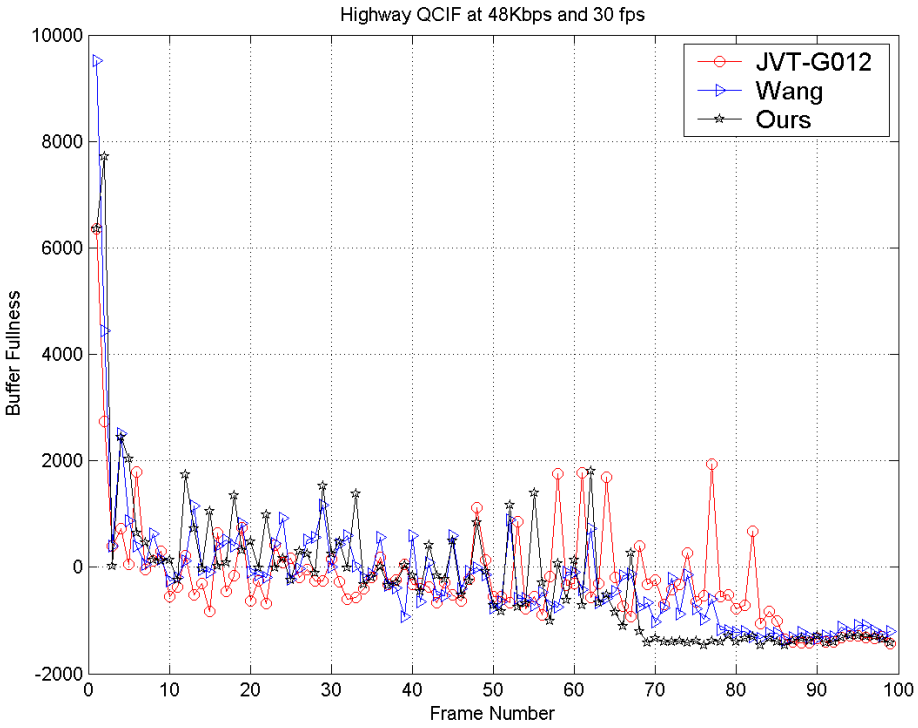
Figure 16. PSNR comparison among JVT-G012, Wang [28], and our proposed algorithm for two standard test sequences: (a) “Mother and Daughter,” (b) “Hall,” and (c) “Highway.”


Mother-Daughter QCIF at 24Kbps and 30 fps
JVT-G012
Wang
Ours
Buffer Fullness
Frame Number
1500
1000
500
0
-500
-1000
0
10
20
30
40
50
60
70
80
90
100

(a)


Hall QCIF at 32Kbps and 30 fps
JVT-G012
Wang
Ours
Buffer Fullness
Frame Number
3500
3000
2500
2000
1500
1000
500
0
-500
-1000
0
10
20
30
40
50
60
70
80
90
100

(b)




(c)

Figure 17. Buffer fullness comparison of JVT-G012, Wang and the proposed algorithm, on three different video sequences: (a) “Mother and Daughter,” (b) “Hall,” and (c) “Highway.”

Table 5. Performance of three algorithms in term of average PSNR, PSNR std. deviation bit rate and ΔR

| Sequence | Method | Average PSNR (dB) | PSNR Std. Deviation | Bit Rates (kbps) | ΔR |
|---|---|---|---|---|---|
| Grandma (24 kbps) | JVT-G012 | 36.17 | 2.91 | 24.21 | 0.88 |
| | Wang [28] | 35.83 | 2.87 | 24.21 | 0.88 |
| | Ours | 36.55 | 2.26 | 24.19 | 0.79 |
| | Gain | +0.38/+0.72 | — | — | — |
| Mother-Daughter (24 kbps) | JVT-G012 | 35.33 | 3.83 | 24.54 | 2.25 |
| | Wang [28] | 35.01 | 3.94 | 25.60 | 6.67 |
| | Ours | 35.48 | 3.61 | 24.54 | 2.25 |
| | Gain | +0.15/+0.47 | — | — | — |
| Miss-America (24 kbps) | JVT-G012 | 39.39 | 2.31 | 24.44 | 1.83 |
| | Wang [28] | 38.94 | 2.69 | 25.25 | 5.21 |
| | Ours | 39.53 | 1.30 | 24.28 | 1.17 |
| | Gain | +0.14/+0.59 | — | — | — |
| Akiyo (32 kbps) | JVT-G012 | 39.70 | 1.35 | 32.12 | 0.38 |
| | Wang [28] | 39.25 | 1.54 | 32.28 | 0.88 |
| | Ours | 39.77 | 1.16 | 32.19 | 0.59 |
| | Gain | +0.07/+0.52 | — | — | — |
| Bridge-far (32 kbps) | JVT-G012 | 39.52 | 0.08 | 32.10 | 0.31 |
| | Wang [28] | 39.67 | 0.10 | 32.22 | 0.69 |
| | Ours | 39.96 | 0.46 | 32.35 | 1.09 |
| | Gain | +0.44/+0.29 | — | — | — |
| Hall (32 kbps) | JVT-G012 | 35.99 | 2.09 | 32.18 | 0.56 |
| | Wang [28] | 35.82 | 2.16 | 32.20 | 0.63 |
| | Ours | 36.16 | 1.93 | 32.17 | 0.53 |
| | Gain | +0.17/+0.34 | — | — | — |
| Carphone (48 kbps) | JVT-G012 | 34.18 | 1.52 | 48.45 | 0.94 |
| | Wang[28] | 33.80 | 2.11 | 49.89 | 3.94 |
| | Ours | 34.21 | 1.26 | 48.25 | 0.52 |
| | Gain | +0.03/+0.41 | — | — | — |
| Highway (48 kbps) | JVT-G012 | 37.08 | 1.95 | 48.70 | 1.46 |
| | Wang [28] | 36.84 | 1.55 | 49.10 | 2.29 |
| | Ours | 37.41 | 1.07 | 48.94 | 1.96 |
| | Gain | +0.33/+0.57 | — | — | — |

Indeed, our algorithm enhanced the PSNR to a greater degree than the two algorithms overall. Comparing Figure 14(c) and Figure 16(c) from the 80$^{th}$ to the 99$^{th}$ frame shows that the degree of PSNR enhancement improved significantly and avoids the local motion problem. Moreover, comparing Figure 14(c) and Figure 16 (c) from the 15$^{th}$ to the 25$^{th}$ frame shows that the enhanced PSNR is maintained, even when the HOD method is applied. The PSNR is close to the two algorithms under smooth conditions and superior when active local motion is produced. Detailed results are given in Table 5. Our algorithm is close to the required bit rate budget, which means that the algorithm can achieve the PSNR enhancement and still match the required bit rate.

To evaluate the bit rate mismatch quantification further, a useful formula [8] is used.

$$\Delta R = \frac{\left| R_{\text{test}} - R_{\text{budget}} \right|}{R_{\text{budget}}} \times 100\% \tag{42}$$

where $R_{test}$ is the bit rate of the test algorithm and $R_{budget}$ is the target bit rate. A small $\Delta$R means that test algorithm can achieve the required bit rate. The gain in Table 5 denotes the comparison of our algorithm to the JVT-G102 algorithm and [28]. In our algorithm, the performance with respect to video quality was enhanced by an average of 0.49 dB and a maximum of 0.72 dB compared to [28] and by an average of 0.21 dB and a maximum of 0.44 dB compared to the JVT-G012 algorithm. Not only did the average PSNR improve dramatically but the standard deviation is lower than that of the existing algorithms. Finally, all PSNR standard deviations are shown and our proposed algorithm maintains a lower deviation than the other two algorithms for test video sequences.

Comparing Tables 4 and 5 shows more excellent results. Although (36) can improve the QP prediction, active local motion will lead to inaccurate QP, which decrease the

PSNR and waste the bit rate budget. To correct the QP prediction, the HODB method is applied to modify QP quickly and accurately. The comparison of calculation units presents the low computation property of the proposed algorithm. Both [28] and our proposed algorithms are built on the JVT-G012, which is compared in terms of addition, subtraction, multiplication, division and square root operations. Thus, the MB level describes and analyzes extra computation in [28] and in our proposed algorithm. In [28], the adaptive Lagrange multiplier $\lambda_i$ in the ith MB is the main contribution and computation. The formula is expressed as follows.

$$\lambda_i = \lambda_i \times \left( \frac{MAD_{frame}}{MAD_{MB}} \right)^{1/4} \tag{43}$$

where $MAD_{frame}$ and $MAD_{MB}$ are the actual MAD in the previous frame and co-located MB respectively. In (43), the numbers of multiplication and square root operations are equal to 1. The addition and subtraction operation for actual MAD calculation are equal to 256 respectively, because the linear regression prediction does not obtain MAD, moreover, the division, operation is 3. The HODB computation is also the main contribution and computation in the proposed algorithm without square root operation. The number of subtraction operations is 256 for two MB operations and the number of addition operations is 256 for histogram accumulation.

Table 6 presents the comparison of five mathematical operations with two different algorithms. The division and square root operation has large computation loading, thus avoiding or reducing the operation requires implementing in hardware or in the firmware level [34]. Table 6 clearly shows that the proposed algorithm is less than [28] in division and square root. Since the digital signal processor (DSP) based system produces multiple multiply-accumulates per cycle, implementing our algorithm in the DSP-based embedded system reduces the addition and multiplication

operations. The histogram function has also been recently built into the DSP core, such as TI TMS320DM6437 [33], due to this very popular function for image and video processing. Thus, HODB computation is not necessary.

Table 6. COMPARISONS OF MAIN COMPUTATION AMOUNTS IN TWO DIFFERENT ALGORITHMS PER MB

| | **[27]** | **Ours** |
|---|---|---|
| # of addition | 256 | 256 |
| # of subtraction | 256 | 256 |
| # of multiplication | 1 | 0 |
| # of division | 3 | 1 |
| # of square root | 1 | 0 |

Figure 18. Visual comparison of the original (top), JVT-G102 (middle), Wang [28] (third), and the proposed algorithm (bottom). These frames are the 81$^{th}$ , 33$^{th}$ 95$^{th}$ and 78$^{th}$ in the test video sequences “Monther and Daughter”, “Hall”,“Highway,” and “Grandma” respectively.

Figure 18 shows actual video sequences to demonstrate that our algorithm suits real

applications. The three algorithms (JVT-G102, [28], and our proposed algorithm) are demonstrated simultaneously when the test video sequences “Mother and Daughter,” “Hall,” “Highway,” and “Grandma” are selected. From visual comparison, our proposed algorithm can select a more precise QP to provide an acceptable visual image quality. It is clear that our proposed algorithm can achieve a higher PSNR than the existing algorithms and reduces the bit rate mismatch ratio.

## 4.4 Summary

This chapter has presented an efficient rate control algorithm for a H.264/AVC video encoder. Through rate control model analysis, an adjustment equation is constructed, based on a statistical model. However, during active local motion the QP will be refined in an incorrect direction. Considering frame complexity and then adaptively enhancing the QP prediction accuracy of the JVT-G102 algorithm, using the HOD method, not only produces a superior result during smooth frames but also generates outstanding results during active local motion. Experimental results show that our algorithm requires no pre-encoding processing and can achieve effective improvements in PSNR—it provided more enhancement than either [28] or the JVT-G102 algorithm. In addition, all algorithms also can be kept on a limited bit rate budget. Thus, our algorithm is suitable for H.264/AVC video rate control even if ported into a real time embedded system.

# Chapter 5 A Temporal-Domain Image Content Inter-Frame Rate Control Algorithm

This chapter proposes a novel rate control for H.264/AVC in low-bit-rate video conferencing. According to the rate-distortion theorem, two concepts are proposed—an approximated rate-distortion equation and a time-domain image content measurement equation—followed by a closed-form equation for quantization step estimation. The main difference between the proposed and Xie's rate control [35] is that the information of time-domain image content is used in this paper, and low-computation complexity is depicted to fit real embedded implementation. Based on this property, the near-stationary image can be reflected directly without frame difference operation to reduce extra encoding bits. To prove that our proposed rate control is suitable for low-bit-rate video conferencing, four typical videos have been selected. Experimental results demonstrate that the proposed rate control scheme can increase the peak signal-to-noise ratio (PSNR) by up to 1.01 dB. These results are based on the JVT H.264/AVC reference model JM15.0. Finally, the proposed rate control provides higher video quality than existing rate control algorithms do.

## 5.1 Review of Related Articles

Xie, in [35], proposed an efficient rate control to improve the current H.264/AVC encoder. An important property of image content is considered to estimate QP. Because the property is valuable for our studies, analyzing more detail in [35] is necessary. The three main contributions are an approximated rate-distortion model, an

estimated texture-complexity (TC) model, and the $Q_{step}$ calculation; these are illustrated as follows.

First, in the approximated rate-quantization model, the original concept is based on H.264/AVC. For mathematical deduction, the distortion model [15] is used and a new approximated rate-quantization model based on low bit rate or high distortion is shown in (44).

$$R(Q_{step}) = \frac{e}{\ln 2} \times \frac{\sigma^2}{Q^2_{step}} \tag{44}$$

where $\sigma^2$ indicates the value of image variance. Second, in the estimated texture-complexity model, Xie defines a new equation to measure texture-complexity as follows.

$$TC = \frac{e}{\ln 2} \times \sigma^2 \tag{45}$$

To circumvent the problem of Lagrange R-D optimization mode selection, the adaptive linear regression TC prediction is also proposed. Finally, $Q_{step}$ is calculated in (46) for current block according to (44) and (45).

$$\begin{cases} TC_{current} = \varphi \times TC_{previous} + \phi \\ Q_{step} = \sqrt{\dfrac{TC_{current}}{R_b}} \end{cases} \tag{46}$$

where $R_b$ is the bits budget of the current basic unit, such as a 4×4 or 8×16 block, and two variables of $\phi$ and $\varphi$ are system parameters.

## 5.2 Our Proposed Algorithm

From the description in chapter 5.1, in the inter-frame-type rate control, the fast and accurate property of QP estimation is extremely difficult because the distortion prediction is not controlled well. Even linear regression can provide a possible

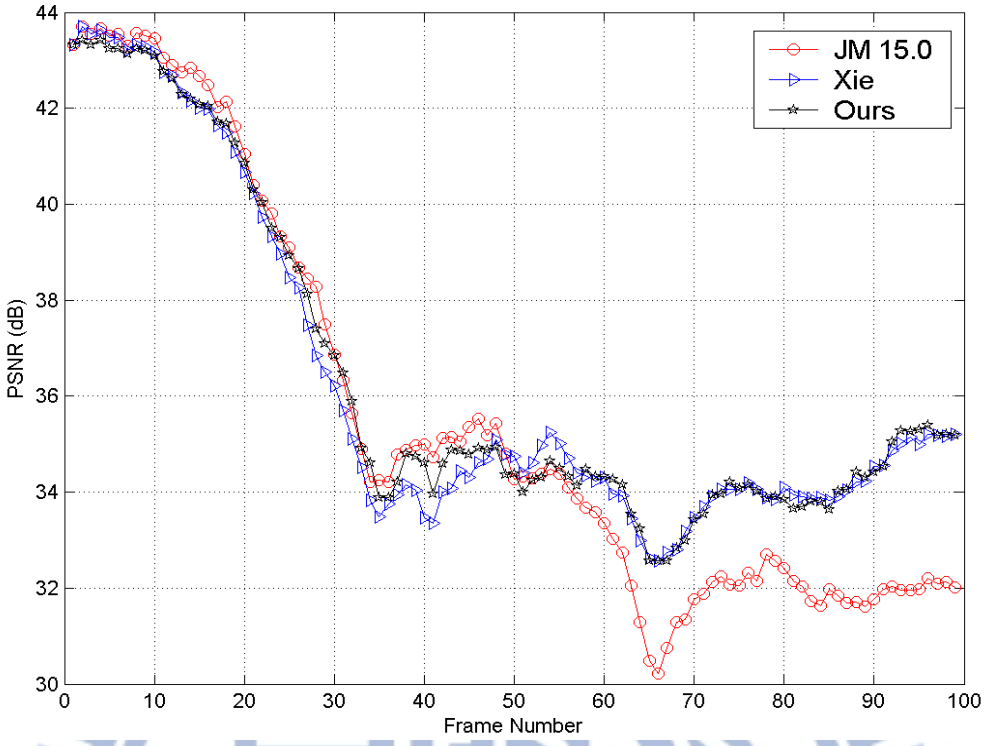
prediction; the vibration and computation complexity are still large. Although Xie can provide a more efficient encoding gain than the current H.264/AVC, a possible drawback of linear regression in $\sigma^2$ obtainment still exists. Even though $\sigma^2$ can be applied after the RDO procedure, TC cannot reflect the image content in the temporal domain. To solve this problem, the estimation of TC based on time-domain image content is applied. This concept is introduced from the rate-distortion theorem. Moreover, emphasizing that the theorem is workable and also adding the temporal-domain information are useful.

Although (19) is a simple and direct formula, the result can directly reflect image distortion. In past approaches, $\sigma^2$ indicated the residue image content between the current and its reconstructed image. However, the residue image is obtained after the quantization procedure, and QP in the quantization procedure is dynamic to fit the rate budget when the rate control is applied. Thus, a typical dilemma of the chicken and egg is performed. To avoid this limitation and to apply an efficient measurement equation, (19) is selected to improve TC estimation. Although (45) can provide a useful TC estimation equation, $\sigma^2$ is not a real calculation under the condition of the chicken and egg. Thus, we can apply (19) for (45) to adjust TC, and two variables of $\sigma^2$ and $\beta^2$ indicate the current and previously reconstructed image variance. Although (19) provides distortion estimation without residue information, the operation of image difference in the temporal domain can yield more and exact information to adjust TC estimation. Moreover, a system parameter $\delta$ is added to enhance the distortion precision. Finally, a new TC, $TC_{adj}$, is constructed, and (46) is also referred to as the QP estimation without linear regression in (47).

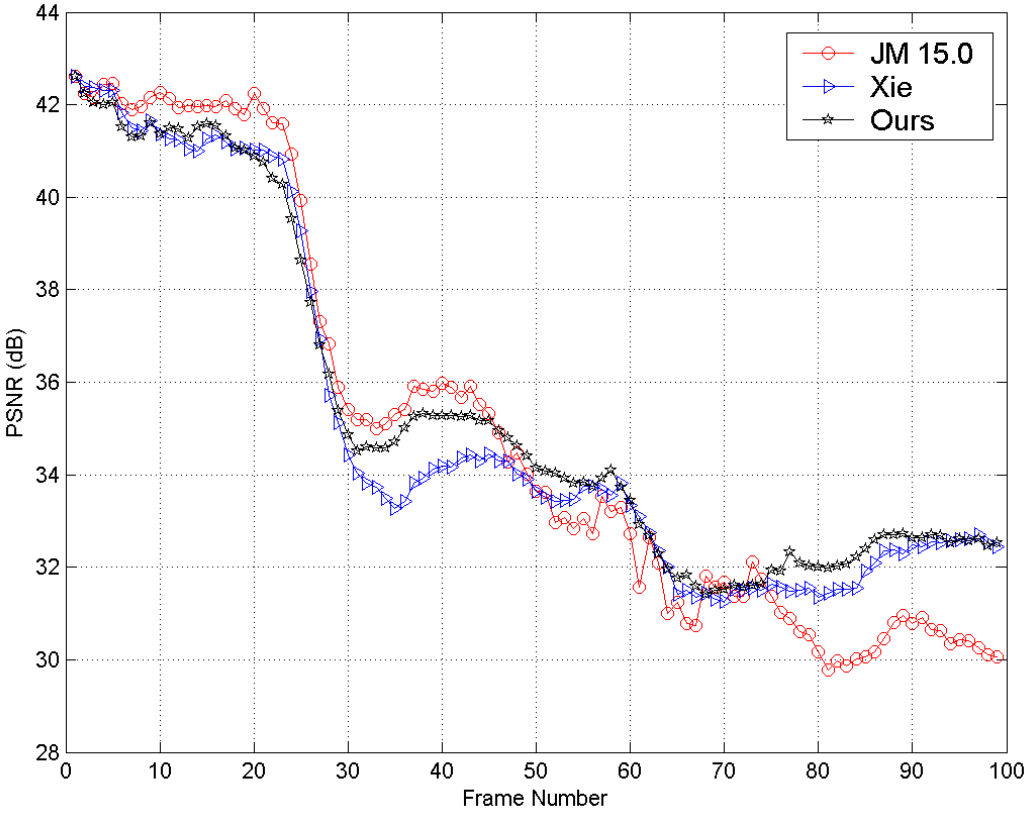
$$\begin{cases} TC_{adj} = \dfrac{e}{\ln 2} \times D \\ D = \delta \times (\sigma^2 - \beta^2) \\ Q_{step} = \sqrt{\dfrac{TC_{adj}}{R_b}} \end{cases} \tag{47}$$

## 5.3 Simulation Results

The H.264/AVC reference software JM15.0 [7] was used to evaluate the proposed rate control algorithm. The evaluation was conducted using the first 100 frames of four QCIF test sequences to construct the video conferencing environment. The test target bit rate is 24 Kbps for “Grandma,” “Mother-Daughter,” “Miss-America,” and “Akiyo” to fit the low bit-rate limitation. Each sequence was coded at 30 fps with the first frame set to I-frame type, and all of the other after-frames set to 99 P-frame types. The reference frame was set to 5, and the search window was set to 15. CAVLC, RDO, and rate control were also enabled. The system parameter $\delta$ was set to 0.9. All other relative parameters were selected equally for [6], [35], and the proposed algorithm. Figure 19(a) and (b) show the frame-by-frame PSNR comparison of three algorithms for the “Grandma” and “Mother-Daughter” video sequences. Figure 20(a) and (b) show the frame-by-frame bit-rate estimation result of different algorithms for the “Grandma” and “Mother-Daughter” video sequences.

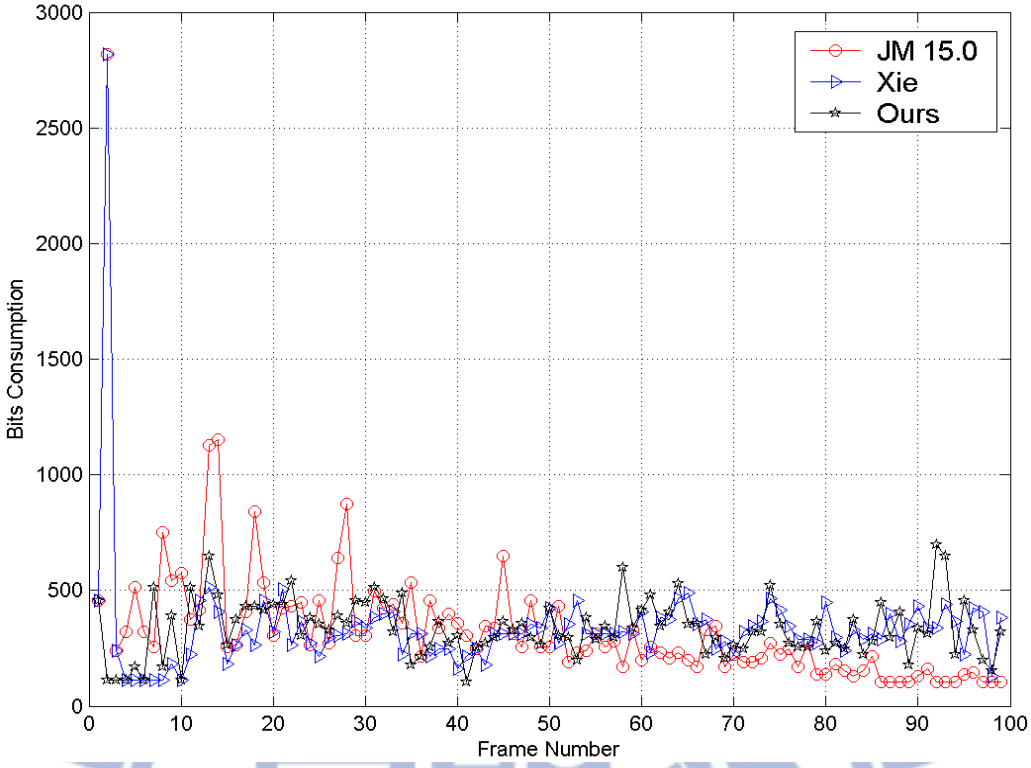



(a)

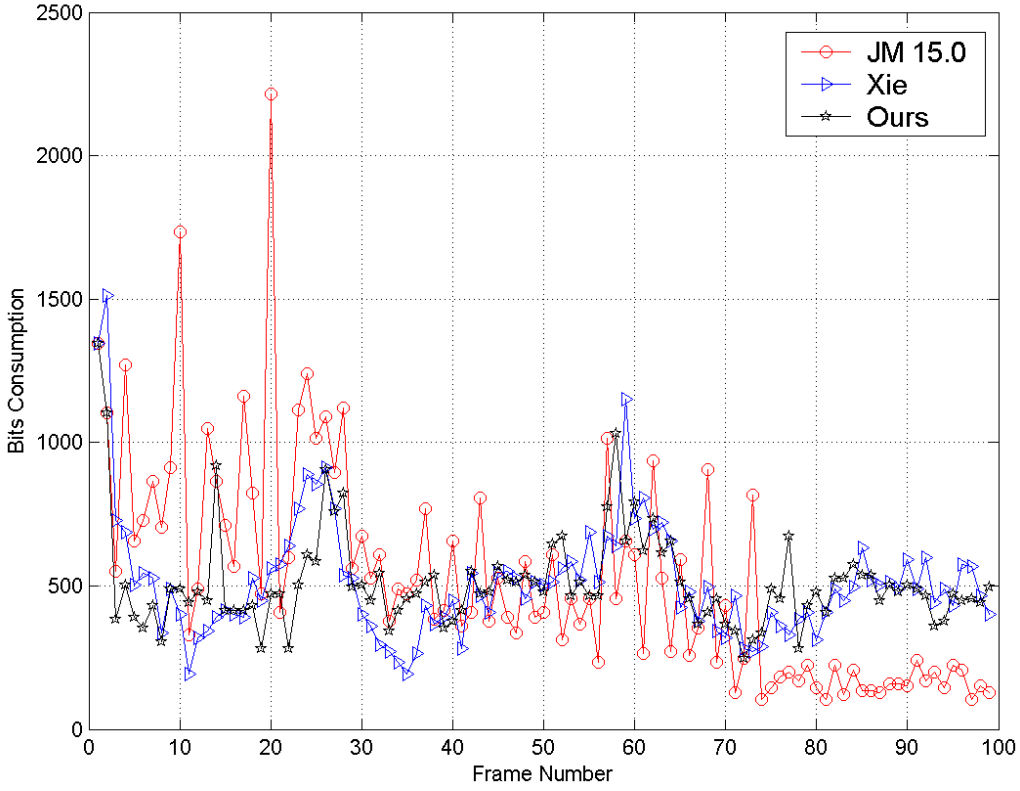



(b)

Figure 19. PSNR comparison among JM15.0, Xie [35] and our proposed method for two standard test sequences. (a) “Grandma” and (b) “Mother-Daughter.”

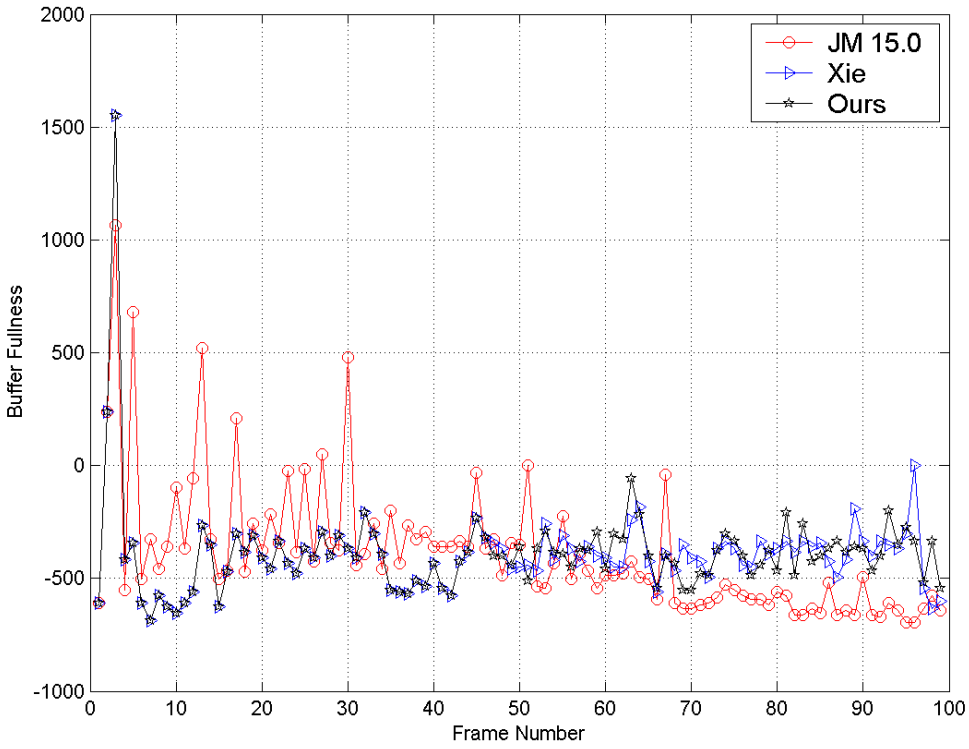



(a)

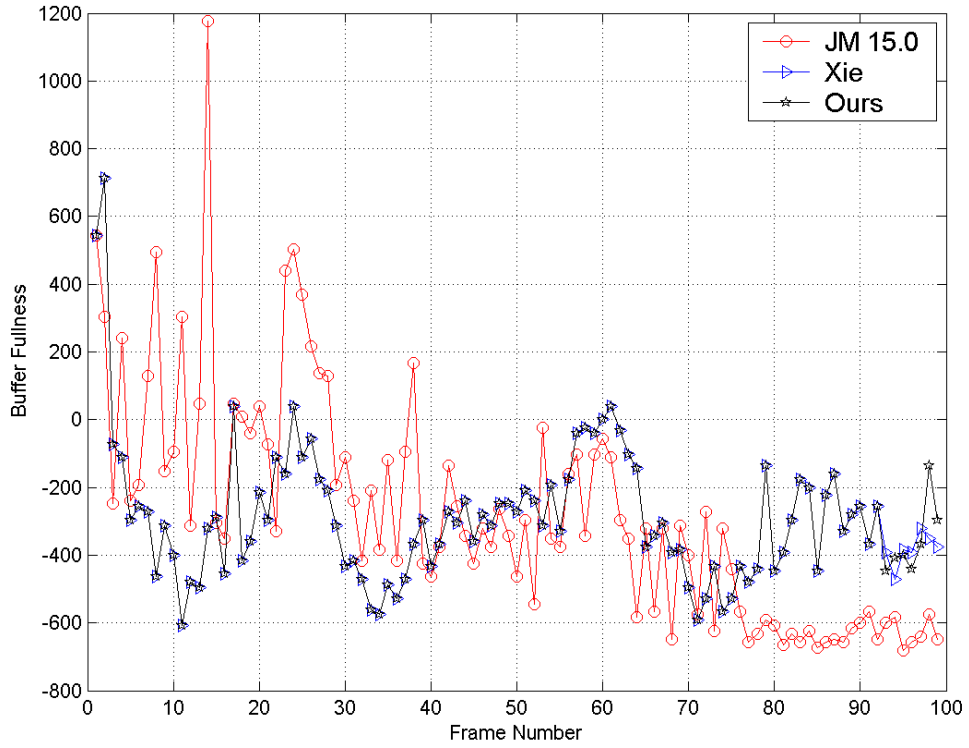



(b)

Figure 20. Bit rate comparison among JM15.0, Xie [35] and our proposed method for two standard test sequences. (a) “Grandma” and (b) “Mother-Daughter.”




(a)




(b)

Figure 21. Buffer fullness comparison of JM15.0, Xie [35] and the proposed algorithm, on two different video sequences: (a) “Grandma” and (b) “Mother-Daughter.”

Two outstanding results are presented. One is from the 40th frame to the 60th frame in Figure 19(a), and the other is from the 30th frame to the 60th frame in Figure 19(b); these show that the smooth property of image quality in our algorithm is more excellent than JVT-G012 and Xie [35]. In detail, even the JM produces higher PSNR from the 40th frame to the 50th frame in Figure 19(a), and from the 30th frame to the 45th frame in Figure 19(b). The JM also produces lower PSNR from the 50th frame to the 60th frame in Figure 19(a), and from the 45th frame to the 60th frame in Figure 19(b). The same status occurs in Xie's algorithm in Figure 19(a) and (b). Another outstanding result from the 60th frame to the 100th frame in Figure 19(a) and from the 75th frame to the 100th frame in Figure 19(b) is equal to Xie's algorithm and JM in PSNR comparison.

Another factor of performance evaluation is to analyze bits consumption. Figure 20(a) and (b) show that the bits vibration in the proposed algorithm is smoother than that of JM and Xie's algorithm. An obvious influence of linear regression is shown in Figure 20(b). A large vibration in JM algorithm is that the linear regression is used twice in QP and distortion estimation, separately. Although Xie's algorithm is only used once in image content estimation for QP estimation, the effect of vibration is also reflected.

For a more detailed comparison and performance analysis, Figure 21 shows the number of bits in the buffer at each frame. In Figure 21(a) and 21(b), almost all encoded frames are underflow because the large number of bits in the initial condition is applied to encoded video sequences. The proposed algorithm is based on the Xie [35] algorithms to modify the distortion prediction only; thus, the underflow status still possibly exists. An excellent result of all three algorithms can meet the real-time requirement of no frame overflow can be obtained.

Table 7. Performance of three algorithms in term of average PSNR, PSNR std. deviation, bit rate and ΔR under initial QP set to 18

| **Initial QP = 18** | | | | |
|---|---|---|---|---|
| **Sequence** | **Method** | **Average PSNR (dB)** | **PSNR Std. Deviation** | **Bit Rate (kbps)** |
| Grandma | JVT-G012 | 35.64 | 14.94 | 24.18 |
| | Xie[35] | 35.90 | 14.90 | 24.08 |
| | Ours | 35.93 | 14.89 | 24.08 |
| | Gain 1 | +0.29 | — | — |
| | Gain 2 | +0.03 | — | — |
| Mother-Daughter | JVT-G012 | 34.04 | 14.81 | 25.08 |
| | Xie[35] | 34.45 | 14.50 | 24.11 |
| | Ours | 34.56 | 14.50 | 24.12 |
| | Gain 1 | +0.52 | — | — |
| | Gain 2 | +0.11 | — | — |
| Miss-America | JVT-G012 | 37.76 | 15.66 | 25.43 |
| | Xie[35] | 37.84 | 15.23 | 24.06 |
| | Ours | 37.90 | 15.23 | 24.10 |
| | Gain 1 | +0.14 | — | — |
| | Gain 2 | +0.06 | — | — |
| Akiyo | JVT-G012 | 37.84 | 15.69 | 24.65 |
| | Xie[35] | 38.04 | 15.47 | 23.30 |
| | Ours | 38.12 | 15.50 | 23.14 |
| | Gain 1 | +0.28 | — | — |
| | Gain 2 | +0.08 | — | — |

Table 8. Performance of three algorithms in term of average PSNR, PSNR std. deviation, bit rate and ΔR under initial QP set to 20

| **Initial QP = 20** | | | | |
|---|---|---|---|---|
| **Sequence** | **Method** | **Average PSNR (dB)** | **PSNR Std. Deviation** | **Bit Rate (kbps)** |
| Grandma | JVT-G012 | 35.51 | 14.84 | 24.16 |
| | Xie[35] | 36.45 | 14.86 | 24.11 |
| | Ours | 36.52 | 14.89 | 24.11 |
| | Gain 1 | +1.01 | — | — |
| | Gain 2 | +0.07 | — | — |
| Mother-Daughter | JVT-G012 | 34.90 | 14.81 | 24.64 |
| | Xie[35] | 35.11 | 14.60 | 24.15 |
| | Ours | 35.17 | 14.61 | 24.13 |
| | Gain 1 | +0.27 | — | — |
| | Gain 2 | +0.06 | — | — |
| Miss-America | JVT-G012 | 38.62 | 15.84 | 24.68 |
| | Xie[35] | 38.87 | 15.58 | 24.09 |
| | Ours | 38.70 | 15.53 | 24.04 |
| | Gain 1 | +0.08 | — | — |
| | Gain 2 | -0.17 | — | — |
| Akiyo | JVT-G012 | 37.92 | 15.65 | 24.25 |
| | Xie[35] | 38.36 | 15.42 | 24.03 |
| | Ours | 38.46 | 15.45 | 23.92 |
| | Gain 1 | +0.54 | — | — |
| | Gain 2 | +0.10 | — | — |

Tables 7 and 8 present more detailed numerical simulation results under a different condition in initial QP selection. The reason is that the initial QP affects the bits budget, because generally large initial QP consumes more bits and reduces bits allocation after encoding a frame. Thus, inter-frame rate control plays an important role. These tables show that the proposed algorithm can provide excellent performance, up to 1.01 dB and 0.11 dB better PSNR than JVT-G012 and [35], respectively, whereas the output bit rate is closer to the target budget. This table also shows two gains, Gain 1 and Gain 2, to depict the performance improvement over JVT-G012 and [35], respectively. Compared with JVT-G012 and [35], the proposed algorithm improves results in most test sequences.

## 5.4 Summary

This chapter presents a novel rate control for H.264/AVC to apply low bit rate video conferencing. This novel rate control algorithm is based on rate-distortion theorem and time-domain image content information. The motivation for this study is to present an efficient distortion measurement without linear regression. Two advantages include lower computation and fast image content reflection. Experimental results show that the proposed algorithm can simultaneously achieve efficient improvement in both the PSNR and the bit rate, even when different initial QP is applied. This algorithm is also constructed on a one-pass scheme; no pre-encoding is required. Numerous simulation results show that the proposed algorithm can also achieve a higher average PSNR and lower bit rate mismatch. Thus, this rate control is better suited to the H.264/AVC rate control than existing algorithms, even in real-time embedded systems.

# Chapter 6 Conclusions and Future work

In this study, we developed a navel distortion prediction method based on the rate distortion theorem for H.264/AVC. According to the results of four applications, we demonstrated that the method was useful, and the results are summarized as follows:

In chapter 2, according to a novel distortion prediction method, we present an efficient intra-frame rate control algorithm to predict accurate quantization parameter by developing variance-based distortion measurements. The advantage is that the accurate distortion is obtained to compare with the mean square error metric.

In chapter 3, we propose a joint rate-distortion optimization for the H.264 rate control algorithm with a novel distortion prediction. The advantage is that the linear regression in distortion prediction may be avoided to compare with the original H.264/AVC.

In chapter 4, we propose a content based intra-frame rate control algorithm for an enhanced JVT-G012 algorithm according to a novel distortion prediction method. The modified histogram was applied to solve the fast object movement problem. The advantage is that the algorithm corresponds to the original JVT-G012 algorithm and the computing complexity is considerably small. Moreover, the computation method is easy for embedded system implementation.

In chapter 5, we propose an efficient image content complexity prediction equation according to a novel distortion prediction method. The advantage is that the accurate distortion prediction can be provided to replace the variance-only method in Xie [35]. Moreover, the image content complexity prediction can be applied directly for the quantization parameter estimation equation in Xie [35].

All of the results in the four applications present excellent PSNR degrees to compare with H.264/AVC. We obtained an efficient distortion prediction method in this study from the original rate distortion theorem. The advantage of the distortion prediction method was verified according to the results from the four applications. Future studies will focus on automatic processing of parameters training and lower computational complexity of quantization parameters.

# Reference


[1] T. Wiegand and G. Sullivan, “The emerging JVT/H.264L video coding standard,” presented at the Tutorial at ICIP 2002, Rochester, NY, Sept. 2002.

[2] H. Kodikara Arachchi and W.A.C. Fernando, “An Intelligent Rate Control Algorithm to Improve the Video Quality at Scene Transitions for off-line MPEG-1/2 Encoders,” *IEEE Trans. Consumer Electron.,* vol. 49, no. 1, pp. 210-219, Feb. 2003.

[3] S. L. P. Yasakethu, W. A. C. Feranado, S. Adedoyin and A. Kondoz, “A Rate Control Technique for Off Line H.264/AVC Video Coding Using Subjective Quality of Video,” *IEEE Trans. Consumer Electron.,* vol. 54, no. 3, pp. 1465-1472, Aug. 2008.

[4] A. Viterbi and J. Omura, Principle of Digital Communication and Coding, New York: McGraw-Hill Electrical Engineering Series, pp.446-448, 1979.

[5] T. Chiang and Y. Q. Zhang, “A new rate control scheme using quadratic rate distortion model,” *IEEE Trans. Circuit Syst. Video Technol.*, vol. 7, no. 1, pp. 246-250, Feb. 1997.

[6] Z. Li, F. Pan, and K. Lim, “Adaptive basic unit layer rate control for JVT,” in proc. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG012, 7th Meeting, Pattaya, Thailand, Mar. 2003.

[7] H.264/AVC Reference Software JM15.0 [Online]. Available: http://iphome.hhi.de/suehring/tml/download/old_jm/

[8] W. Wu and H. K. Kim, “A novel rate control initialization algorithm for H.264,” *IEEE Trans. Consumer Electronics*, vol. 55, no. 2, pp. 665-669, May 2009.

[9] H. Wang and S. Kwong, “Rate-distortion optimization of rate control for H.264 with adaptive initial quantization parameter determination,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 168, no. 1, pp. 140-144, Jan. 2008.

[10] B. Yan and M. Wang, “Adaptive distortion-based intra-rate estimation for H.264/AVC rate control,” *IEEE Signal Process. Lett.*, vol.16, no.3, pp.145-148, 2009.

[11] S. C. Hsia and S. H. Wang, “Adaptive video coding control for real-time H.264/AVC encoder,” *Journal of Visual Communication and Image Representation,* vol. 20, pp. 463-477, 2009.

[12] H. Wang and S. Kwong, “A rate-distortion optimization algorithm for rate control in H.264”, in *Proc. IEEE Conf. on Acoustics, Speech and Signal Processing*, no.1, pp.1149-1152, 2007.

[13] Z. Chen, D. Zhang and K. N. Ngan, “An Efficient Algorithm for H.264/AVC High Definition Video Coding,” *IEEE Trans. Consumer Electron.*, vol. 54, no. 4, pp. 1852-1857, Nov. 2008.

[14] B. Tao, H. A. Peterson, and B. W. Dickinson, “A rate-distortion model for MEPG Encoders,” in *proc. Int. Conf. Image Processing*, Oct. 1997, pp. 338-341.

[15] J. Ribas-Corbera and S. Lei, “Rate control in DCT video coding for low-delay communication,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172-185, Feb. 1999.

[16] J. Dong and N. Ling, “On model parameter estimation for H.264/AVC rate control,” in *proc. IEEE Int. Symp. Circuits Syst.*, New Orleans, LA, May 2007, pp. 289-292.

[17] J. Dong and N. Ling, “A content-adaptive prediction scheme for parameter estimation in H.264/AVC macroblock layer rate control,” *IEEE Trans. Circuit Syst. Video Technol.*, vol. 19, no. 8, pp. 1108-1117, Aug. 2009.

[18] Y. Wu, L. Shouxun, Z. Yongdong, and L. Haiyong, “Optimum bit allocation and rate control for H.264/AVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 705-715, June 2006.

[19] Schäfer and Thomas, Heinrich Hertz Institute, Image Communication JVT Reference Software JM 15.0, 2009.

[20] H. Wang and S. Kwong, “A rate-distortion optimization algorithm for rate control in H.264,” in *proc. Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 1, Apr. 2007, pp.1149-1152.

[21] T. Wiegand, G.J. Sullivan, G. Bjontegaard and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, Aug. 2003.

[22] Y. Zhou, Y. Sun, Z. Feng and S. Sun, “New rate-distortion modeling and efficient rate control for H.264/AVC video coding,” *Signal Processing: Image Communication*, vol. 24, pp. 345-356, 2009.

[23] Y. Liu, Z. G. Li and Y. C. Soh, “A novel rate control scheme for low delay video communication of H.264/AVC standard,” *IEEE Trans. Circuit Syst. Video Technol.*, vol. 17,no. 1, pp. 68-78, Janu. 2007.

[24] C. An and T. Q. Nguyen, “Iterative rate-distortion optimization of H.264 with constant bit rate constraint,” *IEEE Trans. Signal Process.*, vol. 17, pp. 1605-1015, 2008.

[25] H. Song and C. C. Jay Kuo, “Rate control for low-bit-rate video via variable-encoding frame rates,” *IEEE Trans. Circuit Syst. Video T*echnol., vol. 11, no. 4, pp. 512-521, Apr 2001.

[26] H. Song and C. C. Jay Kuo, “A region-based H.263+ codec and its rate control for low bit rate video,” *IEEE Trans Multimedia*., Vol. 6, no 3, pp. 489-500, June 2004.

[27] M. Jaing and N.Ling, “Low-delay rate control for real-time H.264/AVC video coding,” *IEEE Trans Multimedia*., vol. 8, no. 3, pp. 467-477, June 2006.

[28] M. Wang and B.Yan, “On Lagrangian multiplier based joint three-layer rate control for H.264/AVC,” *IEEE Signal Process. Lett.*, vol. 16, no.8, pp. 679-682, Aug. 2009.

[29] D. K. Kwon, M. Y. Chen and C. C. Jay Kuo, “Rate-control for H.264 video with enhanced rate and distortion models,” *IEEE Trans. Circuit Syst. Video T*echnol., vol. 17, no.5, pp. 517-529, May 2007.

[30] S. Zhou, J. T. Li and Y. D. Zhang, “Improvement on rate-distortion performance of H.264 rate control in low bit rate,” *IEEE Trans. Circuit Syst. Video T*echnol., vol. 17, no.8, pp. 996-1006, Aug. 2007.

[31] J. Lee and B. W. Dickinson, “Temporally adaptive motion interpolation exploiting temporal masking in visual perception,” *IEEE Trans. Image Processing*, vol. 3, no.5, pp. 513-526, Sep. 1994.

[32] Wiegand T, Sullivan G J. Draft ITU-T Recommendation H.264 and Final Draft International Standard of Joint Video Specification (ITU-T Recommendation H.264 | ISO/IEC 14496-10 AVC). Joint Video Team of ISO/IEC JTC1/SC29/WG11 and ITU-T SG16/Q.6 Doc. JVT-G050, Pattaya, Thailand, Mar. 2003.

[33] TMS320DM6437 Digital Media Processor (Rev. D), 2008 [Online]. Available: http://www.ti.com

[34] C. H. Kuo, L. C. Chang, K. W. Fan and B. D. Liu, “Hardware/sofrware codesign of a low-cost rate control scheme for H.264/AVC,” *IEEE Trans. Circuit Syst. Video T*echnol., vol. 20, no. 2, pp. 250-261, Feb. 2010.

[35] Z. Xie, Z. Bao, C. Xu and G. Zhang, “Optimal bit allocation and efficient rate control for H.264/AVC based on general rate-distortion model and enhanced

coding complexity measure," *IET Image Process.*, vol.4, no. 3, pp. 172-183, 2010.

[36] International Telecommunication Union, "Video codec for audio-visual services at px64 kbit/s," ITU-T Recommendation H.261 version 2, March 1993.

[37] International Telecommunication Union, "Video coding for low bit rate communication," ITU-T Recommendation H.263 version 2, January 1998.

[38] Consultative Committee for International Telegraph and Telephone, "Description of reference model (RM8)," SG XV WP/1/Q4, June 1989.

[39] International Telecommunication Union, "Video Codec Test Model," ITU-T/SG15, TMN8, Portland, OR, June 1997.

[40] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 8, no. 4, pp. 446-459, August 1998.

[41] E. Viscito and C. A. Gonzales, "Video compression algorithm with adaptive bit allocation and quantization," in Proc. of the SPIE Visual Communications and Image Processing: Visual Communications, Vol. 1605, pp. 58-72, 1991.

[42] ISO/IEC 13818-2:1994, "Information technology-generic coding of moving pictures and associated audio, Part 2: Visual," 1994.

[43] ISO/IEC JTCI/SC29/WG11/No.400/MPEG93/457, "Test Model 5," April 1993.

[44] ISO/IEC 14496-2:1999, "Information technology- coding audio/visual objects, Part 2: Visual," 1999.

[45] ISO/IEC JTCI/SC29/WG11, "MPEG-4 video verification model (VM) v18.0," January 2001.

[46] T. Wiegaad, H. Schwarz, A. Joch, F. Kossentini and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 7, pp. 668-703, July 2003.

[47] T. Wiegaad, G. J. Sullivan and A. Luthra, "Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT G050r1, Geneva, Switzerland, May 2003.

[48] Andrew J. Viterbi and Jim K.Omura, "Principles of Digital Communication and Coding," McGram-Hall, 1979, pp. 447-449.