Rate-Distortion Analysis for H.264/AVC Video Coding and its Application to Rate Control

Siwei Ma, *Student Member, IEEE*, Wen Gao, *Member, IEEE*, and Yan Lu

***Abstract*—In this paper, an efficient rate-control scheme for H.264/AVC video encoding is proposed. The redesign of the quantization scheme in H.264/AVC results in that the relationship between the quantization parameter and the true quantization stepsize is no longer linear. Based on this observation, we propose a new rate-distortion (R-D) model by utilizing the true quantization stepsize and then develop an improved rate-control scheme for the H.264/AVC encoder based on this new R-D model. In general, the current R-D optimization (RDO) mode-selection scheme in H.264/AVC test model is difficult for rate control, because rate control usually requires a predetermined set of motion vectors and coding modes to select the quantization parameter, whereas the RDO does in the different order and requires a predetermined quantization parameter to select motion vectors and coding modes. To tackle this problem, we develop a complexity-adjustable rate-control scheme based on the proposed R-D model. Briefly, the proposed scheme is a one-pass process at frame level and a partial two-pass process at macroblock level. Since the number of macroblocks with the two-pass processing can be controlled by an encoder parameter, the fully one-pass implementation is a subset of the proposed algorithm. An additional topic discussed in this paper is about video buffering. Since a hypothetical reference decoder (HRD) has been defined in H.264/AVC to guarantee that the buffers never overflow or underflow, the more accurate rate-allocation schemes are proposed to satisfy these requirements of HRD.**

***Index Terms*—H.264/AVC, hypothetical reference decoder (HRD), rate control, rate-distortion optimization (RDO), video coding.**

I. Introduction

RATE control plays an important role in all standard-compliant video encoders. Without rate control, the underflow and overflow of the client buffer may occur due to the mismatching between the source bit rate and the available channel bandwidth for delivering a compressed bitstream. In other words, without rate control, any video coding encoder would be practically hard to use. Therefore, the video coding standards usually recommend their own nonnormative rate-control schemes during the standardization process, such as TM5 for MPEG-2 [1], TMN8 for H.263 [2], and VM8 for MPEG-4 [3], which have been developed with regards to these video coding standards.

Nowadays, rate control has become one of the important research topics in the field of video compression and transmission. Besides these standard-recommended rate-control schemes, many improved algorithms have also been developed in the past years. In terms of the transmission style, these rate-control schemes can be classified into two major categories: constant-bit-rate (CBR) control for the constant-channel-bandwidth video transmission [1], [4] and variable-bit-rate (VBR) control for the variable-channel-bandwidth video transmission [4]–[6]. In terms of the unit of rate-control operation, these rate-control schemes can be classified into macroblock- [1], [2], slice-, or frame-layer [3] rate control. These rate-control schemes usually resolve two main problems. The first is how to allocate proper bits to each coding unit according to the buffer status, i.e., rate allocation, and the second is how to adjust the encoder parameters to properly encode each unit with the allocated bits, i.e., quantization parameter adjustment.

The rate allocation in the rate control is usually associated with a buffer model specified in the video coding standard. In the standard, the hypothetical reference decoder (HRD) is usually a normative part to represent a set of normative requirements on bitstreams. It is conceptually connected to the output of an encoder and consists of a decoder buffer, a decoder, and a display unit. A mathematical model, also known as leaky bucket, is usually employed to characterize the hypothetical decoder and its input buffer called coded picture buffers (CPBs). Bits flow into the decoder buffer at a constant rate and are removed from the decoder buffer in chunks. An HRD-compliant bit stream must be decoded in the CPB without overflow and underflow. This requirement can be strictly satisfied by the rate control implemented in the encoder.

The key point of quantization parameter adjustment is to find the relation between the rate and the quantization parameter. Since the source distortion is closely related with the quantization error that is decided by the quantization parameter, the relation between rate and quantization is usually derived based on a rate-distortion (R-D) model [7], [8]. For example, in TM5, a simple linear R-D model, i.e., $R(QP) = X/QP$ [9], is employed, where X is a constant, QP is the quantization parameter and R is the coded bits for the picture. In TMN8 and VM8, the more accurate R-D models, i.e., $R_i = A\left(K\sigma_i^2/QP_i^2 + C\right)$ in TMN8 and $R_i = X_1 \times MAD_i/QP_i + X_2 \times MAD_i^2/QP_i^2$ in VM8, are used respectively. In TMN8, σ_i^2 is the variance of

Manuscript received March 20, 2004; revised November 12, 2004. This work was supported in part by National Fundamental Research and Development Program (973) of China under Contract 2001CCA03300 and by the National Science Foundation of China under Contract 60333020. This paper was recommended by Associate Editor F. Pereira.

S. Ma and W. Gao are with the Digital Media Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China (e-mail: swma@jdl.ac.cn; wgao@jdl.ac.cn).

Y. Lu is with Microsoft Research Asia, Beijing, 100080, China (e-mail: yanlu@microsoft.com).

Digital Object Identifier 10.1109/TCSVT.2005.857300

1051-8215/$20.00 © 2005 IEEE

residual coefficients in a macroblock. A, K, and C are constants. In VM8, MAD_i is the mean absolute difference of a residual macroblock, X_1 and X_2 are model parameters. TMN8 and VM8 can achieve more accurate rate control and provide better performance than TM5 at the price of having relatively higher computational complexity. In [10]–[12], the relation between rate and quantization parameter is indirectly represented with the relation between rate and ρ, where ρ indicates the percent of zero coefficients after quantization. In [13] and [14], a modified linear R-D model with an offset indicating the overhead bits, i.e., $R = (X/QP + L)$, is used for rate control on H.261/H.263.

H.264/AVC is the most up-to-date video coding standard jointly developed by ISO/IEC and ITU-T. The official name is Advanced Video Coding (AVC) of MPEG-4 (part 10 of MPEG-4) in ISO/IEC and H.264 in ITU-T, respectively [15]. The coding efficiency offered by H.264/AVC is much higher than any other existing video coding standards. Similar to the previous standards, H.264/AVC also recommends a rate-control scheme. In the current H.264/AVC test model [16], the recommended rate-control technique was developed partly based on our previous work [17]–[19]. Although it can improve the coding efficiency of fixed quantization and reach the target bit rate accurately, it is still based on the R-D model in VM8. However, the quantization scheme in H.264/AVC has been significantly changed, which leads to the nonlinear relation between the quantization parameter and the true quantization stepsize. Therefore, it is highly desirable to develop a more efficient R-D model for the rate control in H.264/AVC encoder.

In this paper, a more efficient method to realize the rate control on H.264/AVC encoder is proposed. The proposed technique is based on the statistical and theoretical rate-distortion analysis on H.264/AVC. Considering the HRD specification in H.264/AVC, we extend the previous work to strictly satisfy the HRD requirements in the scheme of rate allocation [20]. The target bit for each picture is clipped with an upper and lower bound to guarantee the CPB neither overflow nor underflow. As for the quantization parameter adjustment, it is based on a new R-D model. The R-D relation in H.264/AVC differs from the previous standards because the quantization scheme in H.264/AVC is significantly changed. The quantization parameter and the true quantization stepsize are no longer linear. Therefore, we have to investigate the true relation between rate and quantization parameter. For this purpose, we have extensively investigated the quantization scheme in H.264/AVC and have obtained useful statistics about the relations between distortion and quantization parameter and then derived a new R-D model. Based on the new R-D model, a macroblock-layer rate control is then proposed and implemented in the H.264/AVC encoder.

In addition, the rate-distortion optimization (RDO) [21], [22] scheme also makes the rate control a difficult task, because both RDO-based coding mode selection and rate control depend on the quantization parameter. In other words, rate control usually requires a predetermined set of motion vectors and coding modes to select the quantization parameter, while RDO requires a predetermined set of quantization parameters to select the motion vectors and coding modes. The interference between rate control and RDO has been discussed in [23], whereas it has not provided an efficient way to resolve the interference.

In this paper, we propose a novel way to tackle the interference between rate control and RDO. Briefly, the predicted quantization parameter for the RDO-based mode selection is calculated using the proposed R-D model. After the optimal mode selection the estimated quantization parameter is refined according to the macroblock information, and an estimated R-D cost is then calculated based on the R-D model. If the refined quantization parameter can reach a better R-D cost than the predicted quantization parameter, the refined quantization parameter is used to do RDO again. A threshold can be used to control the number of macroblocks that need to do the second RDO mode selection. Therefore, the proposed technique is a one-pass rate control at frame level or a partial two-pass rate control at macroblock level.

The remainder of this paper is organized as follows. In Section II, the RDO scheme and HRD specification in H.264/AVC are briefly reviewed to reveal their relationship with rate control. In Section III, the R-D relation is statistically and theoretically analyzed, from which a novel R-D model is derived for H.264/AVC. Based on the proposed rate-quantization model, an efficient rate-control algorithm is detailed in Section IV. In Section V, we present some experimental results to evaluate the proposed algorithm. Finally, Section VI concludes the paper.

II. RDO AND HRD IN H.264/AVC

In H.264/AVC encoder, there are two important issues (i.e., RDO and HRD) closely related to rate control. To better understand the proposed technique, we briefly review the RDO scheme and HRD specification in H.264/AVC, respectively.

A. R-D Optimization

The coding mode of a macroblock in H.264/AVC can vary from the set $\{\text{INTRA4} \times 4,\ \text{INTRA16} \times 16,\ \text{INTER16} \times 16, \text{INTER16} \times 8,\ \text{INTER8} \times 16,\ \text{INTER8} \times 8, \text{INTER8} \times 4,\ \text{INTER4} \times 8,\ \text{INTER4} \times 4,\ \text{SKIP},\ \text{DIRECT}\}$. The Lagrangian method is used to find the optimal motion vector for inter coded block and the optimal coding mode for any block. It provides high performance in solving the optimal bit allocation to the motion vectors and the residual coding in the encoder.

For a block in an inter frame, the rate-constrained motion estimation is first done to find the optimal motion vector by minimizing

$$J(mv, \lambda_{\text{MOTION}}) = D(s, c(mv)) + \lambda_{\text{MOTION}} R(mv - pmv) \tag{1}$$

where mv is the motion vector gotten by motion estimation, pmv is the predicted motion vector, and λ_{MOTION} is the Lagrange multiplier. The rate term $R(mv - pmv)$ represents the motion information. $D(s, c(mv))$ is the sum of absolute differences (SAD) or SATD: the sum of absolute differences after Hadamard transform) between the original video signal s and the coded video signal c. For multireference prediction, the reference frame that minimizes the following expression will be selected as the final reference frame:

$$J(\mathrm{REF}|\lambda_{\mathrm{MOTION}}) = D(s, c(\mathrm{REF}, mv(\mathrm{REF}))) + \lambda_{\mathrm{MOTION}}(R(mv(\mathrm{REF}) - pmv(\mathrm{REF})) + R(\mathrm{REF})) \quad (2)$$

where $D(s, c(\mathrm{REF}, mv(\mathrm{REF}))$ is $SA(T)D$ between original signal s and the reference video signal c from reference frame REF.

Afterwards, the rate-constrained mode selection is performed to choose the optimal coding mode by minimizing

$$J(s, c, \mathrm{MODE}|QP, \lambda_{\mathrm{MODE}}) = D(s, c, \mathrm{MODE}|QP), + \lambda_{\mathrm{MODE}} R(s, c, \mathrm{MODE}|QP) \quad (3)$$

where the distortion $D(s, c, \mathrm{MODE}|QP)$ is measured as the sum of squared errors between the original block s and the reconstructed block c, and QP is the quantization parameter. $R(s, c, \mathrm{MODE}|QP)$ is the rate obtained after run-level variable-length coding.

In (1) and (2), the Lagrange multiplier $\lambda_{\mathrm{MOTION}}$ and λ_{MODE} have the following relation with QP:

$$\lambda_{\mathrm{MOTION}} = \sqrt{\lambda_{\mathrm{MODE}}} = m \times 2^{QP/6} \quad (4)$$

where m is a constant. As is well known, $R(s, c, MODE|QP)$ is also associated with the macroblock quantization parameter QP. It means that the macroblock quantization parameter QP would affect the motion vector and mode selection. With the different QP, the different motion vectors and modes might be selected.

B. Hypothetical Reference Decoder

The operation of HRD in H.264/AVC video encoding is briefly reviewed here. More detailed description can be found in [15]. In [15], the basic operation unit in the HRD buffer is an access unit. An access unit may be a coded picture or a slice data partition. For the sake of simplicity, an access unit denotes a coded picture in this paper. As shown in Fig. 1 [27], R is the bit rate at which the CPB is filled, $b(n)$ is the size in bits of picture n, $t_{ai}(n)$ is the time when the first bit of picture n enters the CPB, called the initial arrival time of picture n, $t_{af}(n)$ is the time when the last bit of picture n enters the CPB, called the final arrival time of picture n, and $t_r(n)$ is the time when the picture n is removed from the CPB, called the removal time. According to the specification of H.264/AVC, $t_{ai}(n)$ and $t_{af}(n)$ can be computed by

$$t_{ai}(n) = \begin{cases} 0, & n = 0, \text{ the first picture} \\ t_{af}(n-1), & t_{af}(n-1) \geq t_{ai,\mathrm{earliest}}(n) \\ t_{ai,\mathrm{earliest}}(n), & t_{af}(n-1) < t_{ai,\mathrm{earliest}}(n) \end{cases} \quad (5)$$

where $t_{af}(n) = t_{ai}(n) + b(n)/R$. The removal time $t_r(n)$ of picture n can be computed by

$$t_r(n) = \begin{cases} \frac{\mathrm{initial\_cpb\_removal\_delay}}{90\,000}, & n = 0 \\ t_r(0) + t_c \times \mathrm{cpb\_removal\_delay}(n), & n \neq 0. \end{cases} \quad (6)$$

$t_{ai,\mathrm{earliest}}(n)$ is the earliest time when picture n can reach the buffer or, in other words, it indicates the display time offset from the first picture in the sequence. In particular, $t_{ai,\mathrm{earliest}}(n)$ can be computed as follows:

$$\begin{aligned} t_{ai,\mathrm{earliest}}(n) &= t_c \times \mathrm{cpb\_removal\_delay}(n) - t_0 \\ &= t_r(n) - t_r(0) - t_0 \end{aligned}$$

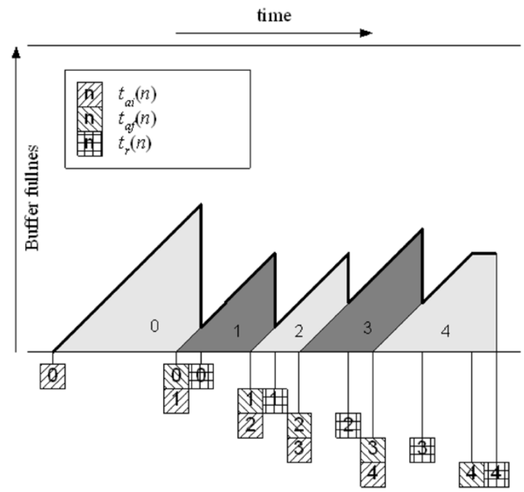

Fig. 1. HRD (coded picture buffer occupancy in the HRD model).

where t_0 is initial_cpb_removal_delay_offset/90 000 [20]. initial_cpb_removal_delay represents the delay between the time when the first bit of the coded data arrives in the CPB and the time when the first time removal of coded data from the CPB. initial_*cpb*_removal_delay_offset denotes the time offset from initial_cpb_removal_delay. The sum of initial_cpb_removal_delay and initial_cpb_removal_delay_offset shall be constant and the bits flowing into the CBP at this time interval shall not exceed the CPB size. cpb_removal_delay specifies how many clock ticks between the recent two times of data removal. t_c is time of a clock tick [20].

To ensure that the CPB does not overflow and underflow, the bits allocated to picture n by rate control must not surpass an upper bound U_n. For CBR, a lower bound L_n must be set to ensure the bit stream enters into the CPB continuously. In other words, the following equation must be true:

$$t_{af}(n) \geq t_{ai,\mathrm{earliest}}(n+1). \quad (7)$$

From (5) and (7), we have

$$b(n) \geq (t_{ai,\mathrm{earliest}}(n+1) - t_{ai}(n))R. \quad (8)$$

Therefore, the lower bound is

$$L_n = \max((t_{ai,\mathrm{earliest}}(n+1) - t_{ai}(n))R, 0). \quad (9)$$

According to the rate-control considerations in [20], it is stated that the CPB should neither overflow nor underflow if the following equation remains true:

$$te[b(n)] \leq t_r(n) - t_{ai}(n) \quad (10)$$

where $be[t]$ and $te[b]$ denote the bit equivalent of a time t and the time equivalent of a number of bits b, respectively, and

$$te[b] = \frac{b}{R}, \quad be[t] = tR. \quad (11)$$

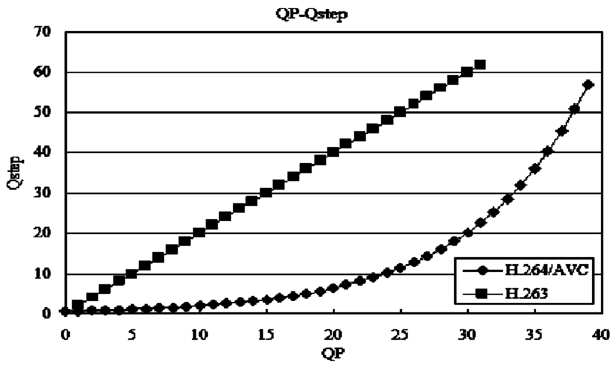

Fig. 2. Relation between QP and Q_{step} in H.264/AVC.

From (6) and (7), we have

$$b(n) = be[te[b(n)]] \leq be[t_r(n) - t_{ai}(n)]. \tag{12}$$

Therefore, the upper bound U_n for the picture n is

$$U_n = (t_r(n) - t_{ai}(n))R. \tag{13}$$

This means that the picture n must be in the buffer when it is removed from the buffer. Since sometimes the rate control is not very accurate, further limits need to be done on U_n or L_n through the bit allocation of rate control. Usually the bits allocated to a picture are clipped to $[\alpha L_n, \beta U_n]$, where $\alpha \geq 1.0$ and $\beta \leq 0.9$.

III. Proposed Rate-Qstep Model

Before we present the proposed rate-control technique, we first make a study of the quantization scheme and the R-D relation in H.264/AVC. Before starting the discussion, we would like to clarify the difference between quantization parameter and quantization stepsize. Quantization parameter denotes the quantization scale indirectly, whereas the quantization stepsize is the true value used in quantization. In the previous video coding standards, the relation between quantization parameter and quantization stepsize is usually linear. For example, in H.263 quantization scheme, in terms of the quantization parameter QP ranging from 1 to 31, a coefficient COF is quantized to

$$\text{LEVEL} = \begin{cases} \frac{|\text{COF}|}{(2\times QP)} \\ \frac{|\text{COF}-\frac{QP}{2}|}{(2\times QP)}. \end{cases} \tag{14}$$

In this case, the quantization stepsize $Q_{\text{step}} = 2QP$.

However, in H.264/AVC, the relation between QP and Q_{step} is that $Q_{\text{step}} = 2^{(QP-4)/6}$, as shown in Fig. 2. This change comes from the integer transform and division-free quantization scheme used in H.264/AVC, i.e., the integer transform [24], [25] is

$$Y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} [X] \begin{bmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{bmatrix}. \tag{15}$$

Since the integer transform is not a unitary matrix, Y must be normalized with

$$W = [Y] \otimes [E] = [Y] \begin{bmatrix} a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \\ a^2 & ab & a^2 & ab \\ ab & b^2 & ab & b^2 \end{bmatrix} \tag{16}$$

where $a = 1/2$ and $b = 1/\sqrt{10}$. In H.264/AVC, the transform normalization is in conjunction with quantization for the purpose of being division-free, and therefore the division operation is replaced with multiply and right-shift operations. Concretely, with regard to quantization parameter QP, the quantization is implemented as

$$\begin{aligned} \text{LEVEL}_{i,j} &= \text{round}\left(\frac{W_{i,j}}{Q_{\text{step}}}\right) \\ &= \text{round}\left(Y_{i,j} \times \frac{E_{i,j}}{Q_{\text{step}}}\right) \\ &= \text{round}\left(Y_{i,j} \times \frac{S_{i,j}}{2^{\text{qbits}}}\right) \end{aligned} \tag{17}$$

where $qbits = 15 + \text{floor}(QP/6)$. $S_{i,j}$ is defined by matrix A as

$$A = \begin{bmatrix} 13\,107 & 5243 & 8066 \\ 11\,916 & 4660 & 7490 \\ 10\,082 & 4194 & 6554 \\ 9362 & 3647 & 5825 \\ 8192 & 3355 & 5243 \\ 7282 & 2893 & 4559 \end{bmatrix}$$

$$\begin{aligned} S_{i,j} &= A(QP\%6), r \\ r &= 0, \quad (i,j) = \{(0,0), (0,2), (2,0), (2,2)\} \\ r &= 1, \quad (i,j) = \{(1,1), (1,3), (3,1), (3,3)\} \\ r &= 2, \quad \text{for others.} \end{aligned}$$

In the remaining part of this section, we present the R-D analysis as well as the relation between rate and quantization parameter in H.264/AVC. It has been proved in [26] that the relation between the peak SNR PSNR and the quantization parameter QP is linear as follows:

$$\text{PSNR} = 10 \log\left(\frac{255^2}{\text{MSE}}\right) = l \times QP + b \tag{18}$$

where l and b are the constants. An example of this linear relation is shown in Fig. 3. Supposing the MSE is used as the distortion measure for a frame distortion D, we can derive the relation of the distortion D and the quantization parameter QP as

$$D = \text{MSE} = \frac{(255^2)}{10^{((lQP+b)/10)}}. \tag{19}$$

Statistics have shown that, when QP increases by 1, PSNR usually loses 0.6–0.75 dB. As for the relation between R and QP in H.264/AVC, we also know that QP increases by 1 with a corresponding 12.5% bit-rate reduction. To further reveal the relation of $R - QP$ and $R - Q_{\text{step}}$, we have done experiments on a large number of sequences. Figs. 4 and 5 show the relation between R and $1/QP$ and the relation between R and $1/Q_{\text{step}}$ on the News sequence, respectively, wherein R is the number

TABLE I
EXPERIMENTAL RESULTS ON NEWS

QP	18	20	22	24	26	28	30	32	36	40
R	105.7	83.88	64.65	48.11	36.07	26.53	19.17	13.69	6.944	3.324
$\rho_{R\text{-}1/QP}$	0.991									
$\rho_{R\text{-}1/Qstep}$	0.999									
$\sum(R\text{-}R')^2_{1/QP}$	236.003									
$\sum(R\text{-}R')^2_{1/Qstep}$	15.1208									

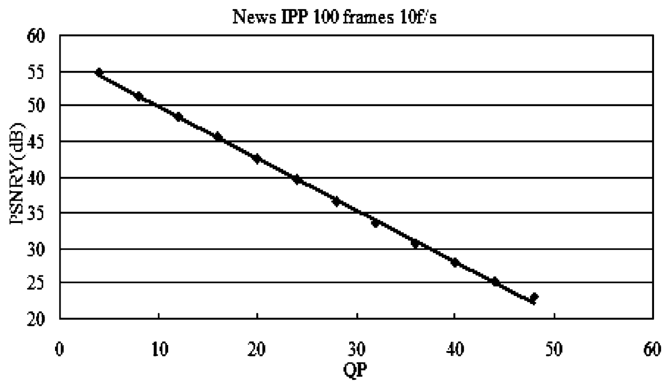

Fig. 3. PSNR of News IPP 100 frames 10 f/s QP range from 4 to 48 with a stepsize of 4.

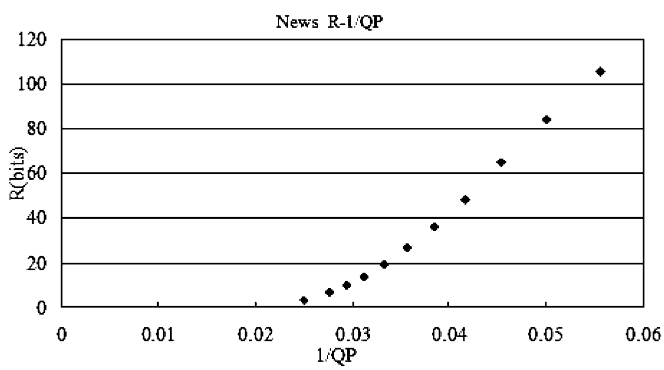

Fig. 4. Relation between coded coefficient bits R and $1/QP$ for News.

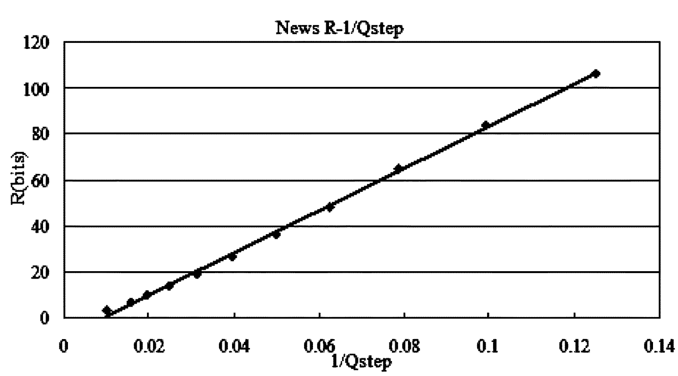

Fig. 5. Relation between coded coefficient bits R and $1/Q_{\rm step}$ for News.

of bits for luminance and chrominance coefficients. Table I provides the statistical results. According to Table I, the correlation coefficient $\rho_{R-1/QP}$ for R and $1/QP$ is 0.991, whereas the correlation coefficient $\rho_{R-1/Q_{\rm step}}$ for R and $1/Q_{\rm step}$ is 0.999. The sum of squared errors for the linear approximation

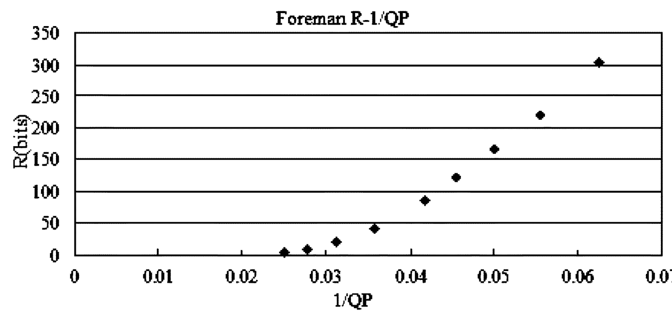

Fig. 6. Relation between coded coefficient bits R and $1/QP$ for Foreman.

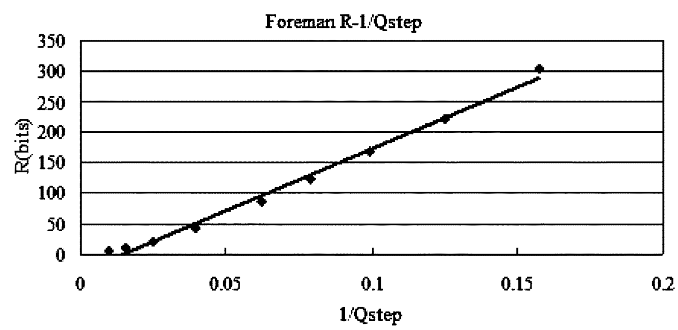

Fig. 7. Relation between coded coefficient bits R and $1/Q_{\rm step}$ for Foreman.

of $R-(1/QP)$ is also much larger than that of $R-(1/Q_{\rm step})$. Conclusively, the relation between R and $1/Q_{\rm step}$ can be more accurately approximated by a linear model rather than the relation between R and $1/QP$. More experimental results on the other sequences also show the similar results, e.g., the statistical results on Foreman sequence shown in Figs. 6, 7, and Table II, respectively.

As a consequence, we can draw a conclusion that the relation between R and $1/Q_{\rm step}$ can be taken as linear in H.264/AVC. The relationship between rate and quantization stepsize (herein referred to as the R-Qstep model) is then denoted as

$$R_i^t = \frac{K^t \mathrm{SAD}_i}{Q_{\mathrm{step},i}} + C^t \; t = I, P, B \tag{20}$$

where R_i^t is the estimated number of coded bits of a macroblock and SAD_i is the SAD of a motion-compensated macroblock. The first item reflects the bits used to code the transform coefficients. The second item is the bits used to code the header information of a macroblock. Compared to the previous linear models in terms of the quantization parameter [1], [13], [14], the proposed linear model in terms of quantization stepsize is more accurate for H.264/AVC. For intra frame, a fixed quantization parameter QP computed as the average quantization parameter

TABLE II
EXPERIMENTAL RESULTS ON NEWS

QP	16	18	20	22	24	28	32	36	40
R	302.61	220.7	167.49	123.46	86.26	42.85	20.43	10.16	4.91
$\rho_{R\text{-}1/QP}$	0.983								
$\rho_{R\text{-}1/Qstep}$	0.996								
$\sum(R\text{-}R')^2{}_{1/QP}$	3071.4								
$\sum(R\text{-}R')^2{}_{1/Qstep}$	680.5643								

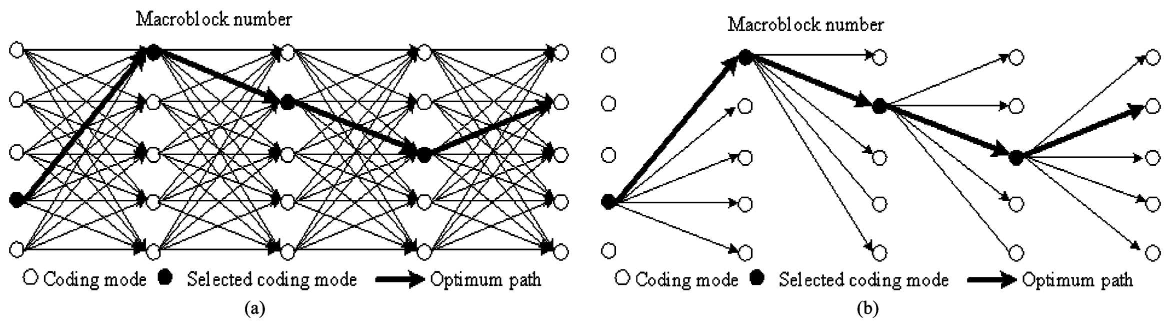

Fig. 8. Joint mode selection and rate control for MPEG-2 encoder. (a) Optimal path found by full search. (b) Near optimal path found by greedy search.

of coded P frames in previous GOP (group of pictures) is used to code the whole frame.

IV. Proposed Rate Control

In [28], the coding efficiency of MPEG-2 encoder is improved by jointly optimizing coding mode selection and rate control. The solution to this optimization problem is viewed as finding an optimum path in a trellis, as shown in Fig. 8(a). In the trellis, each node in a stage denotes a coding mode for the current macroblock, and each link entering the node is assigned a cost to denote the cost used to code the macroblock with the mode. In [28], the cost is the bits used to code the macroblock under a given distortion. The optimum path can be found by full search in the trellis, but it is too complex for the practical applications. Therefore, in [28], a greedy approach is proposed to get the near optimal solution, shown as Fig. 8(b). In the greedy search process, an estimated distortion is given for a macroblock according to the macroblock spatial-masking-activity and the mode with the least coded bits is selected as the best mode. After one pass coding of all the macroblocks in the frame, the distortion should be adjusted to reach the target bitrate and another one or more pass coding may be needed to reach the target bitrate. In [28], the rate control is reached by one or more pass coding for a frame to reach the target bitrate. It is not very suitable for real coding due to its high complexity.

While having the same merit of jointly optimizing coding mode selection and rate control, the proposed algorithm that resolves coding mode selection and rate control at the macroblock level can further improve the coding efficiency. In the previous section, a linear R-Qstep model has been proposed to represent the relation of rate and quantization stepsize in H.264/AVC. In this section, an efficient rate control scheme for H.264/AVC is presented. Both the RDO and HRD are considered in the implementation.

TABLE III
NOTATIONS USED IN PROPOSED RATE-CONTROL ALGORITHM

Variable	*Definition*
t	Picture type, $t = P, B$
n^t	The counter for the coded t type picture in the current GOP
j^t	The counter used for updating rate distortion model parameter
i	The counter for the coded macrobock number in the current picture
N^t	The number of t type pictures in the current GOP
K^t	Rate distortion model parameter for the t type picture
C^t	Rate distortion model parameter for the t type picture
KK_n	The midterm value of model parameter, used to update K^t
CC_n	The midterm value of model parameter, used to C^t
$SAD_n{}^t$	Average SAD of the coded n th t type picture in the current GOP
SAD_i	SAD of the macrobock i
T	Target bits allocated to the current picture
T_i	Target bits for the remained not coded macroblocks in the current picture
$R_n{}^t$	Coded bits of the n th t type current picture in the current GOP
R_i	Coded bits for the macroblock i in the current picture
QP	Quantization parameter for the current macroblock
Q_{step}	Quantization step for the current macroblock
QP_{prev}	Average quantization parameter of the previous coded picture
D_i	Distortion measure for the macroblock i
$D_i{}'$	Estimated distortion measure for the macroblock i
$R_i{}'$	Target bits for the macroblock i
$Cost$	Rate distortion cost for the current macroblock
$Cost'$	Estimated rate distortion cost computed by rate distortion model
M_i	Remained not coded macroblocks in the current picture
$X^t{}_n$	Complexity measure for the n th t type picture in the current GOP

Before describing our proposed rate-control algorithm, we would like to give a summary of the frequently used notations in rate control. These notations are listed in the Table III. The proposed rate-control algorithm is performed in the following steps.

Step 1) **Bit allocation.**

In this step, a target bit is allocated to each picture in a group of pictures (GOP). For the first $I/P/B$ picture, bit allocation is not performed, and a fixed

TABLE IV
EXPERIMENTAL RESULTS ON TEST SEQUENCES ($\epsilon = 0$)

Sequence	Sequence Type	Target Bit-rate (kbps)	Rate Control Scheme	Coded Bit-rate (kbps)	PSNRY (dB)	Improvement over (dB)		
						AVC-TM	TM5	Fixed-QP
News	QCIF 10f/s, IPP	10.19	Proposed	10.19	28.78	0.48	0.98	0.78
			AVC-TM	10.23	28.30			
			TM5	10.20	27.80			
			Fixed-QP	10.19	28.00			
Foreman	QCIF 15f/s, IBBP	96.63	Proposed	97.76	36.35	0.06	0.50	0.25
			AVC-TM	97.76	36.29			
			TM5	98.00	35.85			
			Fixed-QP	96.63	36.10			
Tempete	CIF 30f/s, IPP	1073.72	Proposed	1074.68	34.94	0.03	0.34	0.30
			AVC-TM	1074.80	34.91			
			TM5	1073.91	34.60			
			Fixed-QP	1073.72	34.64			
Mobile	CIF 30f/s, IBBP	1325.66	Proposed	1325.54	34.23	0.08	0.48	0.46
			AVC-TM	1325.61	34.15			
			TM5	1326.36	33.75			
			Fixed-QP	1325.66	33.77			
Football	SD 30f/s, IBBP	3140.92	Proposed	3148.18	28.22	0.16	0.24	0.17
			AVC-TM	3156.19	28.06			
			TM5	3160.23	27.98			
			Fixed-QP	3140.92	28.05			
FlowerGarden	SD 30f/s, IBBP	3515.03	Proposed	3528.95	32.41	0.23	0.38	0.21
			AVC-TM	3530.46	32.18			
			TM5	3536.07	32.03			
			Fixed-QP	3515.03	32.20			
Crew	HD 30f/s, IBBP	2940.47	Proposed	2933.48	38.82	0.12	0.32	0.16
			AVC-TM	2944.85	38.70			
			TM5	2948.54	38.50			
			Fixed-QP	2940.47	38.66			

QP_0^t is then used. After coding the picture, parameters X_0^t, K^t and C^t are initialized, respectively, as follows:

$$\begin{aligned} X_0^t &= R_0^t Q_{\text{step},0}^t = R_0^t \times 2^{\left(\frac{(QP_0^t-4)}{6}\right)} \\ C^t &= B_{\text{head}} \\ K^t &= B_{\text{coeff}} \frac{Q_{\text{step},0}^t}{\text{SAD}_0^t} \\ &= B_{\text{coeff}} \times \frac{2^{\left(\frac{(QP_0^t-4)}{6}\right)}}{\text{SAD}_0^t} \end{aligned} \tag{21}$$

where X_0^t is global complexity measure for a picture as defined in TM5, R_0^t is the coded bits of the picture, SAD_0^t is the average SAD of all macroblocks in the picture, B_{head} is the average header bits for a macroblock, including motion and mode information, and B_{coeff} is the bits used to code luminance and chrominance coefficients.

For the other frames, let T_n^t denote the target bits for the nth t-type picture in a GOP. T_n^t is computed using the method in TM5. In this paper, T_n^t should be clipped to achieve the number of bits allocated to this picture, i.e.,

$$T = \begin{cases} \beta U_n, & T_n^t > \beta U_n, \beta \le 1 \\ L_n, & T_n^t < L_n \\ T_n^t, & \text{other.} \end{cases} \tag{22}$$

In our experiments, $\beta = 0.9$. Set $j^t = 0$ for the first t-type picture, which is used to update K^t in the R-D model. Set $K_1 = KK_0 = K^t, C_1 = CC_0 = C^t$.

Step 2) **Initialization for the current macroblock.**

In this step, we initialize some parameters. Let $i = 1$ for the first macroblock. M_i is the number of remaining uncoded macroblocks in the current frame. Assume that T_i is the number of available bits for encoding the remaining uncoded macroblocks in this frame and $T_1 = T$.

Step 3) **First RDO-based coding mode selection.**

In this step, a predicted quantization parameter QP is used to perform RDO-based coding mode selection. If the current macroblock is the first one in the frame, QP is set to be the average quantization parameter QP_{prev} of the previous frame; otherwise, if $T_i/M_i < C^t$, $QP = QP_{\text{prev}}$, else we calculate Q_{step} based on the proposed R-Qstep model for the current macroblock as follows:

$$Q_{\text{step}} = \frac{K^t \text{SAD}_{i-1}}{\left(\frac{T_i}{M_i - C^t}\right)}. \quad (23)$$

Therefore, $QP = \text{round}(6\log_2 Q_{\text{step}} + 4)$. Further, the calculated QP is clipped to be

$$QP = \min(\max(QP_{\text{prev}} - 3, QP), QP_{\text{prev}} + 3). \quad (24)$$

The predicted quantization parameter QP is then used to perform mode selection for the current macroblok. Afterwards, an R-D cost is computed in terms of the current macroblock as

$$\text{Cost} = D_i + \lambda_{\text{MODE}} R_i = \text{MSE}_i + \lambda_{\text{MODE}} R_i. \quad (25)$$

R_i and D_i denote the number of coded bits and the associated distortion in terms of the selected mode for the current macroblock, wherein the distortion D_i is indicated with the MSE, MSE_i, calculated with

$$D_i = \text{MSE}_i = \frac{255^2}{10^{((lQP+b)/10)}}. \quad (26)$$

Step 4) **Second RDO-based coding mode selection.**

Calculate SAD'_i in terms of the selected coding mode and then compute a new quantization parameter QP' using SAD'_i in the same way as Step 3). Thus, an estimated Cost' is computed for the new QP' as follows:

$$\text{Cost}^t = D'_i + \lambda'_{\text{MODE}} R'_i \quad (27)$$

where $\lambda'_{\text{MODE}} = m^2 \times 2^{QP/6}$ and $D'_i = D_i/10^{l(QP'-QP)/10}$. For simplicity, set $l = -0.5$. If $T_i > 0$, the target bits are used to calculate the R-D cost and $R'_i = T_i/N_i$; otherwise the coded bits of first RDO-based coding mode selection are used to estimate the R-D cost and $R'_i = R_i$. If $\text{Cost}' < \epsilon\text{Cost}$, the new quantization parameter QP' is selected to perform RDO-based coding mode selection again.

Step 5) **Counter updating.**

In this step, the remaining bits and the number of not coded macroblocks of the frame are updated as follows:

$$T_{i+1} = T_i - R_i \text{ and } M_{i+1} = M_i - 1.$$

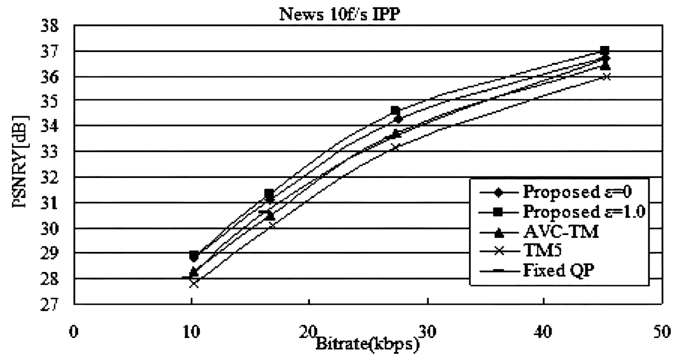

Fig. 9. PSNR curve of news sequence with QCIF format at 10 fps.

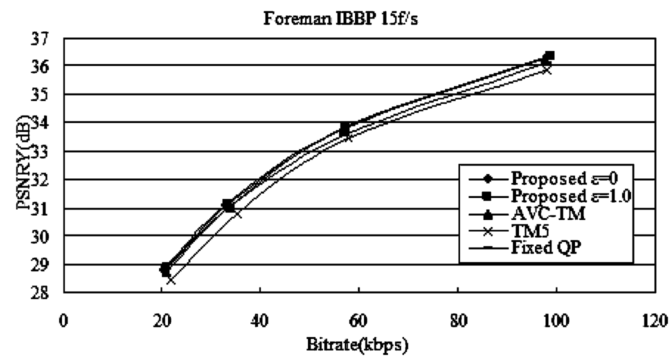

Fig. 10. PSNR curve of Foreman sequence with QCIF format at 15 fps.

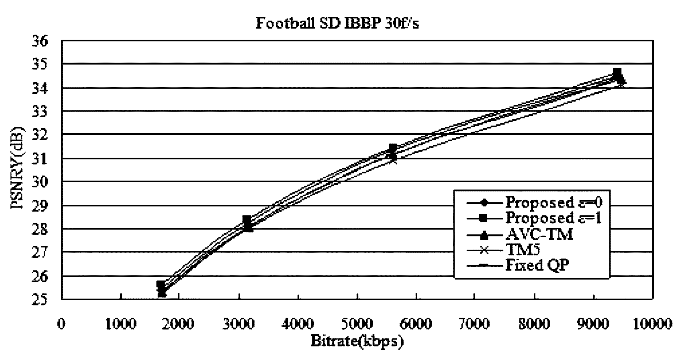

Fig. 11. PSNR curve of football sequence with SD format at 30 fps.

Assume that MB_CNT is the total number of macroblocks in the frame. If $i = MB\_CNT$, all macroblocks in the frame are coded; otherwise, let $i = i + 1$, and go to Step 2).

Step 6) **R-D model parameter updating.**

The R-D mode parameters K and C updating process is similar to [2]. However, in this paper, it is performed at frame level rather than macroblock level for the sake of simplicity.

First, calculate

$$K^t = \frac{R_C Q_{\text{step}}}{\text{SAD}_n^t \times MB\_CNT}$$

and

$$C^t = \frac{R_n^t - R_C}{MB\_CNT} \quad (28)$$

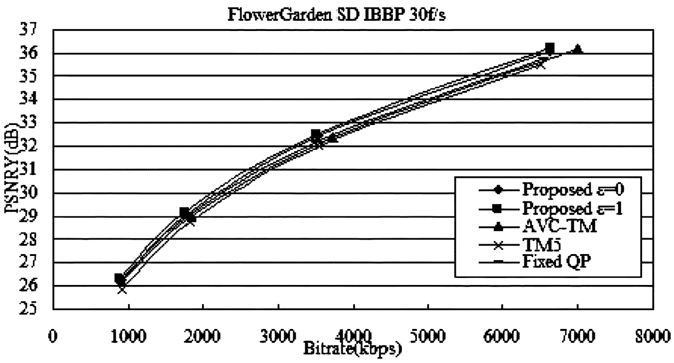

Fig. 12. PSNR curve of flowergarden sequence with SD format at 30 fps.

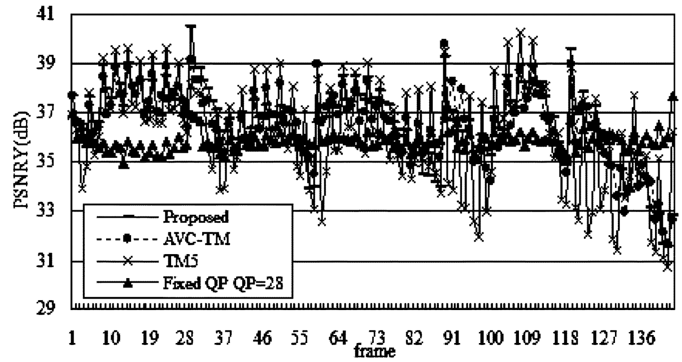

Fig. 14. PSNR curve of Foreman sequence with QCIF format at 15 fps.

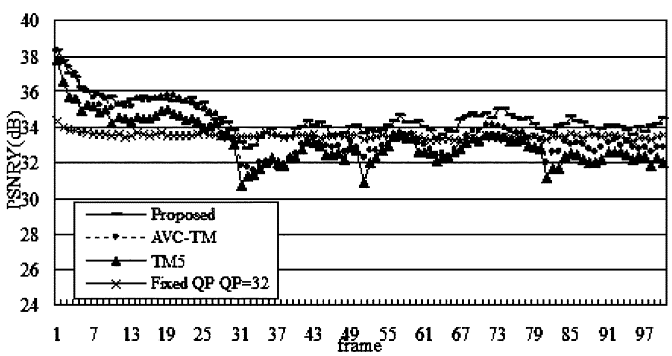

Fig. 13. PSNR curve of News sequence with QCIF format at 10 fps.

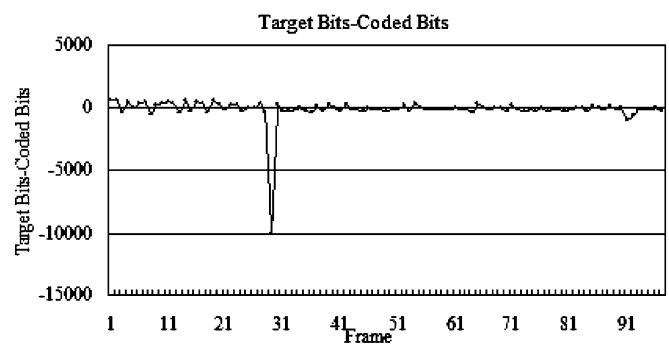

Fig. 15. Bit error between target bits and coded bits for scene cut (at frame 30) test.

where R_n^t is the coded bits of the just coded picture and R_C is the number of bits spent for the luminance and chrominance. If $K^t \geq 0$ and $K^t \leq \pi \log_2 e$, set $j^t = j^t + 1$ and

$$KK_{j^t} = KK_{j^t-1}\frac{(j^t-1)}{j^t} + \frac{K^t}{j^t}. \tag{29}$$

CC_{n^t} is updated with

$$CC_{n^t} = CC_{n^t-1}\frac{(n^t-1)}{n^t} + \frac{C^t}{n^t}. \tag{30}$$

Then, K^t and C^t are updated as a weighted average of the initial estimates and the current average, i.e.,

$$\begin{aligned} K^t &= KK_{j^t}\left(\frac{n^t}{N^t}\right) + K_t\frac{(N^t-n^t)}{N^t}, \\ C^t &= CC_{n^t}\left(\frac{n^t}{N^t}\right) + C_1\frac{(N^t-n^t)}{N^t}. \end{aligned} \tag{31}$$

From the above steps, we can summarize that the proposed algorithm performs one-pass operation at frame level and a partial two-pass operation at the macroblock level. Since the second RDO-based coding model selection in Step 4) is a conditional operation, the percentage of such macroblocks can be controlled by parameter ϵ; consequently, the computing complexity is controllable as well. In other words, the second RDO-based coding model selection can be discarded in the rate-control implementation if setting $\epsilon = 0$. In this case, the coding efficiency may be slightly decreased while the accuracy of rate control is still doing well.

V. Experimental Results

In order to evaluate the performance of the proposed algorithm, we have performed some experiments on some test sequences. The H.264/AVC reference software[1] has been employed as the test platform. Several schemes including the proposed rate-control scheme, TM5 with some parameter adjustment, the current rate control scheme in H.264/AVC test model (AVC-TM) [16], and fixed quantization parameter (Fixed-QP) coding, i.e., without rate control, are tested and compared. The complexity of the proposed scheme can be controlled by parameter ϵ, e.g., $\epsilon = 0$ means that any macroblock is coded with only one RDO-based coding mode selection, and hence the proposed rate control becomes fully a one-pass algorithm. The proposed algorithm is tested with the different values of ϵ.

Table IV illustrates the coding results on some test sequences. The target bit rate is selected according to Fixed-QP coding. In these experiments, set $\epsilon = 0$ in the proposed rate-control implementation. The table shows that the proposed algorithm can accurately control the bit rate at different resolution and frame rate and, at the same time, achieve better coding efficiency than any other coding scheme. Compared to AVC-TM, the proposed algorithm can achieve a maximum gain of 0.48 dB and an average gain of 0.2 dB on the News sequence. The complexity of the proposed rate control is also much lower than that of AVC-TM in that it employs a simple linear prediction rather than the complicated MAD prediction as used in the latter. Compared to the

[1][Online]. Available: http://www.jdl.ac.cn/user/swma/jm61a_rc.zip

Fig. 16. Visual quality comparison between fixed QP coding (left) and proposed rate-control scheme (right) at scene cut position.

optimized TM5 implementation and Fixed-QP coding, the proposed algorithm can improve the coding efficiency with the average PSNR gains of 0.46 and 0.33 dB, respectively. Figs. 9–12 show the R-D curves achieved from the proposed rate control, TM5, AVC-TM, and Fixed-QP coding, respectively. These figures further prove that the proposed algorithm can achieve better performance than any other coding scheme. The experimental results of the proposed algorithm with $\epsilon = 1$ are also shown. Notice that $\epsilon > 1$ is unnecessary in the proposed algorithm. In the experiments, about 25% of the macroblocks have performed the second RDO-based coding mode selection with $\epsilon = 1$. In this case, the average gain is 0.17 dB over the proposed algorithm with $\epsilon = 0$, i.e., full one-pass operation. The improvement is due to the fact that the partial two-pass rate control at the macroblock level can further refine the bit allocation.

The proposed R-D model developed according to the quantization scheme in H.264/AVC leads to these improvements of the proposed algorithm over the other schemes. For the News sequence with small motion, the proposed algorithm can reach much better results. The performance is almost the optimum in terms of bit allocation. Since the motion vectors also cost the coded bits, the motion in the sequence may influence the rate-control scheme as well. However, for the Foreman sequence with high motion, the performance of the proposed scheme is still better than that of the other schemes, which demonstrates that the proposed R-D model is sufficiently accurate even for the sequence with high motion. We also make an implementation of TM5 on JM6.1a software with some parameter adjustment, which still leads to the performance loss and large errors in rate control, because it does not define a dynamic rate of the quantization parameter of each macroblock. Figs. 13 and 14 show the PSNR per frame for some test sequences. These figures further indicate that the proposed rate control shows better performance and has improved the coding efficiency of the original AVC encoder. The proposed rate-control scheme also performs well for

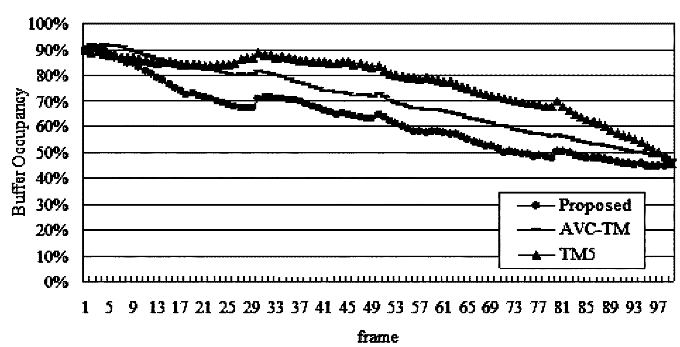

Fig. 17. Buffer occupancy of News sequence at 27 kbps, QCIF, IPP, and 10 fps.

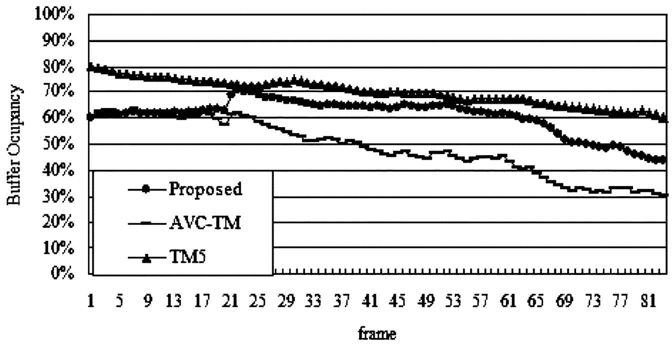

Fig. 18. Buffer occupancy of Foreman sequence at 45 kbps, QCIF, IPP, and 15 fps.

the video sequences with scene cuts. A merged sequence by News and Foreman is used for test and the scene cut happens at the 30th frame (count from 0).

Fig. 15 shows the mismatch between the numbers of target bits and coded bits. At the scene cut, the mismatch is somehow large, whereas the mismatches at the subsequent frames are reduced immediately. Fig. 16 shows the visual quality comparison at the scene cut. After the scene cut, proposed rate control can also reach approximated visual quality compared with fixed QP coding.

Figs. 17 and 18 show the buffer occupancy in terms of the proposed rate-control scheme. According to these plots, the proposed rate control can maintain suitable buffer occupancy levels. In other words, the proposed rate-control algorithm can prevent the buffer from overflow or underflow.

VI. Conclusion

This paper has presented an efficient rate-control algorithm for H.264/AVC video encoder. The proposed algorithm is the extension of our previous work adopted in the H.264/AVC test model. In this paper, a more accurate linear R-D model is developed based on the statistical and theoretical analysis on the quantization scheme as well as the other coding tools in H.264/AVC. The proposed rate-control algorithm is implemented by considering both the generated bit rate and the optimal coding mode selection. In other words, the R-D optimization is performed in conjunction with the proposed rate-control algorithm. In addition, the requirements of HRD on rate control have been satisfied with some clipping operations in the proposed algorithm. Experimental results have shown that the proposed algorithm can generate the bit stream very close to the target bit rate, and meanwhile its coding efficiency is better than the current rate control scheme in H.264/AVC test model as well as the coding with the fixed quantization parameter. Experimental results have also shown that the coding efficiency can be further improved with a little computing complexity increase when using partial two-pass processing.

References

[1] Test Model 5. [Online]. Available: http://www.mpeg.org/MPEG/MSSG/tm5

[2] J. Corbera and S. Lei, “Rate Control for Low-Delay Video Communications,” ITU Study Group 16, Video Coding Experts Group, Portland, Documents Q15-A-20, 1997.

[3] T. Chiang and Y. Zhang, “A new rate control scheme using quadratic rate distortion model,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 287–311, Apr. 1997.

[4] N. Mohsenian, R. Rajagopalan, and C. A. Gonzales, “Single-pass constant- and variable-bit-rate MPEG-2 video compression,” *IBM J. Res. Develop.*, vol. 43, no. 4, pp. 489–509, Jul. 1999.

[5] P. H. Westerink, R. Rajagopalan, and C. A. Gonzales, “Two-pass MPEG-2 variable-bit-rate encoding,” *IBM J. Res. Develop.*, vol. 43, no. 4, pp. 471–488, Jul. 1999.

[6] M. Hamdi, J. W. Roberts, and P. Rolin, “Rate control for VBR video coders in broad-band networks,” *IEEE J. Sel. Areas Commun.*, vol. 15, no. 6, pp. 1040–1051, Aug. 1997.

[7] R. M. Gray and D. L. Neuhoff, “Quantization,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[8] J. Chen and H. Hang, “Source model for transform video coder and its application-part I, II,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 2, pp. 287–311, Apr. 1997.

[9] J. Katto and M. Ohta, “Mathematical analysis of MPEG compression capability and its application to rate control,” in *Proc. IEEE ICIP*, vol. II, 1995, pp. 555–559.

[10] S. Milani, L. Celetto, and G. A. Mian. A rate control algorithm for the H.264 encoder. presented at *Proc. 6th Baiona Workshop on Signal Processing for Communications*. [Online]. Available: http://baiona-workshop.org/Proceedings/milani/pdf

[11] Z. He and S. K. Mitra, “A unified rate-distortion analysis frame work for transform coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 12, pp. 1221–1236, Dec. 2001.

[12] Z. He, “ρ-domain rate-distortion analysis and rate control for visual coding and communication,” Ph.D. dissertation, Elect. Comput. Eng., Univ. California, Santa Barbara.

[13] J. Wang, Z. Chen, Y. He, and Y. Chen, “A MAD-Based Rate Control Strategy,” Document JVT-D070, Klagenfurt, Austria, 2002.

[14] C. Wong, O. C. Au, B. Meng, and H. Lam, “Novel H.26X optimal rate control for low-delay communications,” in *Proc. ICICS-PCM*, Singapore, Dec. 2003, pp. 90–94.

[15] T. Wiegand, G. Sullivan, and A. Luthra, “Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264ISO/IEC 14 496-10 AVC),” Document JVT-G050r1, Geneva, 2003.

[16] *Draft ISO/IEC 14 496-5:2002/PDAM6*, ISO IEC TC JTC 1/SC29 N5821, 2003.

[17] S. Ma, W. Gao, F. Wu, and Y. Lu, “Rate control for AVC video coding scheme with HRD considerations,” in *Proc. IEEE Int. Conf. Image Processing*, Barcelona, Spain, Sep. 2003, pp. 793–796.

[18] S. Ma, W. Gao, P. Gao, and Y. Lu, “Rate control on advanced video coding (AVC) standard,” in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. I, Bangkok, Thailand, May 2003, pp. 892–895.

[19] S. Ma, F. Wu, and Z. Li, “Draft for Adaptive Rate Control With HRD Considerations,” Document JVT-H017, Geneva, Switzerland, 2003.

[20] T. Wiegand, “Editor’s Proposed Draft Text Modifications for Joint Video Specification (ITU-T Rec.H.264 ISO/IEC 14 496-10 AVC), Geneva Modifications, Draft 7,” Document JVT-E146 Annex C, 2002.

[21] G. J. Sullian and T. Wiegand, “Rate-Distortion optimization for video compression,” *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[22] T. Wiegand and B. Girod, “Lagrange multiplier selection in hybrid video coder control,” in *Proc. Int. Conf. Image Processing*, Thessaloniki, Greece, Oct. 2001, pp. 542–545.

[23] M. Gallant, G. Cote, and S. Wenger, “Test Model Issue—Performance of Rate-Distortion Optimization for Fixed Bit Rate Encoding,” Document Q15-F-35, Korea, 1998.

[24] H.264/MPEG-4 Part 10: Transform & Quantization. [Online]. Available: http://www.vcodex.com

[25] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, “Low-Complexity transform and quantization in H.264/AVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 598–603, Jul. 2003.

[26] K. Takagi, Y. Takishima, and Y. Nakajima, “A study on rate distortion optimization scheme for JVT coder,” in *Proc. SPIE*, vol. 5150, 2003, pp. 914–923.

[27] E. Viscito, “HRD and related issues,” in *4th Meeting Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)* , Klagenfurt, Austria, Jul. 22–26, 2002, Paper no. JV-D131.

[28] H. Sun, W. Kwok, M. Chien, and C. H. J. Ju, “MPEG coding performance improvement by jointly optimizing coding mode decisions and rate control,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 3, pp. 449–458, Jun. 1997.

Siwei Ma (S'03) received the B.S. degree from Shandong Normal University, Jinan, China, in 1999. He is currently working toward the Ph.D. degree at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing.

His research interests include image and video coding, video streaming, and transmission.

Wen Gao (M'99) received the M.S. degree and the Ph.D. degree in computer science from Harbin Institute of Technology, Harbin, China, in 1985 and 1988, respectively, and the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He was a Research Fellow with the Institute of Medical Electronics Engineering, University of Tokyo, in 1992, and a Visiting Professor with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, in 1993. From 1994 to 1995, he was a Visiting Professor with the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge. Currently, he is the Director of the Joint R&D Lab (JDL) for Advanced Computing and Communication, Chinese Academy of Sciences, a Professor with the Institute of Computing Technology, a Professor of computer science with the Harbin Institute of Technology, and an honor Professor of computer science at City University of Hong Kong. He has published seven books and over 200 scientific papers. His research interests are in the areas of signal processing, image and video communication, computer vision, and artificial intelligence.

Dr. Gao chairs the Audio Video coding Standard (AVS) workgroup of China. He is the head of the Chinese National Delegation to MPEG working group (ISO/SC29/WG11). He is also the Editor-in-Chief of the *Chinese Journal of Computers* and the general Co-Chair of the IEEE International Conference on Multimodel Interface in 2002.

Yan Lu received the B.S., M.S., and Ph.D. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 1997, 1999, and 2003, respectively.

From 1999 to 2000, he was a Research Assistant with the Computer Science Department, City University of Hong Kong. From 2001 to 2003, he was with the Joint R&D Lab (JDL) for Advanced Computing and Communication, Chinese Academy of Sciences, Beijing, China. Since April 2004, he has been with Microsoft Research Asia, Beijing, as a Postdoctoral Researcher. His research interests include video compression, video segmentation, and pattern recognition.