Rate-Distortion Optimization of Rate Control for H.264 With Adaptive Initial Quantization Parameter Determination

Hanli Wang, Member, IEEE, and Sam Kwong, Senior Member, IEEE

Abstract—A rate-distortion (R-D) optimization rate control (RC) algorithm with adaptive \( Q_p \) initialization is presented for H.264. First, a linear distortion-quantization (D-Q) model is introduced and thus a close-form solution is developed to derive optimal quantization parameters (\( Q_p \)) for encoding each macroblock. Then we exploit to determine the initial \( Q_p \) efficiently and adaptively according to the content of video sequences. The experimental results demonstrate that the proposed algorithm can achieve better R-D performance than that of other two RC algorithms including the algorithm JVT-G012 which is the current recommended RC scheme implemented in the H.264 reference software JM9.5.

Index Terms—H.264, initial quantization parameter, rate control, rate-distortion optimization.

I. INTRODUCTION

RATe control (RC) aims to achieve good perceptual quality given the transmission bandwidth constraint. Usually, RC regulates the coded bit stream by adjusting \( Q_p \) whilst optimizing the video presentation quality. To achieve this, the rate-quantization (R-Q) model is often employed for representing the coded bits by means of \( Q_p \) and other parameters such as the variance of a residual macroblock (MB) \cite{1}, the mean absolute difference (MAD) of a residual MB \cite{2}–\cite{6}, and the percentage of zero quantized coefficients \cite{7}. Unfortunately, using other parameters such as MAD for R-Q modelling causes the chicken-and-egg dilemma \cite{8} to the state-of-the-art video coding standard H.264 \cite{9} because the Lagrangian method \cite{10} employed in H.264 needs \( Q_p \) available before mode decision but until the end of mode decision, RC cannot access the statistics such as MAD for determining \( Q_p \).

It is worthwhile to point out that several RC algorithms have been developed to overcome the H.264 RC chicken-and-egg dilemma. Ma et al. \cite{11} propose a two-pass RC scheme with each scheme applying a TM5 \cite{2} based method. If the first pass fails to obtain an appropriate \( Q_p \), the second pass is implemented as a refinement. The encoding complexity is consequently increased. Later, Ma et al. \cite{12} further develop an improved partial two-pass RC algorithm for RC performance improvement with a linear R-Q model being proposed. In \cite{8}, a one-pass RC algorithm JVT-G012 is proposed with a linear MAD prediction model being used to circumvent the chicken-and-egg dilemma and the conventional MPEG-4 Q2 R-Q model \cite{3} being employed to calculate \( Q_p \) given the allocated bits and predicted coding complexity. Due to its efficiency, JVT-G012 has been adopted in H.264 reference softwares. In \cite{13}, a RC algorithm is proposed for H.264 low delay video communications, where an improved MAD prediction model and a linear R-Q model are applied. In \cite{14}, the authors propose a linear distortion-quantization (D-Q) model for H.264 and hence develop a R-D joint optimized solution to \( Q_p \) determination.

The above RC algorithms focus on how to apply R-Q or D-Q models for improving the H.264 RC performance; however, another important factor influencing the RC performance, i.e., how to determine the initial \( Q_p \), is not well addressed. Traditionally, the start \( Q_p \) value is determined only depending on the bits per pixel (BPP) as JVT-G012 does. Although this initialization scheme is simple, yet it is not accurate enough. A better initialization scheme should not only depend on BPP, but also on the content of the video sequence. For instance, given the same BPP, a large start \( Q_p \) value is desired for video sequences with complex spatial details or high motion types; whereas, for video sequences with simple spatial contents or low motions, a small start \( Q_p \) value should be given. Unfortunately, there are few works about how to adapt the start \( Q_p \) value to the video contents.

In order to address the problems mentioned above, in this paper, we propose a R-D joint optimized RC algorithm with an adaptive initial \( Q_p \) determination scheme by extending our previous work \cite{14}. Firstly, the R-D joint optimized solution \cite{14} is applied to compute \( Q_p \) for each MB. Secondly, for determining the initial \( Q_p \), we utilize an entropy information and an INTRA16 dc mode \cite{15} based complexity measure of the first encoded frame. Thus, the proposed initial \( Q_p \) determination scheme is adaptive and easy to implement. We evaluate the proposed RC algorithm in two conditions by disabling or enabling the adaptive initialization scheme. The experimental results demonstrate that the proposed RC algorithm without enabling the adaptive initialization scheme can achieve better R-D performance than that of the algorithm \cite{12} and JVT-G012 \cite{8}. When the initialization scheme is enabled, the R-D performance of the proposed algorithm can be further improved. The rest of the paper is organized as follows. Section II presents the linear D-Q model and the R-D joint optimized \( Q_p \) solution. The initial \( Q_p \) determination scheme is also proposed herein. Then, we describe the proposed RC algorithm in Section III. Experimental results are shown in Section IV. Finally, Section V concludes this paper.

II. R-D JOINT OPTIMIZATION FOR \( Q_p \) DERIVATION AND INITIAL \( Q_p \) DETERMINATION

A. \( Q_p \) Calculation Based on R-D Optimization

In \cite{14}, a linear D-Q model is formulated as

\[
\text{MSE} = \rho Q_{\text{step}}^2
\]

where the mean squared error (MSE) represents the quality distortion, \( \rho \geq 0 \) is the model parameter; \( Q_{\text{step}} \) is the quantiza-
tion step size. In H.264, $Q_{\text{step}}$ doubles in size for every increment of 6 in $Q_P$. Fig. 1 illustrates the fitting accuracy of the linear D-Q model for two benchmark video sequences. Compared with the traditional distortion measure which is formulated as $\text{MSE} = Q_P^2 / 12$ [16], the linear D-Q model is simplified but suitable for H.264.

With the proposed linear D-Q model formulated in (1) and an improved MPEG-4 Q2 R-Q model [3], we develop a close-form solution to $Q_P$ determination in [14] as

$$Q_P^* = \left\{ \begin{array}{ll}
-\frac{\beta_i}{2\alpha_i} & 
\left( T + \frac{\beta_j^2}{2\alpha_j} - \gamma_i \right) \frac{\sum_{j=1}^{N} M_j - \sum_{j=1}^{N} H_j^p}{\alpha_i \sum_{j=1}^{N} M_j^{-1}} \right)^{-1} \\
+ \frac{1}{M_i} &
\end{array} \right. \quad (2)$$

where $Q_P^*$ is the resultant optimal quantization step size for MB $i$; $N$ is the total number of MBs in a frame; $\alpha_i$, $\beta_i$, and $\gamma_i$ are the improved MPEG-4 Q2 model parameters for MB $i$; $T$ is the bit budget for the remaining MBs of a frame; $H_j^p$ is the predicted header bit budget of the MB $j$ which is equal to that of the collocated MB in the previous frame of the same type; $M_i$ and $M_j$ are the predicted MAD values for MB $i$ and MB $j$, respectively, which are computed by the linear MAD prediction model [8] for circumventing the H.264 RC chicken-and-egg dilemma. Due to the simplicity, the sliding-window based data selection mechanism [4] is used to update the model parameters of the improved Q2 model and the linear MAD model according to the linear regression. It is observed that the D-Q model parameter $\rho$ is dissolved in (2). When $Q_P^*$ is obtained from (2), we can deduce the optimal quantization parameter $Q_P^*$ from $Q_P^*$ to encode the $i$th MB.

B. Initial $Q_P$ Determination

At the beginning of RC, a start $Q_P$ value must be given for encoding the first I-frame. Usually, the start $Q_P$ is determined only based on BPP as JVT-G012 [8] does as below

$$Q_P^* = \left\{ \begin{array}{ll}
40 & \text{BPP} \leq \text{Th}_1 \\
35 & \text{Th}_1 < \text{BPP} \leq \text{Th}_2 \\
25 & \text{Th}_2 < \text{BPP} \leq \text{Th}_3 \\
20 & \text{Th}_3 < \text{BPP} \leq \text{Th}_4 \\
15 & \text{BPP} > \text{Th}_4 \\
\end{array} \right. \quad (3)$$

where $Q_P^*$ is the start $Q_P$ value to encode the first I-frame; BPP = $u / (f \times N_p)$ with $u$, $f$ and $N_p$ being the bitrate, frame rate and total number of pixels in a frame, respectively; $\text{Th}_i$, $1 \leq i \leq 4$, are the predetermined thresholds and $\text{Th}_i < \text{Th}_j$ if $i < j$. For example, the threshold set is recommended as {0,0.1,0.3,0.6} for QCIF sequences and {0.1,0.2,0.6,1.2} for CIF sequences in JM9.5 [17].

Although the initialization scheme formulated in (3) is simple, yet it does not consider the difference between video sequences. A good initialization scheme should adapt itself to the specific content of video sequences, for example, given the same BPP, a large initial $Q_P$ is preferable for a video sequence with complex spatial contents or high motions; whereas, a video sequence with simple spatial contents or low motions results in a small initial $Q_P$ value. To achieve this, it is better to utilize the information of every frame within a video sequence to be encoded. However, this is impossible for real-time communication services where only the first I-frame is available for determining the initial $Q_P$ value.

In this work, we propose an initial $Q_P$ determination scheme which can be used for real-time communication applications by utilizing the entropy information denoted as EI and the INTRA16 dc mode [15] based complexity measure denoted as IM of the first I-frame. The EI and IM are defined as

$$\text{EI} = -\sum_{x=0}^{L-1} p(x) \log_2 p(x) \quad (4)$$

$$\text{IM} = \frac{\sum_{k=0}^{N-1} \sum_{j=0}^{M-1} |I_k(i,j) - I^*_k|}{1000 \times N} \quad (5)$$

where $L - 1$ is the maximum gray level, e.g., $L = 256$ for 8-bit pixel resolution; $p(x)$ is the probability of occurrence of gray level $x$; $N$ is the number of MBs in a frame; $I_k(i,j)$ is the pixel value at $(i,j)$ of the $k$th MB; $I^*_k$ is the predicted compensation value resulted from the INTRA16 dc mode [15].

In order to model the initialization scheme based on the parameter set {BPP, EI, IM}, three benchmark video sequences, namely “News,” “Foreman,” and “Mobile Calendar” are used because these three video sequences have different motion types from low to high and different spatial details from simple to complex and thus can represent a large number of video sequences. A large range of BPP values from 0.02 to 1.0 is employed to model the initialization scheme. For each of the test BPP, we try all the possible initial quantization parameters and record the best one which results in the best R-D point. Then, the relation between the best initial $Q_P$ and the parameter set {BPP, EI, IM} are studied. For example, the relation between the best initial $Q_P$ and BPP for the three CIF video sequences are shown in Fig. 2.
where we can partition the BPP values into three regions: region-1 (BPP \( \geq 0.4 \)), region-2 (0.4 > BPP \( \geq 0.2 \)), and region-3 (0.2 > BPP). In region-1 and region-2, two linear models are applied. In region-3, a second-order model is applied. Within each of the regions, the best initial \( Q_p \) increases along with the decrease of BPP. This is reasonable because the smaller the BPP is, the larger the initial \( Q_p \) is. The slope of the linear model in region-1 is smaller than the slope of the linear model in region-2. And the trend that the initial \( Q_p \) increases when BPP decreases is more obvious in region-3. It can be well understood that when BPP decreases (i.e., the allocated bit budgets decrease), the initial \( Q_p \) increases linearly. When the BPP decreases to some extent (e.g., BPP = 0.4), the initial \( Q_p \) increases linearly in a sharper manner. If the BPP further decreases to 0.2, the initial \( Q_p \) increases more sharply in a quadratic manner because the allocated bit budgets become much scarcer. Moreover, from Fig. 2, we see that the curves of these three video sequences are approximately superposable if moved vertically, i.e., after adding 5 to all the initial quantization parameters of “News” sequence, we can get the approximate curve of “Foreman” sequence; and after adding 12 to all the initial quantization parameters of “News” sequence, we can get the approximate curve of “Mobile Calendar” sequence. Those differences, i.e., 5 between “News” and “Foreman,” 12 between “News” and “Mobile Calendar,” depend on the complexity difference between video sequences. As a result, the following empirical rule is developed to determine the optimal initial \( Q_p \) as

\[
\mathcal{F}(\text{BPP}) + G(\text{E}, \text{I}, \text{M})
\]

where

\[
\mathcal{F}(\text{BPP}) = \begin{cases} 
     a_{11}\text{BPP} + a_{12}, & \text{BPP} \geq 0.4 \\
     a_{12}\text{BPP} + a_{22}, & 0.4 > \text{BPP} \geq 0.2 \\
     a_{31}\text{BPP}^2 + a_{32}\text{BPP} + a_{33}, & 0.2 > \text{BPP}
\end{cases}
\]

(7)

\[
G(\text{E}, \text{I}, \text{M}) = \max\{d_1\text{E}^2 + d_2\text{E} + d_3, e_1\text{I}^2 + e_2\text{I} + e_3\}
\]

(8)

for QCIF sequences, and

\[
\mathcal{E} = \{-15.21, 25, -30.41, 31, 462.47, -189.86, 44.60, -15.25, 241.81, -944.42, -0.76, 13.50, -49.18\}
\]

for CIF sequences.

### III. PROPOSED RATE CONTROL ALGORITHM

The proposed RC algorithm includes three different coding granularities, including the group of picture (GOP) level, frame level and MB level. At the GOP level and the frame level, the proposed RC algorithm uses the way of [12] to allocate target bits and perform the post-encoding regulation. Moreover, as used in [12] and [8], fixed \( Q_p \) values are employed to encode I-frames, the first P-frame and the first B-frame. For the first I-frame, the adaptive \( Q_p \) initialization scheme discussed in Section II-B is applied. In the following, due to the space limit, we focus on a step-by-step description of the proposed algorithm at the MB level for both P-frames and B-frames, which is also discussed in [14].

**Step 1) Initialization:** Let \( T_{fr} \) be the the target bit budget for the current frame. \( Q^*_p \) be the average \( Q_p \) value of the previous frame with the same frame type as the current frame, \( N \) be the total number of MBs in a frame. Let \( i = 1 \) and go to Step 2.

**Step 2) R-Q and MAD model update:** Update the improved MPEG-4 Q2 model parameters \( \alpha_k \) and \( \gamma_k \) and the linear MAD model parameters. If \( \alpha_k = 0 \), go to Step 5; otherwise, go to Step 3.

**Step 3) Parameter calculation:** Calculate \( M_j \) for \( i \leq j \leq N \). Compute \( \Omega = T_a + ((/\rho_j/\alpha_k)\gamma_k) \sum_{j=i}^{N} M_j - \sum_{j=i}^{N} H_j^p \). If \( \Omega < 0 \), go to Step 5; otherwise, go to Step 4.

**Step 4) Optimal \( Q^*_p \) determination:** Compute the optimal quantization step size \( Q^*_p \) for the current MB by using (2). Then deduce the quantization parameter \( Q^*_p \) from \( Q^*_p \). \( Q^*_p \) is adjusted as \( Q^*_p = \max\{Q^*_p - 2, \min\{Q^*_X + 2, Q^*_p\}\} \) to maintain the smoothness of visual quality. Then it is further bounded by \( Q^*_p = \max\{1, \min\{51, Q^*_p\}\} \) to get a finally selected and valid \( Q_p \). Go to Step 6.

**Step 5) \( Q^*_p \) determination:** If \( T_a \leq 0 \), let \( Q^*_p = Q^*_X + 2 \). Else if \( T_a > 0 \) and \( T_a - \sum_{j=i}^{N} H_j^p \leq 0 \), let \( Q^*_p = Q^*_X + 1 \). Else if \( T_a > 0 \) and \( T_a - \sum_{j=i}^{N} H_j^p > 0 \), let \( Q^*_p = Q^*_X - 1 \). Then, \( Q^*_p \) is further bounded by \( Q^*_p = \max\{1, \min\{51, Q^*_p\}\} \). Go to Step 6.

**Step 6) MB encoding:** Use \( Q^*_p \) to encode the current MB \( i \). Go to Step 7.

**Step 7) Post-Encoding:** After encoding the \( i \)th MB, record the actual MAD, total consumed bits and consumed header bits. The remaining target bits \( T_{fr} \) is updated by subtracting the total encoded bits of MB \( i \) from it. Go to Step 8.

**Step 8) Loop condition:** Let \( i = i + 1 \). If \( i \leq N \), go back to Step 2 to encode the next MB. Otherwise, the encoding process for the current frame comes to an end and the buffer fullness is updated. If the buffer fullness is larger than a predetermined threshold, called safety margin, for instance 80% of the buffer size, the next frame will be skipped.
IV. EXPERIMENTAL RESULTS

The H.264 reference software JM9.5 [17] is performed to evaluate the proposed RC algorithm. Fast motion estimation is enabled with 1/4 pixel resolution. The number of reference frames is set to 1 and the motion search range is 16. The R-D optimization is enabled and CABAC coding is enabled. As used by JVT-G012 [8] in JM9.5, the buffer size is set to $2.56 \times u$, where $u$ is the target bitrate. Six benchmark video sequences are tested with the parameter settings as follows. Three video sequences are of CIF format, they are “Foreman,” “News,” and “Garden.” The test target bitrates (units: kbps) for CIF video sequences are 2800, 2400, 2100, 1800, 1500, 1200, 1000, 800, 700, 600, 500, 400, and 300. The test frame rate (units: fps) for CIF format is 30. The other three video sequences are of QCIF format, they are “Coastguard,” “Paris,” and “Mobile Calendar.” The test target bitrates (units: kbps) are 380, 360, 340, 320, 280, 240, 200, 180, 150, 120, 100, 80, 60, 40, and 20. The test frame rate is 10 fps. All the six video sequences have 118 frames to be encoded and the GOP structure is IBBPBB with the period of I-frames equal to 4.

The algorithms [8] and [12] are utilized for comparison with the proposed RC algorithm. In order to evaluate the proposed initial $Q_p$ determination scheme, we test the proposed RC algorithm under two conditions. Under the first condition, the proposed initial $Q_p$ determination scheme is disabled, and the proposed RC algorithm employs the same $Q_p$ initialization method as that of JVT-G012 [8]. We denote this algorithm as PRC-1. Under the second condition, the proposed initial $Q_p$ determination scheme is enabled for performing the proposed RC algorithm as denoted as PRC-2.

First, the $Q_p$ calculation scheme in (2) is analyzed by comparing it with the finally selected $Q_p$ resulted from Section III Step 4. The MB percentage for the difference between the finally selected $Q_p$ and the calculated $Q_p$ is illustrated in Fig. 3 for two sequences. From the results, we observe that the MB percentage is over 50% on average for the difference between the final $Q_p$ and the calculated $Q_p$ being zero.

The video quality is evaluated in terms of peak signal-to-noise ratio (PSNR, units: dB). Note that the frame-skipping method is implemented for all the four algorithms. If the current buffer fullness exceeds 80% of the encoder buffer size, the encoder will skip encoding the next frame until the buffer fullness is lower than 80% of the encoder buffer size. When frame skipping occurs, the decoder displays the previous encoded frame in place of the skipped one. Therefore, the previous encoded frame is used in the PSNR calculation.

Due to the space limit, the R-D curves of only four sequences are shown in Fig. 4, where we can see that the proposed PRC-1 algorithm can achieve better R-D performance than that of the algorithm [12] and JVT-G012 [8]. When the proposed initial $Q_p$ determination scheme is enabled, the proposed PRC-2 algorithm can further improve the R-D performance as compared with PRC-1. As shown in the R-D curves, the algorithm [12] and JVT-G012 [8] degrade the R-D performance seriously as compared with the proposed algorithm for some video sequences. This is due to the reason that the algorithms [12] and [8] skip several frames during encoding those video sequences. For example, in Fig. 4(b), a significant performance drop is observed when JVT-G012 is used to encode the “Garden” sequence at the target bitrate 700 kbps, since 3 frames are skipped to prevent the buffer overflow. In order to show the number of frames skipped for each video sequence, we count up the number of frames skipped at each test target bitrates and give the results in Table I, where we can see that the proposed PRC-1 skips less amount of frames than the algorithms [12] and [8] and the proposed PRC-2 does not skip frames.

In order to evaluate the PSNR and bitrate performances quantificationally, the following two measurements are used in this paper

$$\Delta P = P_t - P_b, \Delta R = \frac{|R_t - R_b|}{R_t} \times 100\%$$

(9)

where $P_t$ is the PSNR performance of the test algorithm, $P_b$ is the PSNR performance resulted from JVT-G012 [8]; $R_t$ is the bitrate performance of the test algorithm and $R_b$ is the target bitrate. Due to the space limit, we average the $\Delta P$ and $\Delta R$ values obtained at various target bitrates for each of the test video sequences and list the average results in Table I.

From the average $\Delta P$ result, we can see that the proposed PRC-1 can achieve better PSNR results than the two algorithms [12] and [8]. When the proposed $Q_p$ initialization scheme is enabled, the PSNR performance is further improved by PRC-2. As for the average $\Delta R$ result, all the four algorithms obtain similar bitrate performances. In addition, we illustrate the buffer occupancy for two test cases in Fig. 5. From these plots, we can see that the propose algorithm PRC-1 can maintain suitable buffer occupancy levels which is similar to that of the algorithm [12], and both of these two algorithms achieve better performances than the algorithm [8] in terms of prevention of the buffer overflow (the buffer occupancy is higher than 80%) and underflow (the buffer occupancy is lower than 0%). When the proposed $Q_p$ initialization scheme is enabled, the stability of the buffer occupancy is further improved by PRC-2 as compared with PRC-1. Moreover, the frame-to-frame PSNR curves are shown in Fig. 6, where we can see that similar or better results can be obtained by the proposed PRC-1 algorithm than [12] and [8]. When the adaptive initialization scheme is enabled, the performance is further improved by PRC-2 as compared with PRC-1.
V. CONCLUSION

In this paper, a R-D joint optimized RC algorithm with adaptive \( Q_p \) initialization is presented for H.264. A linear model is proposed to model the D-Q relation and a close-form solution is applied to calculate optimal \( Q_p \) values for MB encoding. The proposed RC algorithm also employs an adaptive initial \( Q_p \) determination scheme which not only depends on BPP but also considers the specific content of video sequences. The experimental results have demonstrated that the proposed RC algorithm outperforms the other two RC algorithms [12] and [8] in terms of R-D performance, the number of frames skipped and the buffer overflow and underflow prevention.

ACKNOWLEDGMENT

The authors would like to thank Dr. S. Ma and Dr. Y. Lu for the useful discussion and the source code of the RC algorithm [12].

REFERENCES