A RATE-DISTORTION OPTIMIZATION ALGORITHM FOR RATE CONTROL IN H.264

Hanli Wang and Sam Kwong

Department of Computer Science, City University of Hong Kong, Hong Kong, P. R. China
E-mail: warghl@cs.cityu.edu.hk, cssamk@cityu.edu.hk

ABSTRACT

This paper presents a novel rate-distortion (R-D) joint optimization rate control (RC) algorithm for H.264 encoding. For RC in H.264, one of the most important topics is to model the R-D characteristics accurately. To achieve this, an efficient linear model is proposed to model the distortion-quantization (D-Q) relation. With the proposed linear D-Q model, a closed-form solution is derived to calculate the optimal quantization parameter for encoding each macroblock (MB). The proposed RC algorithm can be applied to both P-frames and B-frames. It is shown by experimental results that the proposed algorithm can control the bit rates accuratly with the R-D performance better than that of the RC algorithm JVT-G012 implemented in the H.264 reference software JM9.5.

***Index Terms*—** H.264, rate control, rate-distortion optimization, video encoding.

1. INTRODUCTION

Rate control plays a key role in video coding standards. The goal is to achieve good perceptual quality given the target bitrates. In general, a typical RC scheme includes two basic parts. One is bit allocation in which the best video quality is achieved by effectively distributing bit budgets among frames or MBs. The other is bit allocation achievement, i.e., to choose a quantization parameter (Q_p) for achieving the target bits for the current frame or MB accurately. In order to accomplish optimal RC performance, the rate-quantization (R-Q) model is usually used for modelling the encoded bits in terms of Q_p. Besides Q_p, other parameters are also required to represent the R-Q model, e.g. the mean absolute difference (MAD) of a residual MB in [1, 2]. For the video coding standards H.263 and MPEG-2/4, it is not difficult to obtain MAD to perform RC. However, it is not the case for the newest video coding standard H.264 [3] which applies the Lagrangian method to calculate the R-D cost for high coding performance. The Lagrangian method requires Q_p to be known before motion estimation and mode decision, thus leading to a chicken and egg dilemma because until the end of motion estimation and mode decision, RC cannot access the statistics such as MAD, which are essential to calculate Q_p.

Recently, several RC algorithms have been proposed to circumvent the dilemma for H.264 encoding. In [4], Ma *et al.* propose a two-pass RC scheme for H.264 encoding, with each scheme applying a TM5 [1] based method. If the first pass fails to obtain an appropriate Q_p, the second pass will be carried out as a refinement. As a result, the computational complexity of video encoding is increased. In [5], an improved partial two-pass RC algorithm is presented and a linear R-Q model is proposed. However, the linear R-Q model is not accurate enough to model the R-Q relation and thus leads to a degradation of RC performance. In [6], a one-pass RC algorithm JVT-G012 is proposed to execute RC for H.264 encoding. In this algorithm, a linear MAD model is used to predict coding complexity for the current frame or MB, and the conventional MEPG-4 Q2 R-Q model [2] is employed to calculate Q_p given the allocated bits and predicted coding complexity. Due to its simplicity, JVT-G012 has been recommended as a benchmark and adopted in H.264 reference softwares. However, JVT-G012 fails to achieve optimum bit allocation since the distortion is not explicitly considered into the R-Q model.

In order to further improve the H.264 RC performance, a R-D joint optimization algorithm is presented in this work. After conducting extensive experiments on a large number of video sequences, we study the statistics and derive a linear D-Q model for modelling the D-Q relation in H.264. The proposed model can be applied to both P-frames and B-frames. Based on an improved MPEG-4 Q2 R-Q model and the proposed D-Q model, a novel R-D joint optimization RC algorithm is proposed and implemented into the H.264 reference software JM9.5 [7]. The experimental results demonstrate that the proposed RC algorithm can achieve better R-D performance than that of JVT-G012 [6]. The rest of this paper is organized as follows. In Section 2, the linear D-Q model is presented. Then, we propose the joint optimization RC algorithm in Section 3. Experiments are shown in Section 4. Finally, Section 5 concludes this paper.

2. PROPOSED LINEAR D-Q MODEL

2.1. Relation between Q_p and Q_{step} in H.264

First, the difference between the quantization parameter Q_p and the quantization step size Q_{step} in H.264 should be clarified. Q_p denotes the quantization scale indirectly, whereas

---

This work is supported by City University of Hong Kong Strategic Grant 7001697.

Q_{step} is the true value used in quantization. In previous video coding standards, the relation between Q_p and Q_{step} is usually linear. In H.264, However, the relation is non-linear as shown in Fig. 1. A total of 52 values of Q_p ranging from 0 to 51 are supported in H.264. Q_{step} doubles in size for every increment of 6 in Q_p.

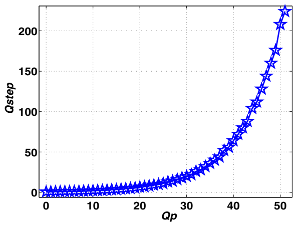


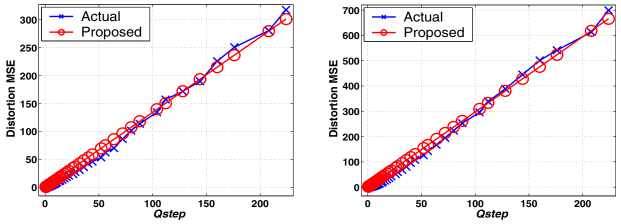
Fig. 1. Relation between Q_p and Q_{step} in H.264.

2.2. Proposed D-Q Model

Based on our observations on a large number of benchmark video sequences which are encoded by using H.264 JM9.5 [7], we find that there is a linear relation between the quality distortion in terms of mean squared error (MSE) and Q_{step}. This linear relation can be formulated as

$$MSE = \rho Q_{step} \tag{1}$$

where ρ is the model parameter. For illustration, the fitting accuracy of the linear D-Q model is shown in Fig. 2 for two benchmark video sequences. And the proposed D-Q model can be applied to both P-frames and B-frames.



(a) “Foreman” (b) “Mobile Calendar”

Fig. 2. Fitting accuracy of the proposed linear D-Q model .

3. PROPOSED RATE CONTROL ALGORITHM

3.1. Q_p Determination based on R-D Joint Optimization

In this work, an improved MPEG-4 Q2 model is employed to model the R-Q relation as formulated as

$$\frac{R_t}{M} = \frac{\alpha}{Q_{step}^2} + \frac{\beta}{Q_{step}} + \gamma \tag{2}$$

where R_t is the number of bits used for texture coding of an MB, M is the corresponding MAD, α, β and γ are the model parameters. To address the chicken and egg dilemma related to MAD, the linear model proposed in [6] is adopted here to predict MAD of the current MB as

$$M = \theta M^p + \phi \tag{3}$$

where M^p is MAD of the colocated MB in the previous frame of the same type, θ and ϕ are model parameters. The sliding-window based data selection mechanism [8] is used to calculate the model parameters of the improved Q2 model and the linear MAD model according to the linear regression analysis. Based on the proposed D-Q model in (1) and the improved R-Q model in (2), a mathematical closed-form solution to the R-D optimized RC is addressed as follows. For notational simplicity, we use Q to denote Q_{step} in the rest of the paper.

Given the target bits T for left MBs of a frame, we want to find an expression for the quantization parameters that minimizes the distortion in (1). This minimization problem can be converted into the following maximization problem as

$$Q_i^*, \cdots, Q_N^* = \underset{\substack{Q_i, \cdots, Q_N \\ \sum_{j=i}^N T_j = T}}{\arg\max} \frac{1}{N-i+1} \sum_{j=i}^{N} \rho_j^{-1} \times Q_j^{-1} \tag{4}$$

where N is the total number of MBs in a frame. Initially, $i = 1$. According to the Lagrange multiplier method and the R-Q model in (2), we can express the optimization problem in (4) in its equivalent form as

$$\begin{aligned} Q_i^*, \cdots, Q_N^*, \lambda^* &= \underset{Q_i, \cdots, Q_N, \lambda}{\arg\max} \left\{ \frac{\sum_{j=i}^N \rho_j^{-1} \times Q_j^{-1}}{N-i+1} + \lambda \left(\sum_{j=i}^N T_j - T \right) \right\} \\ &= \underset{Q_i, \cdots, Q_N, \lambda}{\arg\max} \left\{ \tfrac{1}{N-i+1} \sum_{j=i}^N \rho_j^{-1} \times Q_j^{-1} + \lambda \left[\sum_{j=i}^N (H_j^p \right.\right. \\ &\qquad \left.\left. + \alpha_j M_j Q_j^{-2} + \beta_j M_j Q_j^{-1} + \gamma_j M_j) - T \right] \right\} \end{aligned} \tag{5}$$

where H_j^p is the predicted header bit budget of the jth MB which is equal to that of the colocated MB in the previous frame of the same type.

After mathematical manipulations, we obtain the optimal quantization step size in (6).

$$Q_i^* = \left\{ -\frac{\beta_i}{2\alpha_i} + \frac{\rho_i^{-1}}{\alpha_i M_i} \sqrt{\frac{T + \sum_{j=i}^N \left(\frac{\beta_j^2 M_j}{4\alpha_j} - \gamma_j M_j - H_j^p \right)}{\sum_{j=i}^N \alpha_j^{-1} M_j^{-1} \rho_j^{-2}}} \right\}^{-1} \tag{6}$$

where $1 \leq i \leq N$. Note that when encoding the ith MB, the parameters α_k, β_k, γ_k, ρ_k and M_k for $i < k \leq N$ are not available. To address this problem, we use α_i, β_i, γ_i and ρ_i to approximate the corresponding parameters with index k satisfying $i < k \leq N$. As for M_k, $i \leq k \leq N$, we employ the current MAD model parameters for calculation, i.e., $M_k = \theta_i M_k^p + \phi_i$, $i \leq k \leq N$. As a result, formula (6) can be simplified as

$$Q_i^* \approx \left\{ -\frac{\beta_i}{2\alpha_i} + \frac{1}{M_i} \sqrt{\frac{T + \left(\frac{\beta_i^2}{4\alpha_i} - \gamma_i \right) \sum_{j=i}^N M_j - \sum_{j=i}^N H_j^p}{\alpha_i \sum_{j=i}^N M_j^{-1}}} \right\}^{-1} \tag{7}$$

It is observed that the D-Q model parameter ρ is dissolved in (7). When Q_i^* is obtained from (7), we can deduce ${Q_p}_i^*$ from Q_i^* to encode the ith MB.

3.2. Rate Control Algorithm

The framework of the proposed RC algorithm consists of three different coding granularities, including the GOP level, frame level and MB level. The theoretical foundation behind the proposed RC algorithm is the aforementioned R-D optimization method formulated in (7) which can be applied to both P-frames and B-frames. Similar to [5], the proposed RC algorithm employs the TM5 based method for allocating target bits at the GOP level and frame level. We also follow the way of [5] to use adaptive fixed Q_p values to encode I-frames, the first P-frame and the first B-frame. In the following, due to the space limit, we focus on a step-by-step description of the proposed algorithm at the MB level.

Step 1. Allocate the target bits to the current frame, T. Denote Q_p^a as the average Q_p value of the previous frame with the same frame type as the current frame. Let $i = 1$ and go to Step 2.

Step 2. Update the improved MPEG-4 Q2 model parameters α_i, β_i and γ_i. Update the linear MAD model parameters θ_i and ϕ_i. If $\alpha_i = 0$, go to Step 5; otherwise, go to Step 3.

Step 3. Use θ_i and ϕ_i to calculate M_j for $i \le j \le N$. Compute $\Omega = T + \left(\frac{\beta_i^2}{4\alpha_i} - \gamma_i\right)\sum_{j=i}^{N} M_j - \sum_{j=i}^{N} H_j^p$. If $\Omega < 0$, go to Step 5; otherwise, go to Step 4.

Step 4. Compute the optimal quantization step size Q_i^* for the current MB by using the formula (7). Then deduce the quantization parameter ${Q_p}_i^*$ from Q_i^*. ${Q_p}_i^*$ is adjusted as ${Q_p}_i^* = \max\{Q_p^a - 2, \min\{Q_p^a + 2, {Q_p}_i^*\}\}$ to maintain the smoothness of visual quality. Then, it is further bounded by ${Q_p}_i^* = \max\{1, \min\{51, {Q_p}_i^*\}\}$ to get a valid Q_p value for H.264 encoding. Go to Step 6.

Step 5. If $T \le 0$, let ${Q_p}_i^* = Q_p^a + 2$. Else if $T > 0$ and $T - \sum_{j=i}^{N} H_j^p \le 0$, let ${Q_p}_i^* = Q_p^a + 1$. Else if $T > 0$ and $T - \sum_{j=i}^{N} H_j^p > 0$, let ${Q_p}_i^* = Q_p^a - 1$. Then, ${Q_p}_i^*$ is further bounded by ${Q_p}_i^* = \max\{1, \min\{51, {Q_p}_i^*\}\}$. Go to Step 6.

Step 6. Use ${Q_p}_i^*$ to encode the current MB i. Go to Step 7.

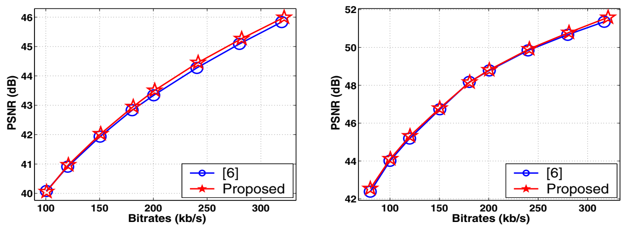
Step 7. After encoding the ith MB, record the actual MAD, total consumed bits and consumed header bits. The remaining target bits T is updated by subtracting the total encoded bits of MB i from it. Go to Step 8.

Step 8. Let $i = i + 1$. If $i \le N$, go back to Step 2 to encoder the next MB. Otherwise, the encoding process for the current frame comes to an end and the buffer fullness is updated. If the updated buffer fullness is larger than a predetermined threshold, called safety margin, for instance 80% of the buffer size, the next frame will be skipped.

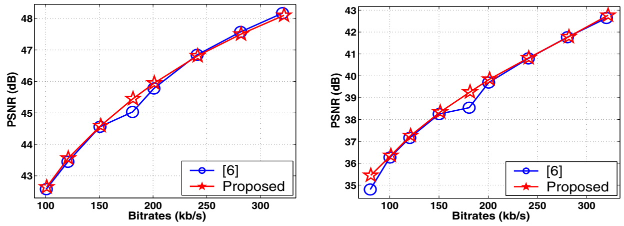
4. EXPERIMENTAL RESULTS

The H.264 reference software JM9.5 [7] is performed to evaluate the proposed RC algorithm. The 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 for mode selection and CABAC coding is enabled. Six benchmark video sequences are used, each of which is of QCIF format (176×144). They are “Foreman”, “News”, “Carphone”, “Coastguard”, “Mobile Calendar” and “Paris”. These video sequences have different motion types from low to high and different spatial details from simple to complex. All the video sequences have 118 frames to be encoded and the GOP structure is IBBPBB with the period of I-frames equal to 4. The test frame rate (units: fps) is 10. And various test target bitrates (units: kb/s) are employed.

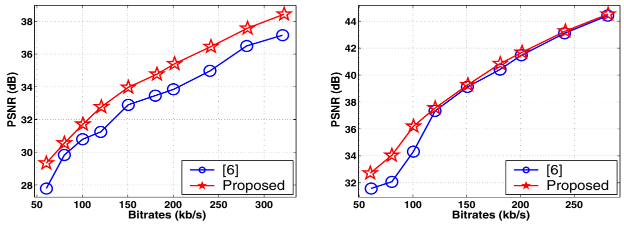
The RC algorithm JVT-G012 [6] is utilized for comparison with the proposed RC algorithm. The video quality is evaluated in terms of Peak Signal to Noise Ratio (PSNR, units: dB). Note that the frame-skipping method is implemented for both of the two test algorithms. If the current buffer fullness exceeds 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.



(a) “Foreman” (b) “News”



(c) “Carphone” (d) “Coastguard”



(e) “Mobile Calendar” (f) “Paris”

Fig. 3. R-D curves.

The R-D curves are shown in Fig. 3, where we can see that the proposed RC algorithm can achieve better R-D performance than that of JVT-G012. As obvious as shown in the R-D curves, JVT-G012 will degrade the R-D performance seriously as compared with the proposed algorithm for some test video sequences, for instance, “Mobile Calendar” and “Paris”. This is due to the reason that JVT-G012 skip several frames during encoding those video sequences. 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 1, where we can see that the proposed algorithm does not skip frames.

Table 1. Comparison of Total Number of Frames Skipped.

Video Sequence	[6]	Proposed
“Foreman”	0	0
“News”	0	0
“Carphone”	0	0
“Coastguard”	5	0
“Mobile Calendar”	46	0
“Paris”	14	0
Total	**65**	**0**

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

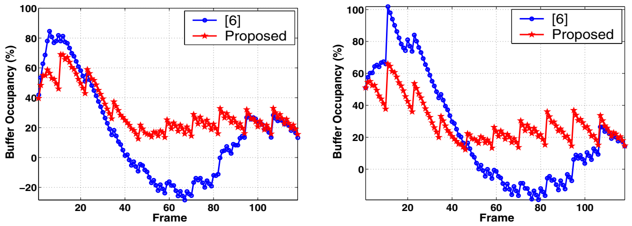
$$\Delta P = P_t - P_b, \quad \Delta R = \frac{R_t - R_b}{R_b} \times 100\% \tag{8}$$

where P_t is the PSNR performance of the test algorithm, P_b is the PSNR performance resulted from JVT-G012; R_t is the bitrates performance of the test algorithm and R_b is the target bitrates. The average ΔP and ΔR values for the test video sequences are shown in Table 2.

Table 2. Comparison of average ΔP (**dB**) and ΔR (%).

Video Sequence	ΔP (dB)		ΔR (%)	
	[6]	Proposed	[6]	Proposed
“Foreman”	0	0.12	0.16	0.68
“News”	0	0.10	-0.05	0.31
“Carphone”	0	0.08	0.43	0.74
“Coastguard”	0	0.22	0.19	0.78
“Mobile Calendar”	0	1.25	0.23	0.64
“Paris”	0	0.70	0.59	0.61
Average	**0**	**0.41**	**0.26**	**0.63**

From the average ΔP and ΔR results, we can see that the proposed algorithm can achieve better PSNR results than that of JVT-G012, while both of the two algorithms obtain similar bitrates performances. In addition, we show the buffer occupancy for some test cases in Fig. 4. From the plot, we can see that the propose algorithm can maintain suitable buffer occupancy levels which is better than JVT-G012 in terms of prevention of the buffer overflow (the buffer occupancy is higher than 80%) and underflow (the buffer occupancy is lower than 0%).



(a) “Mobile Calendar”, 180 kb/s (b) “Paris”, 80 kb/s

Fig. 4. Buffer occupancy.

5. CONCLUSIONS

In this paper, a novel RC algorithm is presented to improve the R-D performance of H.264 encoding. A linear model is firstly proposed to model the D-Q relation by studying the statistics on a large number of video sequences. The proposed D-Q model can be employed for both P-frames and B-frames. Then, based upon the proposed linear D-Q model and an improved MPEG-4 Q2 R-Q model, we deduce a simple closed-form solution to the problem of calculating optimal Q_p which is used to encode the current MB prior to mode decision. The experimental results have demonstrated that the proposed RC algorithm is more accurate and robust than JVT-G012 [6], the current standardized RC algorithm, in terms of R-D performance, the number of frames skipped and the buffer overflow and underflow prevention.

6. REFERENCES

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

[2] T. Chiang and Y. Q. Zhang, “A New Rate Control Scheme Using Quadratic Rate Distortion Model”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246-250, Feb. 1997.

[3] “Advanced Video Coding for Generic Audiovisual Services”, *ITU-T Rec. H.264*, Mar. 2005.

[4] S. Ma, W. Gao, Y. Lu, and D. Zhao, “Improved Rate Control Algorithm”, *Document JVT-E069*, Geneva, Oct. 2003.

[5] S. Ma, W. Gao, and Y. Lu, “Rate-Distortion Analysis for H.264/AVC Video Coding and Its Application to Rate Control”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533-1544, Dec. 2005.

[6] Z. Li, F. Pan, K. P. Lim, et al., “Adaptive Basic Unit Layer Rate Control for JVT”, *Document JVT-G012-r1*, Thailand, Mar. 2003.

[7] H.264/AVC Reference Software JM9.5, [Online] Available: http://iphome.hhi.de/suehring/tml/.

[8] H. J. Lee, T. Chiang, and Y. Q. Zhang, “Scalable Rate Control for MPEG-4 Video”, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878-894, Sept. 2000.