

A Novel Two-Stage Rate Control Scheme for H.264

Do-Kyoung Kwon, Mei-Yin Shen and C.-C. Jay Kuo

Department of Electrical Engineering and Integrated Media Systems Center

University of Southern California, Los Angeles, CA 90089-2564

E-mails: dokwon@usc.edu, meiyinsh@sipi.usc.edu, cckuo@sipi.usc.edu

Abstract— A new rate control algorithm for the H.264 encoder is developed in this work. There are several unique features in the proposed rate control scheme. First, a two-stage encoding mechanism is used to resolve the problem of inter-dependency between R-D optimization (RDO) and rate control (RC) in H.264. Second, it is emphasized that the bits of the header information may occupy a larger portion of the total bit budget especially when encoding at low bit rates in H.264. To address this issue, a rate model for the header information is established so as to estimate header bits more accurately. Third, a new source rate model is proposed to estimate the quantization parameter (QP) from the residual signal. Built upon the above ideas, a rate control algorithm is developed for the H.264 baseline-profile encoder under the constant-bit-rate (CBR) constraint. It is shown by experimental results that the proposed scheme can control bit rates accurately with the R-D performance better than that of the rate control algorithm implemented in H.264 JM8.1a.

I. INTRODUCTION

The H.264 video standard has gained much attention recently due to its substantial coding gain over existing coding standards. An efficient rate control algorithm for the H.264 encoder is important to its practical use under various transmission channels. Among several new coding technologies adopted, the RDO process with various intra- and inter-prediction modes and multiple reference frames provides a major contribution to the high coding efficiency of H.264. However, it also makes rate control of H.264 difficult due to the inter-dependency problem between RDO and RC, which is described as a “chicken and egg” dilemma in [1].

Several model-based rate control algorithms have been proposed for H.264. In [1], the mean of absolute differences (MAD) of each basic unit in a current frame was estimated by the MAD of the collocated basic unit in a previous frame using a linear model. Then, a quadratic rate model was employed to determine the QP of a basic unit, which can be either a frame, a slice or an MB. In [2], the residue of each picture was first estimated by performing RDO with a reduced set of intra- and inter-prediction modes and the number of reference frame was limited to one. After that, the standard deviation of the estimated residue was fed into the H.263 TMN8 R-D models to determine a QP. In [3]–[5], similar rate control algorithms were also proposed for H.264.

However, the above algorithms have the following limitations. First, the exact residual signal is not available before RDO. Second, the header information such as MB modes, MVs and reference frames are not available before RDO, either. Most previous research attempted to resolve the RDO-RC inter-dependency problem by examining the MAD or variance of the residual signal. However, due to various coding

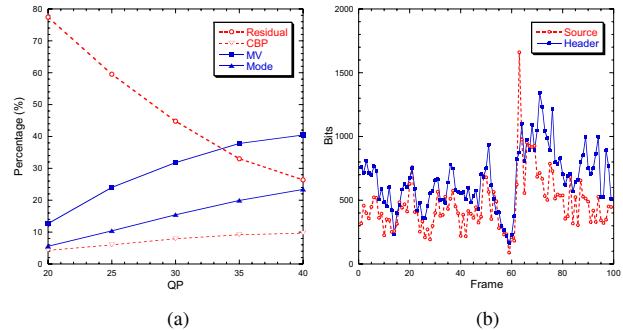


Fig. 1. (a) The percentage of header (MV + Mode) bits at various QP values and (b) header bits when $QP=35$ as a function of the frame number. P frames of the QCIF “Foreman” sequence are encoded with a single reference frame.

options, bits associated with the header information vary a lot from frame to frame, and they may occupy a larger portion of the total bit budget than source bits. The impact of header bits becomes more obvious at low bit rates. Therefore, an accurate estimate of header bits is also critical to rate control of H.264. These issues are addressed in detail in this work.

The rest of this paper is organized as follows. In Sec. II, a two-stage encoding scheme is presented. For the H.264 baseline-profile encoder, details of the proposed rate control algorithm with the new header and source rate models are described in Sec. III. Experimental results are presented in Sec. IV. Finally, concluding remarks are given in Sec. V.

II. TWO-STAGE ENCODING SCHEME

The inter-dependency of RDO and RC is the major difference between the rate control problem for H.264 and prior standards. It affects header bits for the header information as well as source bits for residual signals. Thus, the decoupling of RDO and RC is a key issue in the rate control of H.264. In this section, we propose a two-stage encoding scheme to decouple these two operations.

A. Two Observations

While it has been known that residual signals or the information related with them are essential to model-based rate control algorithms, the header information has been thought to be less important in previous standards such as MPEG-1/2, H.263 and MPEG-4. In the previous standards, the amount of header bits nearly constant and small as compared with that of source bits. Thus, it can be easily estimated by the average header bits of previous frames. In contrast, the size of

header bits in H.264 can no longer be ignored. Fig. 1 shows the percentage of header bits at various QP values for the QCIF “Foreman” sequence. It also shows the variation of header bits when it is encoded using the QP value of 35. We see from Fig. 1 that header bits occupy a large portion of the total amount of bits. When an input sequence is encoded at very low bit rates, the amount of header bits even exceeds that of source bits. It is also worthwhile to point out that the size of header bits fluctuates significantly from frame to frame, which means header bits cannot be estimated simply by the average value of previous frames.

To resolve the inter-dependency of RDO and RC, we propose a two-stage encoding scheme based on two observations. Let QP_n and QP_{n-1} be the average QP values of the n -th and the $(n-1)$ -th frames, and QP_1 and QP_2 be QP values for RDO and quantization, respectively. Then, we observe the following two interesting phenomena.

1. The variation between the QP values of two consecutive frames is usually restricted to a small range for smooth video playback; namely,

$$|QP_n - QP_{n-1}| \leq \Delta, \text{ where } \Delta \leq 3. \quad (1)$$

2. The decrease of the coding gain is not much even though QP_1 and QP_2 are different as long as their difference is restricted to a small range. Thus, we have

$$|QP_1 - QP_2| \leq \Delta, \text{ where } \Delta \leq 3. \quad (2)$$

The first observation is often exploited in model-based rate control algorithms [1]–[5] including the proposed one to smooth the quality variation between frames and/or between MBs. In H.264, given a pre-determined QP, the RDO process performs motion estimation and mode decision so as to choose the best set of reference frames, MVs and a MB mode by minimizing the following Lagrange cost:

$$J(S_{QP}) = D(S_{QP}) + \lambda(QP) \cdot R(S_{QP}), \quad (3)$$

where $\lambda(QP)$ is a Lagrange multiplier, which depends on the pre-determined QP value [6]. The second observation says that even though QP_1 and QP_2 are different, the loss in the coding gain after RDO is not significant as long as the difference between them is small. Experiments show that the coding gain loss is around 0.2 dB when $|QP_1 - QP_2| = 3$ and that it is negligible when $|QP_1 - QP_2| = 1$.

B. Two-Stage Encoder Structure

The proposed two-stage encoding scheme is shown in Fig. 2. In the first stage, QP_1 is initially set to be the average QP_2 of a previous frame, and perform RDO for all MBs in a current frame using QP_1 to determine the residual signal and the header information. The residual signals of all MBs go through DCT/Q and IQ/IDCT to get a reconstructed frame, which is required for intra predictions of subsequent MBs. After the first stage, given the target bits to the current frame, QP_2 values of all MBs are determined by the proposed rate control algorithm such that $|QP_1 - QP_2| = 3$ (to be discussed in Sec. III). Finally, the residual signal of each MB are encoded using QP_2 in the second stage.

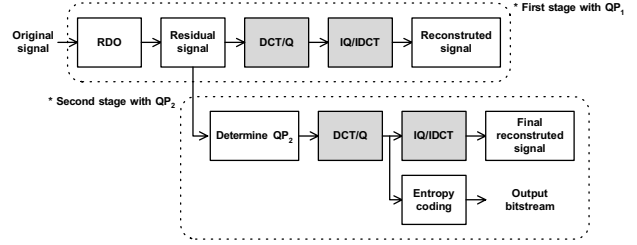


Fig. 2. The proposed two-stage encoding scheme for rate control of H.264.

In the second stage, if an MB is inter-coded, its residual signal is simply re-quantized using QP_2 . However, if an MB is intra-coded, the residual signal should be updated since its neighboring pixels can be different from those in the first stage. For such a case, the residual signal is updated assuming the same intra mode determined in the first stage of encoding. Note that the high coding complexity of the H.264 encoder comes mainly from RDO. The proposed two-stage encoding scheme only requires one extra DCT/Q and IQ/IDCT performed in the first stage. Since RDO is performed only once, the additional encoding complexity demanded by the two-stage encoding scheme is small as compared with the overall H.264 encoding complexity.

III. PROPOSED RATE CONTROL ALGORITHM

A. Rate Model for Header Bits

Let R_T be the total number of bits allocated to a frame, and let R_{src} and R_{hdr} be source and header bits of the frame, respectively. For rate control, the available source bits should be identified first by estimating the required header bits after RDO. A novel rate model for header bits is proposed here. To estimate header bits, every MB in a frame is classified into two classes (*i.e.*, inter MBs and intra MBs) and the header bits of these two classes are estimated separately.

For inter MBs, their header bits have a strong relationship with the number of non-zero horizontal/vertical MV elements, N_{nzMVe} , and the number of MVs, N_{MV} . To give an example, suppose an MB is partitioned into four 8x8 blocks and the four MVs are (4, 1), (2, 0), (3, 7) and (0, 0). Then, $N_{nzMVe} = 5$ (*i.e.* 4, 1, 2, 3 and 7) and $N_{MV} = 4$. We observe from experiments that header bits of inter MBs in a frame can be estimated by the following linear model:

$$R_{hdr,inter} = \gamma \cdot (N_{nzMVe} + \omega \cdot N_{MV}), \quad (4)$$

where γ is a model parameter and ω is a weighting factor that depends on the number of reference frames. The value of ω is set to 0.5, if the number of reference frames is larger than 4. It is set to 0.4, if the number of reference frames is equal to 3 or 4. It is set to 0.3, if the number of reference frames is equal to 1 or 2. The relationship between $R_{hdr,inter}$ and $(N_{nzMVe} + \omega \cdot N_{MV})$ for the “Table Tennis” sequence is plotted in Fig. 3, which confirms the rate model as given in (4). The same relationship is observed for many other sequences. For intra MBs, header bits of intra MBs in a frame can be estimated simply via

$$R_{hdr,intra} = N_{intra} \cdot b_{intra}, \quad (5)$$

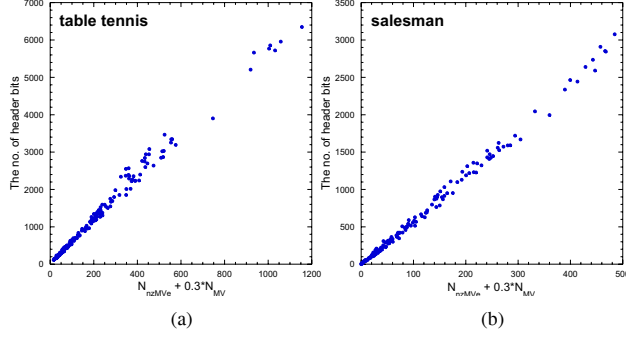


Fig. 3. The relationship between $R_{hdr, inter}$ and $(N_{nzMVE} + \omega \cdot N_{MV})$ for the QCIF sequences, (a) “Table Tennis” and (b) “Salesman”. For each sequence, 30 frames are encoded using every five QP from 15 to 40 with a single reference frame.

where N_{intra} is the number of intra MBs in a frame and b_{intra} is the average number of header bits of intra MBs. Finally, the total number of header bits of a frame is equal to $R_{hdr, inter} + R_{hdr, intra}$.

B. Rate Model for Source Bits

A new source rate model is proposed for H.264 here. The basic unit for DCT and quantization is a 4x4 block in H.264. Accordingly, the 4x4 block is the smallest unit that can be either coded or skipped. For instance, a 4x4 block will be skipped if all of 4x4 DCT coefficients are zero after quantization. Otherwise, it is a coded block. In our proposed scheme, all 4x4 blocks in a frame are first classified into either skipped blocks or coded blocks. Since no bit is required for skipped blocks, the source bits can be estimated by focusing on coded blocks. More specifically, let $SATD_c$ be the sum of absolute transform differences (SATD) of coded 4x4 blocks. Then, the source rate can be estimated via

$$R_{src}(Q) = \alpha \cdot \frac{SATD_c(Q)}{Q}, \quad (6)$$

where α is the model parameter and Q is the quantization stepsize. This is confirmed by the plot in Fig. 4. Note that $SATD_c$ has similar characteristics with the MAD of coded 4x4 blocks. However, it is used in the proposed model due to its slightly better performance in the source rate model.

C. Coded Block Identification

In order to apply the proposed rate model properly, we have to identify whether a 4x4 block will be coded or not at different QP values. Let $X(i, j)$ and $X_q(i, j)$ denote the DCT coefficients at the (i, j) position before and after quantization. In H.264, coefficient $X(i, j)$ is quantized as [7]

$$X_q(i, j) = \text{sign}\{X(i, j)\} \cdot \left[(|X(i, j)| \cdot A(Q_M, i, j) + f \cdot 2^{17+Q_E}) \gg (17 + Q_E) \right], \quad (7)$$

where $Q_M \equiv QP \bmod 6$ and $Q_E \equiv QP/6$. As indicated above, the quantization process is done differently according to a QP value, which makes the ρ -domain rate model [8] difficult to be applied in H.264 rate control. In the proposed algorithm,

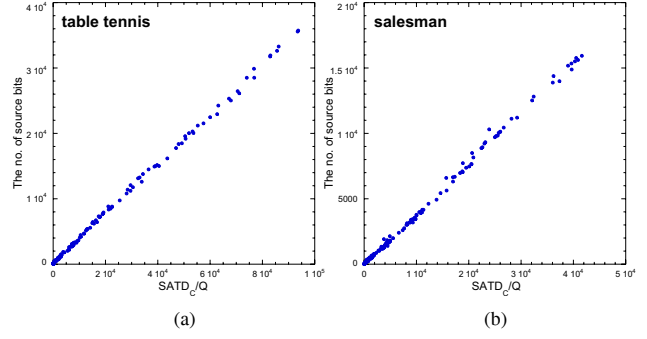


Fig. 4. The relationship between R_{src} and $SATD_c/Q$ for the QCIF sequences, (a) “Table Tennis” and (b) “Salesman”. For each sequence, 30 frames are encoded using every five QP from 15 to 40 with a single reference frame.

coded blocks are identified in the first stage of encoding. To see whether a 4x4 block is coded or not, we first identify X_{max} which satisfies the following condition:

$$X_{max} = \arg \max_{1 \leq i, j \leq 4} |X(i, j)| \cdot A(Q_M, i, j), \quad (8)$$

where $Q_M \equiv QP_1 \bmod 6$. Then, X_{max} is quantized again using all of the QP values ranging from $QP_1 - 3$ to $QP_1 + 3$ since the difference between QP_1 and QP_2 is restricted to 3. If X_{max} is not zero after quantization at the particular QP, the corresponding 4x4 block is a coded block.

D. Rate Control Algorithm

Based on the two-stage encoding scheme along with the new rate models, we propose an algorithm for constant bit rate (CBR) control of the baseline profile encoder, which is useful to real-time applications such as video conferencing. The proposed algorithm is performed as follows:

- STEP 1.** Allocate the target bits to the current frame, R_T .
- STEP 2.** Perform the first stage of encoding using QP_1 to get the residual signal and the header information. The coded 4x4 blocks are identified and $SATD_c(Q)$ of all MBs are computed for $QP \in (QP_1-3, QP_1+3)$.
- STEP 3.** Estimate the header bits using Eqs. (4) and (5) so that the available source bits R_{src} can be obtained.
- STEP 4.** Suppose that the current MB number is m . Let R_{src} be the available source bits before encoding the m -th MB. In this step, QP_2 of the m -th MB is determined as follows. First, compute the total $SATD_c(Q)$ of remaining MBs for $QP \in (QP_1-3, QP_1+3)$. Using (6) and $SATD_c(Q)$ values, we can estimate the source bits, $\hat{R}_{src}(Q)$ for $QP \in (QP_1-3, QP_1+3)$. Next, find the QP that minimizes the distance between R_{src} and the estimated source bits, i.e.,

$$QP_{2,m} = \arg \min_{q \in (QP_1-3, QP_1+3)} |\hat{R}_{src}(Q(q)) - R_{src}|, \quad (9)$$

where $Q(q)$ is a quantization stepsize corresponding to q . Finally, encode the residual signal of the m -th MB using $QP_{2,m}$ and update R_{src} by subtracting the actual source bits of the m -th MB from it.

- STEP 5.** Repeat Step 4 for all MBs in the current frame.

STEP 6. Update model parameters γ and α as well as the buffer status, and then move to the next frame.

IV. EXPERIMENTAL RESULTS

The proposed algorithm is implemented in the H.264 reference software, JM8.1a with the baseline profile. Since the main objective of our experiments is to show the advantages of the two-stage encoding scheme and the proposed rate models, the proposed algorithm is implemented with a simple bit allocation method assuming a single GOP structure. It is compared with that in JM8.1a, which is based on [1]. More specifically, 280 frames are encoded as P frames with a single reference frame except for the first I frame at frame rates of 30. The first I frame is encoded using the QP determined by the same rule in JM8.1a for fair comparison. The target bits for subsequent P frames are allocated by the following rule. Let $N_r(n)$ and $T_r(n)$ be the number of remaining frames and the number of remaining bits before encoding the n -th frame, respectively. Then, the target bits for the n -th frame is simply determined by $T_r(n)/N_r(n)$. The initial values of model parameters γ and α in (4) and (6) are set to 6.0 and 0.4, respectively, for the first P frame in the proposed algorithm.

Various QCIF sequences are encoded at 64 Kbps. The buffer size is set to two times of the channel rate such that no buffer overflow is caused by the RC algorithm in JM8.1a. Table I shows the experimental results, where we can see that the proposed algorithm can meet the target bit rate more closely. Moreover, better picture qualities are achieved by the proposed algorithm. Fig. 5 shows the variations of bits as a function of the frame number. Since the residual signal and the header information are available after the first stage of encoding and the proposed rate models are accurate, the proposed algorithm can control the bit rates of all frames constantly.

TABLE I
COMPARISON OF TWO RATE CONTROL ALGORITHM

Sequences	JM81a		Proposed	
	Bitrate	PSNR	Bitrate	PSNR (gain)
News	64.15	34.48	63.99	34.87 (+0.39)
Carphone	64.09	32.43	64.01	32.74 (+0.31)
Mot & Dau	64.04	37.73	64.00	38.01 (+0.28)
Salesman	64.04	35.36	64.01	35.59 (+0.23)
Foreman	64.12	31.53	64.00	31.79 (+0.26)
Silent	64.03	33.54	64.00	33.56 (+0.02)

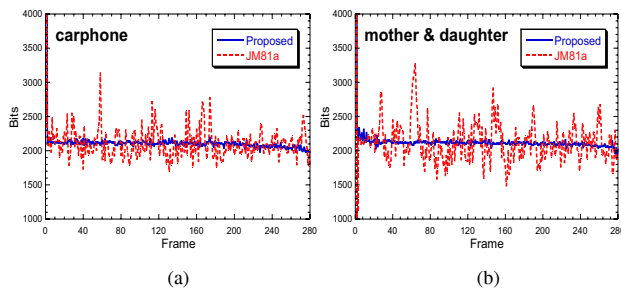


Fig. 5. Bits per frame by the proposed algorithm and that in JM8.1a, respectively, for the QCIF sequences. (a) "Carphone" and (b) "Mot & Dau".

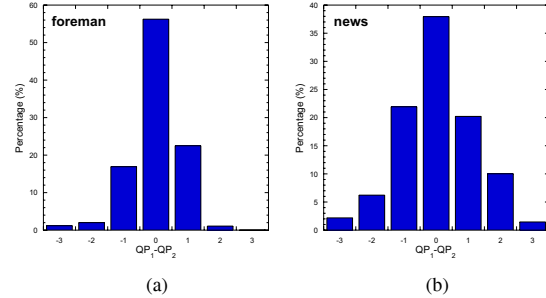


Fig. 6. The distributions of $QP_1 - QP_2$ by the proposed algorithm for the QCIF sequences, (a) "Foreman" and (b) "News" at 64 Kbps.

Fig. 6 shows the distribution of $QP_1 - QP_2$ for different sequences. For the "Foreman" sequence, the probabilities of $|QP_1 - QP_2| \leq 1$ and $|QP_1 - QP_2| \geq 3$ are larger than 95% and less than 2%, respectively. For the "News" sequence, these probabilities are larger than 80% and less than 4%, respectively. Generally, the difference between QP_1 and QP_2 is small for most frames in a sequence whether the sequence is of high activity or not. It implies that the average QP_2 of a previous frame provides a good estimate of QP_1 of a current frame and the coding gain loss by the two-stage encoding is very small in this application with the baseline profile.

V. CONCLUSION

A new model-based RC algorithm for H.264 based on the two-stage encoding scheme was proposed. The RDO-RC inter-dependency problem is resolved by the two-stage encoding scheme at the cost of little extra encoding complexity. Moreover, both source and header bits can be well estimated using the proposed rate models. Thus, the proposed RC algorithm can control bit rates accurately with improved R-D performances as compared with the RC algorithm in JM8.1a. Future work includes rate control for different video coding applications, for example, non-conversational applications that requires a finite GOP structure with B frames.

REFERENCES

- [1] Z. G. Li, F. Pan, K. P. Lim, and S. Rahardja, "Adaptive rate control for H.264," in *Proc. IEEE Intl. Conf. Image Processing*, October 2004, pp. 745–748.
- [2] P. Yin and J. Boyce, "A new rate control scheme for H.264 video coding," in *Proc. IEEE Intl. Conf. Image Processing*, October 2004, pp. 449–452.
- [3] J. Xu and Y. He, "A novel rate control for H.264," in *Proc. IEEE Intl. Symp. Circuits and Systems*, May 2004, pp. III–809–812.
- [4] N. Wang and Y. He, "A new bit rate control strategy for H.264," in *Proc. IEEE Intl. Conf. Infor. Commun. and Signal Processing*, December 2003, pp. 1370–1373.
- [5] S. Kim and Y. S. Ho, "Rate control algorithm for H.264/AVC video coding standard based on rate-quantization model," in *Proc. IEEE Intl. Conf. Multimedia and Expo*, June 2004, pp. 165–168.
- [6] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 7, pp. 688–703, July 2003.
- [7] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 7, pp. 599–603, July 2003.
- [8] Z. He and S. K. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ -domain source modeling," *IEEE Trans. Circuits and Syst. Video Technol.*, vol. 12, pp. 840–849, October 2002.