Analysis of the H.264 advanced video coding standard and an associated rate control scheme

Citation for published version (APA):

Li, P., Lin, W. S., & Yang, X. K. (2008). Analysis of the H.264 advanced video coding standard and an associated rate control scheme. *Journal of Electronic Imaging*, 17(4), 043023-1/15. [043023].
<https://doi.org/doi:10.1117/1.3036181>, <https://doi.org/10.1117/1.3036181>

DOI:

[doi:10.1117/1.3036181](https://doi.org/10.1117/1.3036181)
[10.1117/1.3036181](https://doi.org/10.1117/1.3036181)

Document status and date:

Published: 01/01/2008

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Analysis of the H.264 advanced video coding standard and an associated rate control scheme

Ping Li

Eindhoven University of Technology
Electrical Engineering
Eindhoven, Netherlands 5600 MB
E-mail: p.li@tue.nl

Weisi Lin

Nanyang Technological University
School of Computer Engineering
Singapore

Xiaokang Yang

Shanghai Jiao Tong University
Institute of Image Communication and Information Processing
Department of Electrical Engineering
Minhang District, Shanghai
Shanghai 200240, China

Abstract. An encoded video bitstream is composed of two main components: the coefficient bits representing the discrete cosine transform coefficients, and the header bits representing the header information (e.g., motion vectors, prediction modes, etc.). Compared with previous video standards, the H.264 Advanced Video Coding (AVC) standard has some unique features: (1) the header bits take up a considerable portion of the encoded bitstream; (2) the header bits vary significantly across macroblocks (MBs); and (3) a large number of MBs are quantized to zero and produce zero coefficient bits (zero-coefficient MB). These unique features make most existing rate estimators inaccurate for decision-making processes related to rate-distortion calculation for rate control. This paper analyzes the characteristics of the H.264/AVC bitstream, and reveals that both the header bits and the occurrence of the zero-coefficient MBs are strongly related with motion-compensated residues obtained by $\text{INTER}16 \times 16$. Therefore, two statistical models are proposed for estimating the header bits and separating the zero-coefficient MBs. Based on the proposed models, a rate control scheme is developed for buffer-constrained constant-bit-rate video coding. Experimental results show that the resultant scheme achieves an average of 0.53 dB peak signal-to-noise ratio (PSNR) gain over the original JM6.1e, and less than 2% bit-rate inaccuracy. © 2008 SPIE and IS&T. [DOI: 10.1117/1.3036181]

1 Introduction

The hybrid discrete cosine transform (DCT) based motion-compensated predictive video coding intrinsically produces a variable bit rate. Generally, more bits are allocated to the frame/MB with high activity and fewer bits to the

frame/MB with low activity in order to achieve a consistent video quality. For constant-bit-rate (CBR) transmission, a buffer^{*} is usually used to smooth out the bit-rate fluctuation. Rate control is employed by most video encoders to (1) control the output bit rate to avoid buffer overflow and underflow, and (2) to improve the decoded video quality by appropriately allocating the bits to individual MBs and frames. Rate control in a DCT-based video encoder performs bit allocation by selecting the quantization step size for each MB based on certain source information, including frame/MB complexities, buffer fullness, etc.

The H.264 advanced video coding standard (H.264/AVC),^{2,3} also known as MPEG-4 Part 10, is the latest video coding standard developed by the Joint Video Team (JVT) of the International Standardization Organization (ISO) Moving Picture Experts Group (MPEG) and the International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) Video Coding Experts Group (VCEG). As in other video standards, such as MPEG-2^{4,5} and H.263,⁶ rate control remains an open but important issue for H.264/AVC. A rate control scheme that is able to maximize the video quality and at the same time meet the buffer constraints is much desired for H.264/AVC due to the following three difficulties for rate control in H.264/AVC.

First, compared to previous video standards such as MPEG-2^{4,5} and H.263,⁶ H.264 has a much better choice of prediction modes. This leads to smaller motion-

Paper 08022R received Feb. 7, 2008; revised manuscript received Sep. 4, 2008; accepted for publication Oct. 7, 2008; published online Dec. 11, 2008. This work was done when the authors were with the Institute for Infocomm Research, Singapore.

1017-9909/2008/17(4)/043023/15/\$25.00 © 2008 SPIE and IS&T.

^{*}The selection of the buffer size is dependent on the bandwidth and the maximum delay allowed. A large buffer size tends to ensure a smoother video but causes a longer delay, while a small buffer size guarantees low delay but may be more susceptible to dropping frames due to buffer overflow.¹

Table 1 Percentage of header bits for *news* and *foreman* encoded at different QPs.

QP	Percentage of header bits for <i>news</i>	Percentage of header bits for <i>foreman</i>
24	28.0%	32.4%
32	40.6%	49.6%
40	52.1%	65.4%

compensated residues, and consequently a large number of MBs are quantized to zero. Existing rate estimation models (e.g., the quadratic model⁷) do not consider such zero-coefficient MBs in their estimation. An appropriate modification of the existing models is required for rate-distortion (RD) estimation in H.264.

Second, while DCT coefficients can be represented with many fewer bits than in previous standards, the number of header bits is significantly increased for encoding the increased header information. The header bits consume a large portion of the overall bit budget, especially at low bit rates. Besides, the number of header bits varies greatly among MBs. Consequently, existing rate estimators that usually neglect the header bits or simply represent them by a constant cannot work well. A special investigation into the header bits is necessary for a better RD estimation. Similar problems exist in video standards such as H.263 and MPEG-2, but they are less significant due to the use of relatively simple prediction modes.

Another difficulty of rate control in H.264 is that the rate-distortion optimization (RDO) process is quantization

parameter (QP) dependent. That is, to perform the RDO to determine the best prediction mode for an MB, QP must first be computed by a rate-control scheme. However, to compute the QP by a rate-control scheme, RDO must be conducted to compute the source information (e.g., motion-compensated residues).

1.1 Related Work

Two rate-control approaches have been widely used: the feedback rate control approach, and the rate-and-distortion based approach. In the feedback approach, the coding results of past frames/MBs are used to predict the coding characteristics of future frames/MBs based on the assumption that the temporally close frames or spatially close MBs have similar coding characteristics.^{8,9} As can be expected, the computational complexity of this approach is usually low. However, the past frames/MBs do not always predict well for the future frames/MBs, especially when a scene change occurs. The RD-based approach is usually considered the method that is able to maximize the video quality subject to the given bit budget. RD measurements or estimation of future frames/MBs is fundamental for this approach. To achieve the best RD tradeoff, both the rate and the distortion must be accurately measured or estimated. For example, future frames/MBs may be encoded several times using different QP settings. The QP setting with the minimum rate and distortion can then be selected. The computational complexity of such an approach could become extremely high. A popular approach to reduce the computational complexity relies on RD estimation. With appropriate RD models, rate and distortion of future frames/MBs can be efficiently estimated. The optimization techniques such as Lagrangian optimization¹⁰ and dynamic

Table 2 Occurrences (Occu) of the six prediction modes and their corresponding average header bits (AHB) for *news* and *foreman* encoded at different QPs.

Sequence	QP	Statistics	INT16×16	INT16×8	INT8×16	P8×8	INTRA4×4	INTRA16×16
<i>foreman</i>	40	Occu(%)	63.70	11.37	13.95	2.58	1.31	7.10
		AHB	6.09	15.77	16.53	42.33	9.76	7.80
	32	Occu(%)	44.38	14.20	16.52	18.26	2.79	3.85
		AHB	7.86	18.18	18.09	54.46	9.25	7.87
	24	Occu(%)	25.51	10.55	11.79	46.65	2.82	2.68
		AHB	10.22	19.89	20.54	65.69	7.06	8.39
<i>news</i>	40	Occu(%)	85.32	2.97	4.21	3.42	1.56	2.52
		AHB	1.46	20.19	19.80	56.23	10.92	8.96
	32	Occu(%)	75.85	4.27	5.36	10.96	1.93	1.63
		AHB	1.96	18.26	19.21	66.74	8.52	9.25
	24	Occu(%)	67.30	4.13	3.15	23.09	1.97	0.35
		AHB	2.78	17.00	18.86	72.04	6.35	8.54

Table 3 Percentage of zero-coefficient MBs for *news* and *foreman* encoded at different QPs.

QP	Percentage of zero-coefficient MBs for <i>news</i>	Percentage of zero-coefficient MBs for <i>foreman</i>
24	59.1%	23.0%
32	76.1%	60.6%
40	87.6%	78.8%

programming¹¹ can then be applied to minimize the distortion of the decoded frame/MBs subject to a given bit budget.

The classic RD analysis is performed by analyzing the structure and behavior of each component of the video encoder.^{12–14} This analysis usually provides good theoretical insight into the coding process. However, accuracy may become a problem due to the complexity of the video coding process. Some components, such as quantization, motion compensation, entropy coding, etc., are difficult to analyze. The operational RD model-based approach has proven rather useful for rate control in DCT-based video encoders.^{7,15–20} In Ref. 19, a histogram-based RD model was developed that estimated the rate and distortion based on the number of nonzero DCT coefficients in an MB. As reported in the paper, the RD estimation errors were less than 3%. Using the histogram-based RD model, a rate-control scheme was proposed for low-delay MPEG-2 video coding. Compared with the MPEG-2 TM5, a 0.52 to 1.84 dB peak signal-to-noise ratio (PSNR) improvement was obtained in their simulation. In Ref. 7, a simple quadratic RD model was presented that estimated the rate and distortion of an MB based on the variance of the motion-compensated residues for H.263 low-delay video communications. One advantage of the RD model is that it offered a closed-form solution to the quantization step size after the Lagrangian optimization.

In Refs. 21 and 22, a predictive rate control scheme was proposed for H.264. The general idea is: after pre-encoding of the MB using the QP of a previously encoded MB, the block activity is measured by the sum of absolute difference (SAD). The QP was then determined using a linear model that captures the relation among the QP, the buffer occupancy, and the block activity. The MB was re-encoded if the difference between the two QPs exceeded a specific threshold. In Ref. 23, a rate-control scheme was proposed for bit allocation among frames based on the measurement of frame complexity using the PSNR fluctuation ratio. In Ref. 24, another rate-control scheme was proposed based on the measurement of both the image characteristics and the buffer status. In Ref. 25, a frame-bit allocation scheme was proposed for H.264 via the Cauchy density-based RD models.

Based on our earlier work,²⁶ this paper investigates the characteristics of H.264 RD estimation (especially toward the MB header bits and MBs producing zero-coefficient bits). As a result, we propose two related rate estimators and a frame-layer bit allocation targeted at buffer-

constrained CBR rate control. Based on the observation that both the coefficient bits and the header bits may dominate the output bitstream, we develop the proposed models to estimate the coefficient bits and the header bits separately. The problems of a large header-bits fluctuation and the frequent occurrence of the zero-coefficient MBs are addressed in our work.

The rest of this paper is organized as follows. The characteristics of H.264 are investigated in Sec. 2. The rate and distortion estimators are proposed in Sec. 3, while the frame- and MB-layer bit allocation schemes are presented in Sec. 4. Experimental results are given in Sec. 5. Finally, Sec. 6 concludes this paper.

2 Characteristics of H.264

In H.264, a total of six MB prediction modes—INTER16 × 16, INTER16 × 8, INTER8 × 16, P8 × 8, INTRA4 × 4, and INTRA16 × 16—can be used for an MB in a P frame. In the case of P8 × 8, each of the 8 × 8 blocks can be further partitioned into subblocks of 8 × 4, 4 × 8, or 4 × 4 luminance samples. The MB mode decision is conducted using the RD optimization technique. The best mode is selected by minimizing the following Lagrangian function (RD cost):^{27,28}

$$J_{Mode} = D(Mode|QP) + \lambda_{Mode}R(Mode|QP), \quad (1)$$

where *Mode* is one of the six prediction modes; *QP* is the quantization parameter; $D(Mode|QP)$ is the distortion measured as the sum of the squared difference between the reconstructed and the original MBs; $R(Mode|QP)$ is the rate after entropy coding (both the header bits and coefficient bits); and λ_{Mode} is the Lagrange multiplier, which is computed as

$$\lambda_{Mode} = 0.85 \times 2^{(QP-12)/3}. \quad (2)$$

A large λ_{Mode} tends to choose the prediction mode that produces fewer bits, which is preferred when the buffer level is high. A small λ_{Mode} is preferred for the low buffer levels. λ_{Mode} directly affects the bit rate and distortion of an MB. In Ref. 29, an adaptive Lagrange multiplier adjustment scheme was proposed for the mode decision in H.264.

2.1 Header-Bits Fluctuation

The header bits refer to the bits used to encode the header information, including the prediction mode (MB type), coded block pattern (CBP), QP difference for two consecutive MBs, motion vector difference (MVD), etc.² The header bits fluctuate greatly among MBs that use different prediction modes. For example, if P8 × 8 is used for an MB and if the four 8 × 8 blocks are partitioned into 4 × 4 subblocks, then there will be 16 motion vectors for a single MB. In contrast, if INTER16 × 16 is used, there is only one motion vector. Similar problems also exist in MPEG-2, H.263, etc., but they are less significant due to the use of relatively simple prediction modes. Table 1 shows the bit statistics for the QCIF sequences *news* and *foreman* encoded at different QPs by the reference software.³⁰ As shown in the table, header bits can be comparable to (or even more than) the coefficient bits; and header bits vary greatly with video contents and QPs (and therefore bit rates).

Fig. 1 $C-H$ curves for four sequences encoded at different bit rates: (a) *news* at 27 Kbps, (b) *foreman* at 122 Kbps, (c) *silent* at 12 Kbps, and (d) *paris* at 64 Kbps.

Table 2 shows detailed statistics of the six MB prediction modes when *news* and *foreman* are encoded at different bit rates by the reference software.³⁰ From the table, we observe:

1. The header bits vary greatly with prediction modes. $INTER16 \times 16$ is the most header-bits economical, and $P8 \times 8$ is the most header-bits expensive.
2. $INTER16 \times 16$ is used more often for low bit rates while $P8 \times 8$ is used more often for high bit rates. This can be easily observed if we compare the occurrences of $INTER16 \times 16$ and $P8 \times 8$ at different bit rates for the same video sequence.
3. $INTER16 \times 16$ is used more often for slow-motion video and $P8 \times 8$ is used more often for fast-motion video. This can be easily observed if we compare the occurrences of $INTER16 \times 16$ and $P8 \times 8$ for different sequences at the same QP.

The above observations will be explained below.

The header bits are comprised of the bits for the MB type, 8×8 subpartition type, reference frame, MVD, DQuant, CBP, etc. The bits used for the MB type, 8×8 subpartition type, and MVD depend heavily on the prediction mode. For example, due to the multiple motion vectors for $P8 \times 8$ (up to 16 motion vectors) instead of only one for $INTER16 \times 16$, the header bits for $P8 \times 8$ are much larger than for $INTER16 \times 16$.

Because of the use of multiple motion vectors, $P8 \times 8$ has a better capability than $INTER16 \times 16$ to reduce the prediction residues at the cost of the bits for the header, and $INTER16 \times 16$ has a better capability to reduce the bit rate at the cost of the distortion (residues). The RDO attempts to select a prediction mode that achieves the best tradeoff. For slow-motion videos, where MBs can be easily matched to the corresponding MBs in the reference frame, $INTER16 \times 16$ is the most suitable since both the distortion, which largely depends on the prediction residues, and the rate, which largely depends on the prediction mode, are small. In

contrast, for fast-motion videos $P8 \times 8$ has the advantage since it can significantly reduce the residues, and thus the distortion.

The reason why $INTER16 \times 16$ is used more often for low bit rates and $P8 \times 8$ is used more often for high bit rates is the same as explained above. Due to the large λ_{Mode} in Eq. (1) at low bit rates, the header bits are very expensive for computing the RD cost. Thus, $INTER16 \times 16$ gains the advantage over $P8 \times 8$ because it produces many fewer header bits than does $P8 \times 8$.

2.2 Zero-Coefficient MBs

Compared to other video coding standards, H.264 has a much better choice of prediction modes for the MBs. This usually leads to smaller motion-compensated residues (MCRs). As a result, the DCT coefficients of many MBs are quantized to zero in H.264, especially when the video is encoded at low bit rates. Table 3 shows the percentages of the zero-coefficient MBs for *news* and *foreman* encoded at different QPs by the reference software.³⁰ The table shows that the percentage is very high for most cases. Explicitly considering these zero-coefficient MBs is important for rate

Fig. 2 Four neighboring MBs for predicting the DCT rate of the current MB (CMB), which are denoted as ULMB, UMB, URMb, and LMB respectively.


```

if ( $\sigma_{CMB}^2 < \sigma_Z^2$ )  $ZCOF = 1$ ;
switch (num)
{
  case 4: if ( $\sigma_{CMB}^2 < \min(2.6\sigma_1^2, 1.9\sigma_2^2)$ )  $ZCOF = 1$ ;
  case 3: if ( $\sigma_{CMB}^2 < \min(2.3\sigma_1^2, 1.7\sigma_2^2)$ )  $ZCOF = 1$ ;
  case 2: if ( $\sigma_{CMB}^2 < \min(1.9\sigma_1^2, 1.4\sigma_2^2)$ )  $ZCOF = 1$ ;
  case 1: if ( $\sigma_{CMB}^2 < \min(1.6\sigma_1^2, 1.2\sigma_2^2)$ )  $ZCOF = 1$ ;
  case 0: if ( $\sigma_{CMB}^2 < \min(\sigma_1^2, \sigma_2^2)$ )  $ZCOF = 1$ ;
}

```

Fig. 3 Pseudo codes of the proposed algorithm for separating zero-coefficient MBs.

control in H.264, since it is obvious that the existing rate estimators are designed for only the MBs with nonzero coefficient bits.

3 Rate and Distortion Estimation

Most existing estimators consider only the coefficient bits. The header bits are usually neglected or simply represented by a constant. This is not a problem for MPEG-2, H.263, etc., because the header bits are relatively small in number due to the simplicity of prediction modes in these standards. However, as discussed in Sec. 2, the header bits constitute a significant portion of the H.264 bitstream. Separate estimation of the header bits is necessary. Furthermore, as discussed in Sec. 2, a large portion of the MBs are quantized to zero in H.264. Separation of such zero-coefficient MBs also helps to improve rate control.

3.1 Premeasurement

To estimate the coefficient bits and the header bits, all MBs are preanalyzed using $INTER16 \times 16$. We refer to this step as *premeasurement* or *preanalysis*. In this step, the MCR associated with $INTER16 \times 16$ is obtained for each MB, which is thereafter used for RD estimation.

3.2 Header-Bits Estimation

Based on extensive experiments, we have observed that the number of header bits H_i for an MB is approximately linear to the normalized variance of the MCR obtained in preanalysis. This linear relationship can be represented by the following equation:

$$H_i = C \times [\log(\sigma_i^2)]^2 \quad \text{when } H_i > 10, \quad (3)$$

where σ_i^2 is the variance of the MCR obtained by $INTER16 \times 16$ in preanalysis and C is a constant factor representing the linear relationship between the header bits and the normalized variance. Figure 1(a) shows a C - H curve for *news*, in which C is computed as $H_i / [\log(\sigma_i^2)]^2$. From the figure, we see that C varies little along the H axis. That means C keeps “constant” for most MBs. Thus, the linear relationship between the header bits and the normalized variance captured by Eq. (3) is validated. Figures 1(b)–1(d) depict the C - H curves for several other sequences, and similar observations can be made.

3.2.1 Rationale of header-bits model

As previously discussed, $INTER16 \times 16$ is used to compute the MCR for each MB. A small MCR means a good pre-

Fig. 4 Program flow chart of the proposed rate-control scheme.

diction by $INTER16 \times 16$, and therefore, with a high probability, big-block prediction modes like $INTER16 \times 16$, $INTER16 \times 8$, or $INTER8 \times 16$ will be selected in the RDO process. A large MCR means poor prediction by $INTER16 \times 16$, and thus header bits-expensive modes such as $INTRA4 \times 4$ and $P8 \times 8$ are more likely to be selected. Therefore, the prediction modes largely depend on the MCR by $INTER16 \times 16$. The larger the variance of the MCR, the higher the probability that the header bits-expensive modes such as $P8 \times 8$ will be used. In other words, the header bits increase with the variance of the MCR, as suggested by Eq. (3).

The cases with $H_i < 11$ correspond mostly to MBs where $INTER16 \times 16$ is selected in the RDO process. In the proposed scheme, these small-header-bits MBs are considered as follows: (1) during the encoding of the previous frame, record σ^2 and H for all MBs whose H is smaller than 11; (2) compute the average σ^2 and H , which are termed σ_{trd}^2 and H_{trd} , respectively; (3) during the encoding of the current frame, once $\sigma^2 \leq \sigma_{trd}^2$ is found for an MB, the MB is expected to produce small header bits, and H is directly estimated by H_{trd} . By combining with Eq. (3), the proposed header-bits model can be expressed as.

$$H_i = \begin{cases} H_{trd}, & \sigma_i^2 \leq \sigma_{trd}^2 \\ C \times [\log(\sigma_i^2)]^2, & \text{else.} \end{cases} \quad (4)$$

With the introduction of an intermediate variable v_i defined as

$$v_i = \begin{cases} H_{trd}/C, & \sigma_i^2 \leq \sigma_{trd}^2 \\ [\log(\sigma_i^2)]^2, & \text{else.} \end{cases} \quad (5)$$

Eq. (4) can be rewritten as

$$H_i = C \times v_i. \quad (6)$$

Both C and v_i are adaptively updated during the encoding process, as will be described in Secs. 4.3.1–4.3.3.

Fig. 5 Relation between λ_1 , λ_2 , and the normalized buffer fullness L/M .

3.3 Separation of Zero-Coefficient MBs

The quadratic model proposed in Ref. 7 is used to estimate the coefficient bits. Let F_i denote the bits required for encoding the DCT coefficients of the i 'th MB, and Q_i denote the quantization step size; F_i can be estimated by the following formula:

$$F_i = AK \frac{\sigma_i^2}{Q_i^2}, \quad (7)$$

where A is the number of the pixels in an MB and K could be set to $(e/\ln 2)$. F_i is also referred to as the DCT rate hereinafter. As discussed in Ref. 7, Eq. (7) is derived with the assumption that DCT coefficients are approximately uncorrelated and Laplacian distributed with variance σ_i^2 in H.263. Under this assumption, the DCT rate of an MB can be approximated by the empirical entropy of Q-quantized DCT coefficients. In H.264, the DCT transform and the Q-quantization process are not different from those in H.263. Thus, the same assumption can be made about the DCT coefficients, and Eq. (7) therefore applies to H.264 for DCT rate estimation. The difference with H.264 is that, as was pointed out in Sec. 1, the DCT coefficients of many MBs are quantized to zero due to a smaller residual energy. For these zero-coefficient MBs, Eq. (7) does not apply, so we must separate them before applying Eq. (7).

In the proposed algorithm, zero-coefficient MBs (indicated by ZCOF) are predicted using statistics of the coding results of up to four neighboring MBs, as shown in Fig. 2. The more zero-coefficient MBs in its neighborhood, the higher the probability that the current MB will produce zero coefficients as well. The separation steps are as follows: first, count the number of neighboring zero-coefficient MBs (num); second, compare the variance σ_{CMB}^2 of the current MB with its neighbors' variances. Once we find that σ_{CMB}^2 is smaller than its neighbor's variances by a threshold, we set ZCOF to 1. The pseudo codes of the proposed separation algorithm are shown in Fig. 3, where $\sigma_1^2 = \min(\sigma_{ULMB}^2, \sigma_{URMB}^2)$, $\sigma_2^2 = \min(\sigma_{ULMB}^2, \sigma_{LMB}^2)$, and σ_Z^2 is the average variance of all zero-coefficient MBs in the previous frame.

As shown in Fig. 4, once an MB is predicted to produce a zero DCT rate, we do not compute the QP at all. Instead,

we use the QP of the previously encoded MB. The use of the above rate model at very low bit rates is therefore avoided. As will be discussed in Sec. 3.4, separation of these zero-coefficient MBs is also necessary to prevent the use of the distortion model [Eq. (8)] at low bit rates, where the distortion model does not apply.

3.4 Distortion Estimation

Uniform quantization of the DCT coefficients inevitably introduces distortion for the encoded MBs. In our design, the following typical distortion model⁷ is used to measure the distortion of the encoded MBs:

$$D = \frac{1}{N} \sum_{i=1}^N \alpha_i^2 \frac{Q_i^2}{12}, \quad (8)$$

where N is the number of MBs in a frame and α_i is the distortion weight of the i 'th MB. As explained in Ref. 7, at very low bit rates, there are often MBs whose quantizer step size Q_i is three or four times larger than their respective σ_i . Consequently, the approximation of the mean square error (MSE) by $Q_i^2/12$ is not as effective. The use of the distortion model at very low bit rates should be corrected or avoided. In our proposal, besides separating zero-coefficient MBs to avoid the use of the above model at very low bit rates, the distortion weight α_i is employed to tilt the above distortion model and reduce the quantization overhead at low bit rates. α_i is set using the following formula:

$$\alpha_i = \begin{cases} \sigma_i^{3/4}, & \frac{B}{AN} \leq 0.05 \\ \sigma_i^{1/2}, & \frac{B}{AN} \leq 0.2 \\ \sigma_i^{1/4}, & \frac{B}{AN} \leq 0.5 \\ 1.0, & \text{elsewhere} \end{cases}, \quad (9)$$

where B is the bit budget for the frame to be encoded. In H.264, the QPs of the MBs are differentially encoded. Though the overhead of frequent QP change is negligible at high bit rates, it may seriously affect the video quality at

Fig. 6 (a) PSNR, (b) average QP, (c) frame bits, and (d) buffer level for *news* encoded at 27 Kbps.

low bit rates.⁷ Using Eq. (9), by setting α_i according to the bit rate (B/AN), frequent QP change is prevented at low bit rates. For example, if $\alpha_i = \sigma_i$, the QPs computed by Eq. (12) will be equal for all MBs. When $\alpha_i \approx \sigma_i$ at very low bit rates, the QPs by Eq. (12) will remain close to each other, and thus frequent QP change is avoided. When the bit rate is high (above 0.5 bpp), α_i is set to 1. The distortion of an MB can then be well approximated using $Q_i^2/12$.

Both the DCT rate model [Eq. (7)] and the distortion model [Eq. (8)] are approximations of video signal proper-

Fig. 7 (a) PSNR, (b) average QP, (c) frame bits, and (d) buffer level for *foreman* encoded at 66 Kbps.

ties under certain assumptions. Though such approximations work in general for a video, we cannot expect them to work for every MB. The accuracy of RD models thus relies heavily on the adaptation of their weighting factors to the local video context (see Sec. 4.3). This is reasonable. As pointed out in Sec. 1, due to the varieties of video sources and the complexity of the video coding process, it is difficult to design an analytical model that works for all scenarios. We believe the statistical approach that learns from the actual video signal is more realistic and effective. Like

Table 4 Test sequences.

Test sequence	Size	Frame rate	QP range	Frames encoded	GOP structure
news	QCIF	10	28–40	100	IPPP
container	QCIF	10	28–40	100	IPPP
silent	QCIF	15	28–40	150	IPPP
foreman	QCIF	15	28–40	150	IPPP
paris	CIF	15	28–40	150	IPPP
mobile	CIF	30	28–40	300	IPPP
tempe	CIF	30	28–40	240	IPPP

the large number of statistical context tables in H.264/AVC used for entropy coding with CABAC and CVLC, the proposed models use such a statistical approach. Instead of attempting to describe the video signal analytically, the statistics of past and neighboring frames/MBs are used to predict characteristics of future frames/MBs. The experimental results presented in Sec. 5 validate the effectiveness of this approach.

4 Rate-Control Scheme

Figure 4 depicts the program flow of the proposed rate-control scheme, which is comprised of the following three steps:

1. *Premeasurement* to compute the source information for subsequent RD estimation;
2. *Frame-layer bit allocation* to determine the bit target for each frame;
3. *MB-layer rate control* to compute the QP for each MB.

4.1 Premeasurement

As discussed in Sec. 3.1, we compute the MCRs in this step using $\text{INTER}16 \times 16$ for RD estimation. Besides the residues, the RD cost for each MB is also obtained, which is used for frame-layer bit allocation described in Sec. 4.2. The computational complexity of the premeasurement using $\text{INTER}16 \times 16$ is quite high. However, since the results obtained in premeasurement can be stored for later use in the RDO process, $\text{INTER}16 \times 16$ does not have to be executed again in the RDO process. Thus, premeasurement does not actually increase the overall computational complexity.

4.2 Frame-Layer Control

The proposed frame-layer bit allocation scheme can be divided into two steps. First, determine the frame budget for the best video quality without considering the buffer constraints:

$$B_1 = [1 + (\hat{P} - P_n)/2] \times \frac{\mathbf{J}_{cur} - \mathbf{J}_{prev,0}}{\hat{\mathbf{J}} - \mathbf{J}_{prev,0}} \times \frac{R}{f}, \quad (10)$$

where R is the available channel bandwidth; f is the frame rate; $\mathbf{J}_{cur}$ is the RD cost of the current frame, and it is computed as the sum of the RD costs of all MBs in the current frame; $\hat{\mathbf{J}}$ is the average RD cost of all encoded frames (including current frame); $\mathbf{J}_{prev,0}$ is the sum of the RD cost of all zero-coefficient MBs in the previous frame (all RD costs are obtained in premeasurement using $\text{INTER}16 \times 16$); P_n is the average PSNR of the previous n frames computed in a sliding window; and $\hat{P}$ is the average PSNR of all encoded frames. Using Eq. (10), more bits will be allocated to the frames whose activity (measured by $\mathbf{J}_{cur}$) is high and the predicted PSNR (measured by P_n) is low, and vice versa.

In the second step, the frame budget is adjusted according to the buffer state. If the buffer level is predicted to increase and the current buffer level is above a threshold B_{inc} , or if the buffer level is predicted to decrease and the current buffer level is below a threshold B_{dec} , B_1 is appropriately adjusted using the following algorithm:

```

 $B_2 = \kappa \times B_1;$  //predict the bits for current frame
if ( $B_2 > R/f \& L > 0.2M$ ) //if buffer level predicted to
    increase above 0.2M
{
     $B_2 = R/f + \lambda_1(B_2 - R/f);$  //restrict buffer level increase
     $B_2 = \min(B_2, M - L - 0.5 \times R/f);$ 
}
if ( $B_2 < R/f \& L < 0.2M$ ) //if buffer level predicted to
    decrease below 0.2M
{
     $B_2 = R/f + \lambda_2(B_2 - R/f);$  //restrict buffer level decrease
     $B_2 = \max(B_2, R/f - L);$ 
}
 $B_2 = [0.3R/f, 2.5R/f];$  //clip the bit budget
 $B = B_2 / \kappa;$ 

```

where M is the maximum allowable buffer level; L is the current buffer fullness; B_2 is the number of bits predicted to be generated after encoding the current frame; and κ is the ratio between the bits actually used by a frame and the predicted frame budget, which is updated after encoding each frame. The actual bits used by a frame can be different from the bits that are initially allocated. We will give the explanation in Sec. 4.3.2

Frame-layer bit allocation is based on the following two principles: (1) the bit rate should be allowed to fluctuate with the varying frame complexity for good video quality, provided that the buffer is safe from overflow and underflow; (2) the buffer level increase should be more strictly limited than the buffer level decrease to allow subsequent high-complexity frames to use more bits to maintain a consistent video quality.

As shown by the pseudo codes, if the observed buffer level is above 20% of the maximum buffer size, an appropriate restriction on the buffer level increase is imposed. If the observed buffer level is below 20% of the maximum buffer size, an appropriate restriction on the buffer level decrease is imposed. The extent of restriction depends on λ_1 and λ_2 . The relation among λ_1 , λ_2 , and the normalized buffer fullness (L/M) is shown in Fig. 5. If $L/M > 0.2$, any possible buffer level increase ($B_2 - R/f$) is multiplied by a factor that is smaller than 1. The higher the buffer level L/M is, the smaller λ_1 is and thus the stronger the restriction imposed on the buffer level increase.

4.3 MB-Layer Control

Given the RD models described in Sec. 3, the Lagrangian optimization is used to compute the optimal quantization step sizes $Q_1^*, Q_2^*, \dots, Q_N^*$ for each MB by minimizing the following cost:

$$\text{cost} = D + \lambda \left[\sum_{i=1}^N (F_i + H_i) - B \right] = \frac{1}{N} \sum_{i=1}^N \alpha_i^2 \frac{Q_i^2}{12} + \lambda \left[\sum_{i=1}^N \left(AK \frac{\sigma_i^2}{Q_i^2} + C \times v_i \right) - B \right], \quad (11)$$

where B is the frame budget, and λ is the Lagrange multiplier. By setting partial derivatives of $Q_1, Q_2, \dots, Q_N$, and λ to zero, we have $N+1$ equations with $N+1$ independent variables. By solving the $N+1$ equations, we obtain:

$$Q_i^* = \sqrt{\frac{AK}{B - C \sum_{j=1}^N v_j} \frac{\sigma_i}{\alpha_i} \sum_{i=1}^N \alpha_i \sigma_i}. \quad (12)$$

Given the computed quantization parameters, λ_{Mode} in Eq. (2) can be computed and the RDO can be performed to select the prediction mode for the MB. More details about the above Lagrangian optimization process can be found in Ref. 7. Another method for adaptive adjustment of λ_{Mode} for rate control in H.264 can be found in Ref. 29.

As stated in Sec. 3.3, an algorithm is developed to separate zero-coefficient MBs before applying the quadratic model for rate estimation. As shown in Fig. 4, once an MB is predicted to produce zero coefficient bits, the QP of the

previously encoded MB is used. Otherwise, Eq. (12) is used to compute the QP. Let $S_i = \sum_{j=i}^N \alpha_j \sigma_j$, $T_i = \sum_{j=i}^N v_j$, the Q_i^* for the i 'th MB is computed by:

$$Q_i^* = \sqrt{\frac{AK_i}{B_i - C_i T_i} \frac{\sigma_i}{\alpha_i} S_i}, \quad (13)$$

where B_i is the bit target for the remaining MBs from i to N in the frame; K_i , C_i are the updated values of K , C after encoding the first $(i-1)$ MBs; and $C_i T_i$ is the number of header bits required for the remaining MBs. Obviously, $S_{i+1} = S_i - \alpha_i \sigma_i$ and $T_{i+1} = T_i - v_i$.

4.3.1 Updating B_i

B_{i+1} is updated using the following formula:

$$B_{i+1} = \left(B - \sum_{j=1}^i b_j \right) \times \frac{N-i}{N} + \left(\frac{\sum_{j=i+1}^N J_i}{\sum_{j=1}^i J_i} \times \sum_{j=1}^i b_j \right) \times \frac{i}{N}, \quad (14)$$

where J_i is the RD cost of the j 'th MB; and b_j is the actual number of bits used for the j 'th MB. The first term in the right side of the equation allocates the bits of the remaining MBs based on the initially determined bit budget for the frame. The second term allocates the bits according to the actual RD cost of the MBs after the first i MBs have been encoded, which adapts the bit allocation to the local frame contents. The more MBs encoded, the stronger the adaptation. This is suggested by (i/N) in Eq. (14).

Equation (14) implies that the actual bits used by a frame can be different from the bits that are initially allocated by the frame-layer rate control. This is reasonable. Since more accurate local source information is available, the coding parameters should be adapted during the encoding. This also partly explains the inconsistency between the actual bits and the allocated bits, as shown in Figs. 6(c) and Figs. 7(c).

4.3.2 Updating K_i

K_i is updated using the following steps:

1. Compute K'_i after encoding the current MB:

Table 5 Test conditions.

MV resolution	1/4 pel
Hadamard	ON
Search range	32
Restricted search range	2
Reference frames	1
Symbol mode	CABAC
Slice mode	OFF

Table 6 Results when QP=40 (the QP for the first 'I' frame for the proposed scheme is 36).

Sequence	Scheme	PSNR (dB)	R (bps)	Rp (bps)	UFLW	OFLW	AR	GAIN (dB)	ΔR (%)
news	JM	28.07	10,451						
	ours	29.04	10,296	9,034	3	0	87.7	0.97	-1.48
container	JM	27.94	5,158						
	ours	28.98	5,113	4,029	5	0	93.74	1.04	-0.87
silent	JM	27.97	12,104						
	ours	28.9	12,001	11,019	3	0	89.86	0.93	-0.85
foreman	JM	28.63	22,184						
	ours	28.81	22,257	21,336	3	0	75.47	0.18	0.33
paris	JM	26.78	59,633						
	ours	27.59	58,749	52,812	8	0	88.94	0.81	-1.48
tempeste	JM	26.24	166,109						
	ours	26.36	164,054	156,252	15	0	83.04	0.12	-1.24
mobile	JM	24.51	223,240						
	ours	24.47	221,764	211,978	5	0	80.19	-0.04	-0.66

$$K'_i = \frac{F_i \times (Q_i^*)^2}{256\sigma_i^2}.$$

2. If $K'_i > 0$ and $K'_i \leq 4.5$, compute the average K of the MBs encoded so far:

$$\bar{K}_i = \bar{K}_{i-1}(l-1)/l + K'_i/l,$$

where l is the number of MBs encoded so far whose K'_i are within $[0, 4.5]$.

3. Find the weighted average of the initial estimate K_1 with $\bar{K}_i$:

$$\tilde{K}_i = \bar{K}_i(i/N) + K_1(N-i)/N.$$

The above three steps to update K are reported in Ref. 7. In our implementation, K is further updated according to the MB's local scene activity.

4. Adjust $\tilde{K}_i$ according to the average K of its neighboring MBs, which is denoted as K_{local} :

$$\check{K}_i = 0.9 \times \tilde{K}_i + 0.1 \times K_{local}.$$

5. If there is no scene change between the current and previous frames, adjust $\check{K}_i$ according to the K of its co-located MB in the previous frame, which is denoted as K_{prev} :

$$K_{i+1} = 0.9 \times \check{K}_i + 0.1 \times K_{prev}.$$

We deem that there is a scene change if the ratio of the RD cost of two consecutive frames $\mathbf{J}_{curr}/\mathbf{J}_{prev}$ is greater than 1.2.

4.3.3 Updating C_i

C_i is updated using the following steps:

1. Compute C'_i based on the available information after encoding the current MB:

$$C'_i = \frac{\sum_{j=1}^i (b_j - F_j)}{\sum_{j=1}^i v_j},$$

where $\sum_{j=1}^i (b_j - F_j)$ is the total number of header bits used to encode the first i MBs.

2. Find the average C' of all the encoded MBs in the current frame:

$$C''_i = C'_i \times (i-1)/i + C'_1 \times 1/i.$$

3. Find the weighted average of the initial estimate C_1 with C''_i :

$$C_{i+1} = C''_i \times i/N + C_1 \times (N-i)/N.$$

Table 7 Results when QP=36 (the QP for the first 'I' frame for the proposed scheme is 32).

Sequence	Scheme	PSNR (dB)	R (bps)	R_p (bps)	UFLW	OFLW	AR	GAIN (dB)	ΔR (%)
news	JM	30.64	16,748						
	ours	31.6	16,471	14,655	0	0	83.05	0.96	-1.65
container	JM	30.41	8,116						
	ours	31.3	8,099	6,471	3	0	90.82	0.89	-0.21
silent	JM	30.25	20,156						
	ours	31.15	19,886	18,322	2	0	85.07	0.9	-1.34
foreman	JM	30.85	33,816						
	ours	30.95	33,943	32,523	2	0	66	0.1	0.38
paris	JM	29.41	102,423						
	ours	30.37	100,512	91,798	11	0	84.01	0.96	-1.87
tempeste	JM	28.61	301,769						
	ours	28.54	298,870	287,055	6	0	73.53	-0.07	-0.96
mobile	JM	27.12	405,258						
	ours	27.12	402,631	388,562	3	0	70.63	0	-0.65

As in Eq. (4), C is a constant that describes the linear relationship between the header bits of an MB and the normalized variance v_i .

As we note from Eq. (13), to compute Q_i^* for the current MB, we need to estimate the total number of header bits ($C_i T_i$) required for all the remaining MBs in the current frame. Because of this requirement, our statistical header-bits model becomes useful, though it may not work accurately for individual MBs. Compared with other rate models where the header bits are simply represented by a constant, our model can better reflect the relationship between the header bits and the MB characteristics i.e., the header bits of an MB increase with the variance of the residues in an approximately linear way.

5 Experimental Results

The proposed rate estimators can serve as enhancement modules to be integrated into an existing rate controller. In this section, we implement the proposed rate control scheme in an H.264 reference software (JM).³⁰ In our experiments, only the features enabled in the main profile of H.264 are used. Table 4 lists the test sequences and Table 5 lists the test conditions for our experiments.

Seven sequences were tested in our experiments, each of which was encoded at 4 different bit rates with a QP ranging from 28 to 40. The performance of the proposed algorithm was evaluated against the fixed-QP scheme of the original JM software in terms of both the PSNR improvement and the bit rate. The fixed-QP scheme was first run at

different QPs. The resulting bit rates were thereafter used as target bit rates for the proposed scheme. The differences between the two bit rates and PSNRs were then computed.

Tables 6–9 show the experimental results. In these tables, R is the average bit rate; R_p is the average bit rate of P frames; AR denotes the accuracy of the source information, which will be defined in Sec. 5.1; $OFLW$ denotes the number of buffer overflows that occurred; $UFLW$ denotes the number of buffer underflows that occurred; $GAIN$ denotes the PSNR improvement achieved by the proposed scheme over the JM; ΔR denotes the bit rate inaccuracy, which is computed as $(R_{ours} - R_{JM})/R_{JM} \times 100\%$, where R_{ours} is the bit rate achieved by the proposed scheme and R_{JM} is the target bit rate determined by the JM software.

5.1 Performance in Terms of PSNR

As shown in Tables 6–9, the proposed scheme is able to significantly improve the PSNR for most bit rates and sequences. For the seven sequences tested in this work, the proposed scheme achieved an average of 0.53 dB PSNR improvement over the original JM software. The proposed scheme performed better for low-motion sequences. An average 0.89 dB PSNR improvement was achieved for the low-motion sequences *news*, *container*, *silent*, and *paris*. We also observe that the proposed scheme performed better for low bit rates. As shown in Tables 6–9, the PSNR improvement decreased gradually as the bit rate increased for the same video sequence.

As discussed in Sec. 4, the residue obtained in premeasurement by $INTER16 \times 16$ is used for DCT rate and dis-

Table 8 Results when QP=32 (the QP for the first 'I' frame for the proposed scheme is 28).

Sequence	Scheme	PSNR (dB)	R (bps)	R_p (bps)	UFLW	OFLW	AR	GAIN (dB)	ΔR (%)
news	JM	33.5	27,586						
	ours	34.51	27,158	24,634	0	0	77.78	1.01	-1.55
container	JM	33.14	13,841						
	ours	33.86	13,867	11,547	1	0	86.09	0.72	0.19
silent	JM	32.77	34,558						
	ours	33.76	34,271	31,858	2	0	78.38	0.99	-0.83
foreman	JM	33.17	55,544						
	ours	33.28	55,758	53,690	0	0	53.68	0.11	0.39
paris	JM	32.34	182,965						
	ours	33.39	180,298	168,045	6	0	78.88	1.05	-1.46
tempeste	JM	31.38	616,914						
	ours	31.39	610,268	593,815	13	0	60.57	0.01	-1.08
mobile	JM	30.22	850,171						
	ours	30.41	845,756	827,288	3	0	59.08	0.19	-0.52

tortion estimation. However, both the quadratic model [Eq. (7)] and the distortion model [Eq. (8)] are based on the actual residue. If the RDO process selects a mode other than $\text{INTER}16 \times 16$ for motion estimation, the source information for RD estimations will be inaccurate. We use AR to denote the *accuracy of source information*, which is computed as follows:

$$AR = N_{16 \times 16} / N_{\text{total}} \times 100\%, \quad (15)$$

where $N_{16 \times 16}$ denotes the occurrence of $\text{INTER}16 \times 16$, and N_{total} denotes the occurrence of all prediction modes. As shown in Tables 6–9, AR is higher for low-motion sequences and low bit rates because, as explained in Sec. 2, there are more chances for $\text{INTER}16 \times 16$ to be selected by RDO in such situations.

One other important reason why our scheme performs better for the low-motion video sequences and low bit rates is that, in these cases, more MBs produce zero or very low bit rates. The proposed zero-coefficient-MB-separation algorithm (Fig. 3) and the adaptation of the distortion weight [Eq. (9)], which are specially designed for handling such low bit rates, make a positive effect on improving the video quality.

5.2 Performance in Terms of Bit Rate

A large buffer allows large bit rate fluctuation and is helpful for achieving a constant video quality. However, it is not desirable for streaming the video over CBR channels, because a higher bandwidth and a larger decoder buffer are

required in this case.^{31,32} Appropriate tradeoff between the bit rate variation and the quality fluctuation has to be made.

The “leaky bucket” model^{31,32} is used in this work to control the output bit rate. The leaky bucket model can be characterized by (R, B, Γ) , where B is the encoder buffer size, R is the transmission bit rate, Γ means that the transmission of the bits in the encoder buffer starts Γ seconds after the bits for the first frame enter the buffer. Our buffer control assumes that the bits for the first I frame are, in some way, transmitted to the terminal without pushing them into the encoder buffer (this is the typical assumption for buffer control for low-delay IPPP video). Thus, when the target bit rate is computed for the P frames, the bits for the first I frame must be deducted from the overall bit budget. Suppose we want to encode a video sequence of N frames and the first I frame consumes b_I bits. The target bit rate for the following $(N-1)$ number of P frames is computed as $R_p = (N \times R - b_I \times f) / (N-1)$. In our experiments, the encoder buffer size was selected as 5 times the average P frame size, i.e., $M = 5R_p / f$, which means that the maximum buffer delay was 500 ms for 10-fps video and 167 ms for 30-fps video. The startup encoder buffer delay was set to 2.2 times the frame interval, i.e., $\Gamma = 2.2/f$. Upon the completion of encoding a frame, all bits are pushed into the buffer instantaneously at every frame interval. The bits in the buffer are drained to the channel at a constant bit rate of R_p unless the buffer is empty.

One important criterion to evaluate a rate-control scheme is the occurrence of buffer overflows and under-

Table 9 Results when QP=28 (the QP for the first 'I' frame for the proposed scheme is 24).

Sequence	Scheme	PSNR (dB)	R (bps)	R_p (bps)	UFLW	OFLW	AR	GAIN (dB)	ΔR (%)
news	JM	36.66	44,931						
	ours	37.44	44,248	40,885	0	0	72.49	0.78	-1.52
container	JM	35.94	25,604						
	ours	36.32	25,374	22,144	1	0	80.64	0.38	-0.9
silent	JM	35.68	58,564						
	ours	36.62	57,748	54,194	1	0	70.93	0.94	-1.39
foreman	JM	35.79	97,210						
	ours	35.85	97,587	94,713	0	0	41.71	0.06	0.39
paris	JM	35.5	314,801						
	ours	36.45	309,824	293,442	6	0	74.34	0.95	-1.58
tempeste	JM	34.64	1,260,891						
	ours	34.59	1,245,455	1,224,351	21	0	46.87	-0.05	-1.22
mobile	JM	33.74	1,689,307						
	ours	33.67	1,678,285	1,655,531	8	0	47.97	-0.07	-0.65

flows. Buffer overflow results in discarding the entire encoded frame (frame skipping) and seriously affects the video quality, and therefore it should be avoided as much as possible. However, occasional buffer underflow is allowed since it has a small impact on the video quality, though it will waste some channel bandwidth. Tables 6–9 also show the occurrences of the buffer overflows (OFLW) and buffer underflows (UFLW) observed in our experiments. As shown, no single buffer overflow and only a few buffer underflows occurred in our experiments. The maximum bit-rate inaccuracy ΔR was less than 2% given an output buffer that was equal to 5 times the bandwidth (R_p).

As in Figs. 6(c) and 7(c), the actual bits used by a frame are not equal to the bits allocated to it by the frame-layer rate control, and in some cases, the difference is significant. As explained in Secs. 4.2 and 4.3, the reasons are (1) the MB-by-MB adjustment of the bit budget based on local video context will result in such inconsistency; and (2) scaling of κ at a very high or low buffer level also results in such inconsistency. However, such inconsistency between the used and allocated bits is not an issue as long as it does not cause buffer underflows/overflows or impair the video quality. The more important is that the allocated bits should reflect the real *needs* for bits by a frame, and at the same time keep the buffer safe.

Figures 6 and 7 show the PSNR, average QP, the bits used for each frame, and the buffer level at every frame interval for *news* and *foreman*. The figures show that the bits used for each frame, the PSNR, and the average QP fluctuate with both the frame complexity and the buffer

Fig. 8 QP standard deviation (a) when *news* is encoded at 27 Kbps, and (b) when *foreman* is encoded at 66 Kbps.

fullness. The PSNR by the proposed scheme is better than that by the fixed-QP scheme for most frames, and the buffer level is kept within the given range. Figure 8 shows the QP standard deviation of each frame, where we see, within a frame, the QPs of most MBs fluctuate around its average within the range of -2 to 2 . The observation reflects the nature of video signals, and is desirable in video coding. Since neighboring MBs usually have similar characteristics, we do not expect that QPs will vary significantly across MBs if the proposed RD models are able to estimate the rates correctly.

6 Conclusion

In comparison with existing video coding standards, H.264 has some unique features and their impact on rate control should be addressed. In this paper, we have given a comprehensive analysis of the H.264 encoding process, and then proposed corresponding RD models for H.264 rate control that target buffer-constrained CBR video coding. The contributions of this work include: (1) an algorithm to separate the zero-coefficient MBs to avoid the use of the quadratic rate model and the distortion model at very low bit rates; and (2) a statistical header-bits model to separately estimate the header bits from the residues obtained in the preanalysis. Since header bits take a large portion of the encoded bitstream in H.264, a separate header-bits estimation is necessary. The adaptation of various coding parameters according to the local video context also makes the proposed RD models more accurate. These issues had not been addressed elsewhere (including the JVT-G012 rate-control proposal and the related latest JVT reference software). The proposed models are expected to work alongside the hierarchical bit-allocation scheme in the latest JVT reference software, and this remains as the next step of our research work.

References

1. P. Hsu and K. Liu, "A predictive H.263 bit-rate control scheme based on scene information," in *Proc. IEEE Inter. Conf. Multimedia & Expo*, vol. 3, pp. 1735–1738, Aug. 2000.
2. "Draft ITU-t recommendation and final draft international standard of joint video specification (ITU-t rec. h.264/ISO/IEC 14496-10 AVC)," Doc. JVT-G050, Joint Video Team (JVT) of ISO/ICE MPEG and ITU-T VCEG, (2003).
3. T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Trans. Circuits Syst. Video Technol.* **13**(7), 560–576 (Jul. 2003).
4. Generic coding of moving pictures and associated audio information: video," ISO-IEC 13818-2, ISO-IEC/JTC1/SC29/WG11, Nov. 1994.
5. D. L. Gall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM* **34**, 46–58 (Apr. 1991).
6. "Draft of recommendation h.263: Video coding for low bitrate communication," ITU-T Study Group 15, May 1996.
7. J. Ribas-Corbera and S. Lei, "Rate control in dct video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.* **9**(1), 172–185 (Feb. 1999).
8. W. Ding and B. Liu, "Rate control of mpeg video coding and recording by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.* **6**(1), 12–20 (Feb. 1996).
9. "MPEG-2, test model 5 (tm5)," ISO/IEC/JTC1/SC29/WG11/93-225b (Apr. 1993).
10. J. Choi and D. Park, "A stable feedback control of the buffer state using the controlled langrange multiplier method," *IEEE Trans. Image Process.* **3**, 546–558 (Sep. 1994).
11. A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximations," *IEEE Trans. Image Process.* **3**, 26–40 (Jan. 1994).
12. H. Hang and J. Chen, "Source model for transform video coder and its application-part I: Fundamental theory," *IEEE Trans. Circuits Syst. Video Technol.* **7**(2), 287–298 (Apr.1997).
13. H. Gish and J. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inf. Theory* **IT-14**, 676–683 (Sep. 1968).
14. T. Berger, *Rate Distortion Theory*, Prentice-Hall, Englewood Cliffs, (1984).
15. L. Zhao and C. J. Kuo, "Buffer-constrained r-d optimized rate control for video coding," in *Proc. IEEE Inter. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, III-89-92 Apr. 2003.
16. Y. Kim, Z. He, and S. Mitra, "A novel linear source model and a unified rate control algorithm for h.263/mpeg-2/mpeg-4," in *Proc. IEEE Inter. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, 1777–1780, May 2001.
17. H. Lee, T. Chiang, and Y. Zhang, "Scalable rate control for mpeg-4 video," *IEEE Trans. Circuits Syst. Video Technol.* **10**(6), 87–894 (Sep. 2000).
18. L. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.* **8**(4), 446–459 (Aug. 1998).
19. S. Hong, S. Yoo, S. Lee, H. Kang, and S. Hong, "Rate control of mpeg video for consistent picture quality," *IEEE Trans. Broadcast.* **49**(1), 1–13 (Mar. 2003).
20. T. Chiang and Y. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.* **7**(1), 246–250 (Feb. 1997).
21. S. Ma, W. Gao, P. Gao, and Y. Lu, "Rate control for advance video coding (avc) standard," in *Proc. IEEE Inter. Symposium on Circuits and Systems*, vol. 2, II-892–895, May 2003.
22. 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.* **15**(12), 1533–1544 (Dec. 2005).
23. M. Jiang, X. Yi, and N. Ling, "On enhancing h.264 rate control by psnr-based frame complexity estimation," in *Proc. Inter. Conf. Consumer Electronics*, vol. 3, pp. 231–232, Jan. 2005.
24. S. Miyaji, Y. Takishima, and Y. Hatori, "A novel rate control method for h.264 video coding," in *Proc. IEEE Inter. Conf. Image Processing*, vol. 2, pp. II-309–312, Sept. 2005.
25. N. Kamaci, Y. Altunbasak, and R. Mersereau, "Frame bit allocation for the h.264/avc video coder via cauchy-density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.* **15**(8), 994–1006 (Aug. 2005).
26. P. Li, X. Yang, and W. Lin, "Buffer-constrained r-d model-based rate control for h.264/ avc," in *Proc. IEEE Inter. Conf. Acoustics, Speech and Signal Processing*, vol. 2, pp. 321–324, Mar. 2005.
27. T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. Sullivan, "Rate-constrained coder control and compression of video standards," *IEEE Trans. Circuits Syst. Video Technol.* **13**(7), 688–703 (Jul. 2003).
28. "Doc. jvt-b118r2," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, pp. 693–669, Mar. 2002.
29. M. Jiang and N. Ling, "On Lagrange multiplier and quantizer adjustment for h.264 frame-layer video rate control," *IEEE Trans. Circuits Syst. Video Technol.* **16**(5), 663–669 (May 2006).
30. "JVT test model IM," Doc. JVT-D147, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Jul. 2002.
31. P. Li, W. Lin, S. Rahardja, X. Lin, X. Yang, and Z. Li, "Geometrically determining the leaky bucket parameters for video steaming over constant bit-rate channels," *Signal Process. Image Commun.* **20**(2), 193–204 (Feb. 2005).
32. J. Ribas-Corbera, P. Chou, and S. Regunathan, "A generalized hypothetical reference decoder for h.264/ avc," *IEEE Trans. Circuits Syst. Video Technol.* **13**(7), 674–687 (Jul. 2003).

Ping Li received his BEng and MEng degrees in mechanical engineering from Xi'an Jiaotong University in 1998 and 2000. He received his MEng degree in computer engineering from the National University of Singapore in 2003. From March 2003 to July 2004, he was with the Institute for Infocomm Research, Singapore, working on H.264/AVC video coding. In October 2004, he joined the Philips Electronics Singapore Pte Ltd. and started his work on content adaptive sharpness enhancement for LCD displays. Since May 2005, Ping Li has been with the Video Coding and Architectures group at the Eindhoven University of Technology, The Netherlands, and the Video Processing System Group at Philips Research Europe, The Netherlands. His PhD research is on 3-D geometry reconstruction from multiple images.

Weisi Lin graduated from Zhongshan University, China, with a BSc degree in electronics and a MSc degree in digital signal processing in 1982 and 1985, respectively, from King's College, London University, UK, in 1992. He has taught and researched at Computer Vision Zhongshan University, Shantou University (China), Bath University (UK), the National University of Singapore, the Institute of Microelectronics (Singapore), the Centre for Signal Processing (Singapore), and the Institute for Infocomm Research (Singapore). He has been the project leader of 12 projects successfully delivered in digital multimedia technology development. He also serves as the lab head, of visual processing and is the acting department manager of media processing at the Institute for Infocomm Research. Currently, he is an associate professor at the School of Computer Engineering, Nanyang Technological University, Singapore. His areas of expertise include image processing, perceptual modeling, video compression, multimedia communication, and computer vision. He holds 10 patents, wrote four book chapters, and has published over 130 refereed papers in international journals and conferences. He is a senior member of the Institute of Electrical and Electronics Engineers, a member of Institution of Engineering and Technology, and a Chartered Engineer (UK). He believes that good theory is practical, so he has kept a balance of academic research and industrial deployment throughout his working life.

Xiaokang Yang (M'00, AM'04) received his BS degree from Xiamen University, Xiamen, China, in 1994, his MS degree from Chinese Academy of Sciences, Shanghai, China, in 1997, and his PhD degree from Shanghai Jiao Tong University, Shanghai China, in 2000. He is currently a professor and the deputy director of the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. From August 2007 to July 2008, he visited the Institute for Computer Science, University of Freiburg, Germany, as an Alexander von Humboldt Research Fellow. From September 2000 to March 2002, he worked as a Research Fellow in Centre for Signal Processing, Nanyang Technological University, Singapore. From April 2002 to October 2004, he was a research scientist in the Institute for Infocomm Research (I²R), Singapore. He has published

over 130 refereed papers, and has filed 12 patents. His current research interests include visual processing and communication, media analysis and retrieval, and pattern recognition. He actively participates in the International Standards such as MPEG-4, JVT, and MPEG-21. He received the Microsoft Professorship Award 2006, the Best Young Investigator Paper Award at IS&T/SPIE International Conference on Video Communication and Image Processing (VCIP2003) and awards from A-STAR and Tan Kah Kee foundations. He is currently a senior member of IEEE, a member of Design and Implementation of Signal Processing Systems (DISPS) Technical Committee of the IEEE Signal Processing Society and a member of Visual Signal Processing and Communications (VSPC) Technical Committee of the IEEE Circuits and Systems Society. He was the special session chair of Perceptual Visual Processing of IEEE ICME2006. He is the local co-chair of ChianCom2007 and the technical program co-chair of IEE SiPS2007.