Rate Control for H.264 Video With Enhanced Rate and Distortion Models

Do-Kyoung Kwon, *Student Member, IEEE*, Mei-Yin Shen, *Member, IEEE*, and C.-C. Jay Kuo, *Fellow, IEEE*

Abstract—A new rate control scheme for H.264 video encoding with enhanced rate and distortion models is proposed in this work. Compared with existing H.264 rate control schemes, our scheme has offered several new features. First, the inter-dependency between rate-distortion optimization (RDO) and rate control in H.264 is resolved via quantization parameter estimation and update. Second, since the bits of the header information may occupy a larger portion of the total bit budget, which is especially true when being coded at low bit rates, a rate model for the header information is developed to estimate header bits more accurately. The number of header bits is modeled as a function of the number of nonzero motion vector (MV) elements and the number of MVs. Third, a new source rate model and a distortion model are proposed. For this purpose, coded 4×4 blocks are identified and the number of source bits and distortion are modeled as functions of the quantization stepsize and the complexity of coded 4×4 blocks. Finally, a R-D optimized bit allocation scheme among macroblocks (MBs) is proposed to improve picture quality. Built upon the above ideas, a rate control algorithm is developed for the H.264 baseline-profile encoder under the constant bit rate constraint. It is shown by experimental results that the new algorithm can control bit rates accurately with the R-D performance significantly better than that of the rate control algorithm implemented in the H.264 software encoder JM8.1a.

Index Terms—H.264, optimal bit allocation, rate control, rate-distortion (R-D) optimization.

I. INTRODUCTION

$\mathbf{S}$EVERAL coding parameters, such as macroblock (MB) mode, quantization parameter (QP), frame type and frame rates, need to be determined in the implementation of a video encoder. Even though their choice is not specified in the standard, they are important in that their values have a great impact on the performance of the resulting coder. In addition, the QP has another importance in that it can regulate the encoded bit stream. The QP can be effectively used to control the bit stream such that the coding efficiency is maximized without violating the constraints imposed by the channel rate, encoder/decoder buffer sizes and the constant end-to-end delay. The other coding parameters also should be chosen with a guarantee the constraints are not violated. That means they should be optimized jointly with the QP. For example, methods with joint MB mode and QP decision were studied in [1] and [2]. Frame type and frame rate selection algorithms are studied jointly with QP decision in [3]–[5], respectively. Because of its importance in video coding, rate control has been extensively studied for various video coding standards under various applications [6]–[14].

The H.264 video standard has gained much attention recently due to its excellent coding performance over previous standards. Among several new coding techniques adopted, the rate-distortion (R-D) optimized motion estimation and mode decision, which is called RDO, with various intra- and inter-prediction modes and multiple reference frames contributes significantly to the high coding efficiency of H.264 [15], [16]. In contrast, these features make H.264 rate control more complicated. For example, when applying a model-based rate control approach to an H.264 encoder, the residual information such as the mean absolute difference (MAD) or variance is needed to determine the proper QP. However, the residual information is only available after the RDO process that uses a pre-determined QP to generate it. This inter-dependency between RDO and rate control, which was described as a "chicken and egg" dilemma in [17], makes H.264 rate control more challenging than previous standards.

There are two problems caused by the inter-dependency of RDO and rate control. First, the residual signal and its related information are not available before the mode selection of the RDO process and, as a consequence, it is difficult to estimate the number of source bits and the corresponding distortion. Second, the header information such as MB modes, motion vectors (MVs) and reference frames are not available before the RDO process, either. This makes the number of header bits difficult to estimate. Most previous work [17]–[22] attempts to resolve the inter-dependency between RDO and rate control by examining the MAD or variance of the residual signal only. However, this approach may not be accurate enough. Due to various coding modes allowed in H.264, the number of bits associated with the header information varies a lot from frame to frame. Sometimes, they may even occupy a larger portion than source bits in the total bit budget. The impact of header bits is more obvious at low bit rates. Thus, an accurate estimate of header bits is critical to H.264 rate control.

Several model-based rate control algorithms have been proposed for H.264. In [17], the quadratic rate model proposed in [23] is employed to determine the QP of a basic unit, which can be either a frame, a slice or a MB. Since the residual signal is not available before the RDO process, the MAD of each basic unit in the current frame is estimated by the MAD of the col-located basic unit in the previous frame using a linear model.

The same quadratic rate model and the linear MAD estimation model are used in [19]. However, to improve the performance of [17], a new complexity measure using the MAD ratio and the PSNR drop ratio is developed for the improved bit allocation. In [18], the H.263 TMN8 rate and distortion models proposed in [10] are employed. The residual signal of each frame is first estimated by performing the RDO process with a reduced set of reference frame, intra- and inter-prediction modes. After that, the standard deviation of the estimated residual signal is fed into the H.263 TMN8 rate and distortion models to refine the QP of each MB. Then, the RDO process is performed again using the refined QP. Other H.264 rate control algorithms, e.g., [20] and [21], were developed by following a similar idea. It is worthwhile to point out that none of the above methods address the importance of header bits. Furthermore, most of them do not work well for video that consists of frames of variable characteristics, since the rate control performance degrades due to an inaccurate estimate of the complexity of residual signals under such a scenario.

With respect to the above shortcomings, we propose an enhanced rate control scheme for H.264 video with a few new features. First, the inter-dependency between RDO and rate control is resolved by allowing different QP values at the RDO process and the quantization process, respectively. As long as these QPs are close to each other, we have a good approximation. Second, to address the increased importance of header bits, a header rate model is established so as to estimate header bits more accurately. To be more specific, the number of header bits is modeled as a function of the number of nonzero MV elements and the number of MVs. Third, a new source rate model and a distortion model are proposed. For this purpose, coded 4×4 blocks are identified and the number of source bits and distortion are modeled as functions of the quantization stepsize and the complexity of coded 4×4 blocks. Finally, a R-D optimized bit allocation scheme among MBs is proposed to improve picture quality. Built upon the above ideas, a rate control algorithm is developed for the H.264 baseline-profile encoder under the CBR constraint. It is shown by experimental results that the new algorithm can control bit rates accurately with the R-D performance significantly better than that of the rate control algorithm implemented in the H.264 software encoder JM8.1a.

The rest of this paper is organized as follows. An overview of the proposed encoder structure is proposed in Section II, where the inter-dependency between RDO and rate control in H.264 is decoupled via QP estimation and update. Improved rate and distortion models are described in Section III while the constant bit rate control algorithm for the H.264 baseline-profile encoder is presented in Section IV. Experimental results are provided in Section V. Finally, concluding remarks are drawn in Section VI.

II. PROPOSED ENCODER STRUCTURE

The inter-dependency of RDO and rate control is the main difference between the rate control problem for H.264 and prior standards. In this section, we propose an encoder structure that decouples this dependency via the use of two QP values. In the following discussion, we use QP_1 and QP_2 to denote QPs used in the RDO process and the quantization process, respectively.

If there exists a known QP for RDO and quantization, i.e., $QP = QP_1 = QP_2$, the best set of reference frames, the MVs of partitions and the MB mode for each MB can be determined via RDO by minimizing the following Lagrange cost [24]:

$$J(S_{QP}) = D(S_{QP}) + \lambda(QP) \cdot R(S_{QP}) \qquad (1)$$

where S_{QP} represents a set of MVs, reference frames and the MB mode, and $\lambda(QP)$ is the Lagrangian multiplier. Please note that there are two Lagrange multipliers (one for motion estimation and the other for mode decision) and both of them depend on the pre-determined QP. Although various RDO methods and their low complexity variants such as fast motion estimation and mode decision have been proposed to speed up the encoding process [25], [26], all these algorithms eventually determine MVs, reference frames and a MB mode by minimizing (1).

However, it is difficult to find $QP = QP_1 = QP_2$ *a priori* due to the inter-dependency of RDO and rate control. In practice, we have to begin with one QP value, say QP_1, for the RDO process to determine the set of MVs, reference frames and the MB mode. Then, under such a setting, we can determine the best QP_2 for quantization in the rate control process. If $QP_1 = QP_2$, the process is done. If $QP_1 \neq QP_2$, we may assign $(QP_1 + QP_2)/2$ to $QP_1^{(2)}$, where the superscript (2) denotes that this is the second iteration. After that, with $QP_1^{(2)}$ and RDO, we can determine the corresponding MVs, reference frames and the MB mode. Then, we can find $QP_2^{(2)}$ for quantization. Typically, the gap between $QP_1^{(k)}$ and $QP_2^{(k)}$ becomes smaller as the iteration number k becomes larger. This iteration can continue until the following stopping criterion is met:

$$\left| QP_1^{(k)} - QP_2^{(k)} \right| \leq \delta \qquad (2)$$

where δ is a parameter to be chosen. Since the iteration process is computationally expensive, we would like to find good initial values of QP_1 and QP_2 and determine the proper value of δ to control the complexity while keeping good coding performance.

The following two observations provide useful guidelines in selecting QP_1, QP_2, and δ.

- *Observation 1:*
  The variation between the QPs of two consecutive frames is often restricted to a small range for smooth video playback; namely

  $$|QP_n - QP_{n-1}| \leq \Delta \text{ where } \Delta \leq 3$$

  where QP_n and QP_{n-1} are the average QPs of the nth and the $(n-1)$th frames, respectively.
- *Observation 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, i.e.,

  $$|QP_1 - QP_2| \leq \Delta \text{ where } \Delta \leq 3.$$

Observation 1 is often exploited in model-based rate control algorithms [10], [12], [17]–[20] to smooth the quality variation between frames and between MBs. Fig. 1 provides the evidence to Observation 2. This figure shows R-D curves with fixed QP_1 for RDO and varying QP_2 for quantization of the
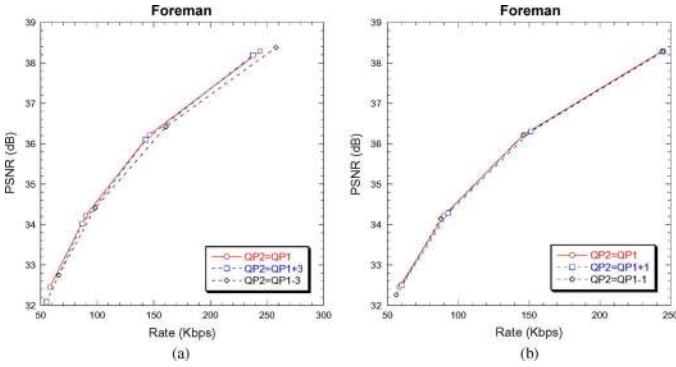
Fig. 1. R-D curves by setting QP_2 values in quantization to (a) QP_1, $QP_1 + 3$ and $QP_1 - 3$ and (b) QP_1, $QP_1 + 1$, and $QP_1 - 1$, where QP_1 is used in RDO.
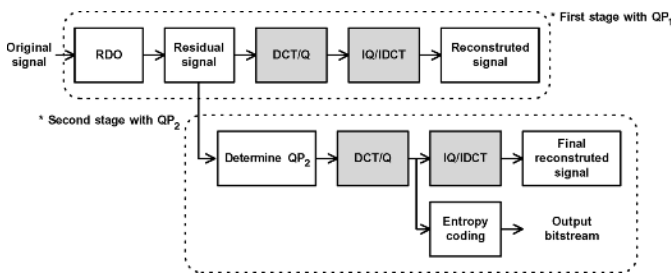

Fig. 2. Proposed encoder structure, where the RDO is performed for all MBs using QP_1 at the first stage and the residual signal of a MB is then quantized using QP_2 at the second stage.

QCIF "Foreman" sequence encoded with five reference frames. We can see that the coding gain loss is around 0.2 dB when $|QP_1 - QP_2| = 3$ and it is almost negligible when $|QP_1 - QP_2| = 1$. Clearly, observation 2 can be used to choose parameter δ in (2). Please note that we observe almost identical results for numerous test sequences and show only "Foreman" as a representative one in Fig. 1.

In Fig. 2, we show the proposed encoding scheme that consists of the following two stages.

- *Stage 1:*

  To encode a new frame, it is natural to choose QP_1 to be the average QP_2 of a previous frame based on observation 1, and perform RDO for all MBs in the current frame to determine the residual signal and the header information. The recursive property of intra-prediction requires previous MBs to be reconstructed. Thus, 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. Throughout stage 1, only QP_1 is used.

- *Stage 2:*

  Given the target number of bits for the frame, the QP_2 value of all MBs is determined by the rate control algorithm (to be discussed in Sections III and IV) and used to quantize the residual signals. If an MB is inter-coded, its residual signal is simply requantized using QP_2. If a 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

using the same intra-mode determined in the first stage of encoding. The final reconstructed frame is obtained via IQ/IDCT for the intra-prediction of subsequent MBs and the inter-prediction of subsequent frames. The output bit stream can also be easily generated via entropy coding as shown in the figure. Throughout stage 2, only QP_2 is used.

After the above two-stage process, we can verify the difference between QP_1 and QP_2. If the difference is less or equal to 3 for a MB, it is done. If not, one possible solution is to do some iterations to narrow down the gap between them. However, the complexity could be too high to be practical. Another solution is to choose QP_2 to be either $QP_1 + 3$ or $QP_1 - 3$ and accept the consequence of R-D performance degradation. The latter solution is adopted in this work. It should be noted that the high coding complexity of the H.264 encoders comes mainly from the RDO process. The proposed two-stage encoding scheme demands only one additional forward and inverse DCT and one more quantization process. The extra computational cost is relatively small with respect to the overall H.264 encoding complexity, which depends on the complexity of RDO, the number of reference frames, etc. For example, as the number of reference frames increases, the impact of the two-stage encoding scheme on the total encoder complexity decreases even more. Since the residual signal and the header information are determined in the first stage and all of them except for residual signals of intra-MBs do not change throughout the encoding process, we conclude that rate and distortion can be accurately estimated based on the information obtained in the first stage.

In the next two sections, we will focus on the selection of QP QP_2 given the information from stage 1 by an improved rate control scheme.

III. ENHANCED RATE AND DISTORTION MODELS

Accurate rate and distortion models have great impact on the rate control performance in model-based rate control algorithms. In this section, enhanced rate models for header bits and source bits and a new distortion model are proposed. They are used to estimate rates and distortion of the residual signal obtained at the first stage of the proposed encoder given in Section II.

A. Rate Model for Header Bits

In the design of model-based rate control algorithms for previous standards such as MPEG-1/2, H.263 and MPEG-4, the number of bits in the header is less critical due to two reasons. First, the amount of header bits is small as compared with that of source bits. Second, the number of header bits is nearly constant for the same type of frames so that it can be easily estimated by the average header bits of previous frames. However, the situation changes for H.264. The average percentages of header bits of P-frames at various QPs are shown in Fig. 3(a) when the QCIF "Foreman" sequence is encoded with a single reference frame by the H.264 software encoder. The frame-by-frame variations of source bits and header bits when the same sequence is encoded using $QP = 35$ is shown in Fig. 3(b). Note that both MV bits and MB mode bits are included in the header bits. Although the coded block pattern (CBP) is also a type of header information, we classify CBP bits to source bits since they have
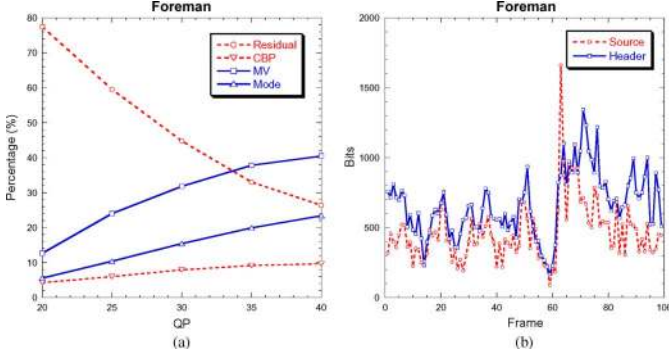
Fig. 3. (a) Average percentages of header bits (including MVs and modes) and source bits (including residual and CBP) of P-frames at various QPs and (b) the source bits and header bits of a P-frame when $QP = 35$ as a function of the frame number.

a strong relationship with residual bits. For example, if there is no residual signal to encode, the number of residual bits and the number of CBP bits are both equal to zero.

As shown in Fig. 3(a), header bits occupy a significant portion of the total amount of bits, and the percentage of header bits increases as the QP value becomes larger. When a sequence is encoded at very low bit rates, the amount of header bits may even exceed that of source bits. Furthermore, we see from Fig. 3(b) that the size of header bits fluctuates a lot from frame to frame, which means header bits cannot be simply estimated by the average value of previous frames. It is also observed that the percentage of header bits increases as the number of reference frames increases. As a result, it is clear that the header information is as important as the residual signal to the rate control of H.264.

Let R_T, R_{src}, and R_{hdr} denote the total number of bits allocated to a frame, and the corresponding source and header bits, respectively. We can first estimate the header bits after the RDO process, and then compute available source bits as

$$R_{src} = R_T - R_{hdr}. \tag{3}$$

Here, we consider a header rate model that estimates the total number of bits required to encode MVs, reference frames and MB modes. Since inter-MBs and intra-MBs are quite different, we estimate their header bits separately.

The header bits of inter-MBs have a strong relationship with the number of nonzero horizontal/vertical MV elements, N_{nzMVe}, and the number of MVs, N_{MV}. To give an example, suppose a MB is partitioned into four 8×8 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 $N_{MV} = 4$. As the number of nonzero MV elements increases, the bits required to encode their values increases. Likewise, as a MB is finely partitioned, there are more MVs associated with it, which results in an increase in header bits. It is observed from experiments that header bits of inter-MBs in the current frame can be estimated by the following model:

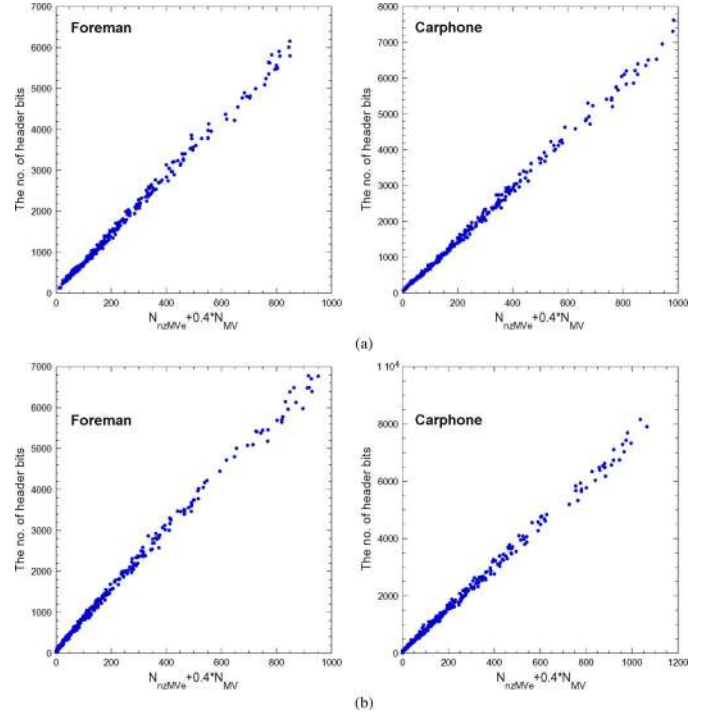$$R_{hdr,inter} = \gamma_1 \cdot N_{nzMVe} + \gamma_2 \cdot N_{MV} \tag{4}$$

Fig. 4. Relationship between $R_{hdr,inter}$ and $(N_{nzMVe} + \omega \cdot N_{MV})$ for test sequences. (a) P-frames. (b) B-frames.

where γ_1 and γ_2 are model parameters. In addition, at the cost of little estimation error, (4) can be simplified as

$$R_{hdr,inter} = \gamma \cdot (N_{nzMVe} + \omega \cdot N_{MV}) \tag{5}$$

where γ is a model parameter and ω is a weighting factor that depends on the number of reference frames used. Note that the number of reference frames is another important factor that affects the number of header bits. Clearly, the required bits to encode the indexes of reference frames increases as the number of reference frames increases. Thus, if more reference frames are used for inter-prediction, a larger weight should be given to N_{MV}. We find empirically that

$$\omega = \begin{cases} 0.5, & \text{if the no. of reference frames} \geq 5 \\ 0.4, & \text{if the no. of reference frames} \geq 3 \\ 0.3, & \text{otherwise.} \end{cases} \tag{6}$$

The relationship between $R_{hdr,inter}$ and $(N_{nzMVe} + \omega \cdot N_{MV})$ for QCIF test sequences is given in Fig. 4. We consider P- and B-frames separately and show their results in Figs. 4(a) and (b), respectively. For each plot, 30 frames are encoded using every three QP values ranging from 15 to 45 with three reference frames (i.e., $\omega = 0.4$). From these plots, we see that both P- and B-frames of all test sequences follow the header rate model given in (5) closely.

For the same experiment, Table I shows the estimation errors in RMSE and the R^2 values of P-frames by (4) and (5), respectively. The R^2 is a quantity used to measure the degree of data variation from a given model [27]. It is defined as

$$R^2 = 1 - \frac{\sum_i (X_i - \hat{X}_i)^2}{\sum_i (X_i - \bar{X})^2} \tag{7}$$

TABLE I
PERFORMANCE COMPARISON OF THE PROPOSED HEADER RATE MODELS
FOR P-FRAMES

Sequences	Eq. (4)		Eq. (5)	
	RMSE	R^2	RMSE	R^2
News	49.7	0.9929	62.8	0.9894
Carphone	86.9	0.9973	88.5	0.9972
Table Tennis	164.3	0.9862	180.8	0.9853
Mot & Dau	39.2	0.9969	84.2	0.9874
Foreman	88.6	0.9961	90.7	0.9960
Silent	92.1	0.9921	95.4	0.9922
Coastguard	97.4	0.9968	125.3	0.9951

where X_i and $\hat{X}_i$ are the actual and estimated values of data point i, respectively, and $\bar{X}$ is the mean of all data points. The estimation $\hat{X}_i$ is obtained using a data model. For any reasonable model, we expect that the second term of the right-hand-side of (7) is less than one. Thus, R^2 takes the value between 0 and 1. The better the model, the closer the R^2 value to 1. As shown in Table I, the estimation error based on the model in (5) is not much larger than that based on the model in (4) for most sequences. Consequently, we use (5) to estimate the number of header bits of inter-MBs for simplicity.

For intra-MBs, the header information includes only MB modes such that the amount of header bits is much smaller than that of source bits. Thus, an exact estimation is not critical as much as inter-MBs. We simply estimate the number of header bits of intra-MBs in the current frame by

$$R_{\text{hdr,intra}} = N_{\text{intra}} \cdot b_{\text{intra}} \qquad (8)$$

where N_{intra} is the number of intra-MBs in the current frame and b_{intra} is the average number of header bits of intra-MBs in previous frames. Then, the total number of header bits of the current frame is equal to the sum of header bits from inter- and intra-MBs in the frame, i.e.,

$$R_{\text{hdr}} = \gamma \cdot (N_{nz\text{MV}e} + \omega \cdot N_{\text{MV}}) + N_{\text{intra}} \cdot b_{\text{intra}}. \qquad (9)$$

B. Rate Model for Source Bits

Two types of rate and distortion models have been considered in q-domain and ρ-domain, respectively. In the q-domain approach, the source rate is modeled as a function of the quantization stepsize and the complexity of the residual signal. Several source rate models were studied in the past, e.g., [8]–[10], [23]. In [23], the source is assumed to be Laplacian distributed, and a quadratic function of the quantization stepsize is employed to estimate source bits. The following quadratic model

$$R_{\text{src}}(Q) = \alpha_1 \cdot \frac{\text{MAD}}{Q} + \alpha_2 \cdot \frac{\text{MAD}}{Q^2} \qquad (10)$$

where α_1 and α_2 are model parameters and Q is the quantization stepsize, has been widely studied. It was adopted as a nonnormative guidance for the rate control implementation in several standards such as MPEG-4 [28], [29] and H.264 [17].

In the ρ-domain approach, where ρ is the percentage of zero DCT coefficients, the number of source bits is modeled as a

linear function of $(1 - \rho)$ [12], [13]. To determine the quantization stepsize Q from ρ, it is possible to find the one-to-one correspondence between them. That is, we can build a DCT coefficient histogram and find the relationship between ρ and Q to determine the corresponding Q. The ρ-domain source rate model is often more accurate than the q-domain source rate model. The inaccuracy of the q-domain source rate model is mainly due to the rough estimate of the complexity of residual signals such as MAD given in (10).

The basic unit for DCT and quantization in H.264 is a 4×4 block, which can be either coded or skipped. A 4×4 block is skipped if all of its 4×4 DCT coefficients are zero after quantization. Otherwise, it is a coded block. It should be noted that no bit is required for skipped block. Therefore, in this work, we consider the complexities of coded blocks only so as to estimate the complexity of the residual signal more precisely. Let SATD_c be the sum of absolute transform differences (SATD) of coded 4×4 blocks. By modifying (10) slightly, we can derive another source rate model

$$R_{\text{src}}(Q) = \alpha_1 \cdot \frac{\text{SATD}_c(Q)}{Q} + \alpha_2 \cdot \frac{\text{SATD}_c(Q)}{Q^2} \qquad (11)$$

where α_1 and α_2 are model parameters and Q is a quantization stepsize. It is clear that the difference between (10) and (11) lies in the replacement of MAD with $\text{SATD}_c(Q)$ to characterize the complexity of the residual signal. Please note that Q in parentheses implies SATD_c depends on the quantization stepsize. The MAD of coded blocks, i.e., MAD_c could be used instead of SATD_c in (11). However, SATD_c is used because of its slightly better performance in the source rate model [26].

It is confirmed by experiments that the estimation error of the quadratic rate model in (10) can be significantly reduced by considering only the complexities of coded 4×4 blocks. Furthermore, it is observed that the second order term Q^{-2} in (11) can be dropped without sacrificing much the modeling performance. As a result, we can obtain a much simplified source rate model as

$$R_{\text{src}}(Q) = \alpha \cdot \frac{\text{SATD}_c(Q)}{Q^p} \qquad (12)$$

where α is a model parameter and p is a value depending on the frame type. It is observed that the best value of p is 1.0 for P- and B-frames and 0.8 for I-frames. The relationship between the source rate and $\text{SATD}_c(Q)/Q^p$ for QCIF test sequences is plotted in Fig. 5. For each plot, 30 frames are encoded using every three QP values ranging from 15 to 45 with three reference frames for P- and B-frames. We see from Fig. 5 that the linear relationship as shown in (12) for all frame types is confirmed.

We study the performance of different source rate models, i.e., the proposed quadratic model in (11), the simplified model in (12), the q-domain model [23] in (10), and the ρ-domain model [12], [13], by comparing the estimation errors and the R^2 values of P-frames in Table II. We see that, for all test sequences except for "Carphone," the simplified model gives smaller RMSE values and larger R^2 values than q-domain and ρ-domain models. Moreover, the estimation errors of the simplified model are only slightly larger than those of the proposed
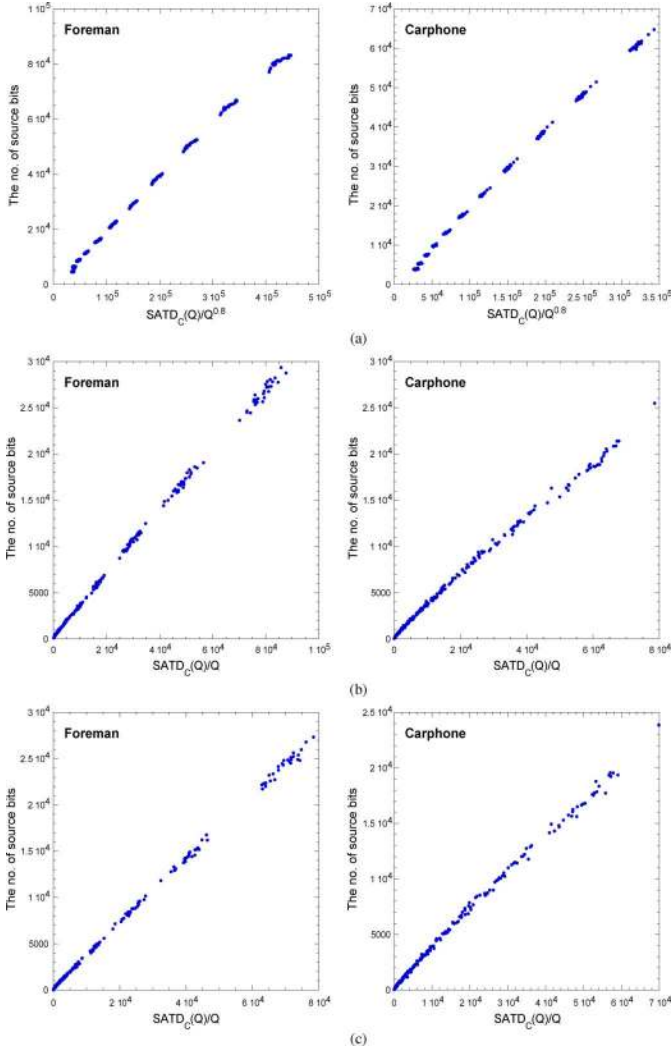
Fig. 5. Relationship between the source rate and $\mathrm{SATD}_c(Q)/Q^p$ for for test sequences. (a) I-frames. (b) P-frames. (c) B-frames.

quadratic model. Consequently, we adopt (12) to estimate source bits in our rate control algorithms for its simplicity.

C. Distortion Model

Generally speaking, distortion can be approximated by an exponential function of source bits. In [13], distortion was modeled as an exponential function of source bits in the ρ-domain, and it was shown that the number of source bits and distortion can be estimated more accurately in the ρ-domain than in the q-domain. However, the inferior performance in the q-domain is mainly attributed to the poor estimation of source bits. Since the source rate model proposed in Section III-B is more accurate, the accuracy of the distortion model is enhanced accordingly. However, in this subsection, we consider a new approach to improve the distortion estimation furthermore.

Let D_c and D_s represent the distortion measures of all coded 4×4 blocks and skipped 4×4 blocks, respectively. Then, the total distortion D is the sum of two distortion measures, i.e.,

$$D(Q) = D_c(Q) + D_s(Q). \tag{13}$$

Since the residual signal of a skipped 4×4 block is not coded, its distortion can be computed directly from its residual signal. Therefore, we need to estimate the distortion of coded 4×4 blocks only. The distortion of coded blocks can be estimated via

$$D_c(Q) = \beta_1 \cdot \mathrm{SATD}_c(Q) \cdot Q + \beta_2 \cdot \mathrm{SATD}_c(Q) \cdot Q^2 \tag{14}$$

where β_1 and β_2 are model parameters. Similar to the source rate model, the quadratic model in (14) can be simplified to be

$$D_c(Q) = \beta \cdot \mathrm{SATD}_c(Q) \cdot Q^p \tag{15}$$

where β is a model parameter and p is the value depending on the frame type. It is observed that the best value of p is 1.0 for P- and B-frames and 1.2 for I-frames.

The distortion of coded 4×4 blocks, measured in terms of the sum of squared errors (SSE), versus the quantity $\mathrm{SATD}_c(Q) \cdot Q^p$ for QCIF test sequences is plotted in Fig. 6. The same experimental conditions with Fig. 5 are applied to Fig. 6. We see an approximately linear relationship between them from this figure, which justifies the model given in (15). However, we should note that we are concerned with the total distortion of the whole frame (including both skipped and coded 4×4 blocks) rather than the distortion of the coded 4×4 blocks only. Fig. 7 plots the relationship between the actual total distortion (i.e., $D_c + D_s$) and the estimated total distortion [i.e., $\hat{D}_c + D_s$, where $\hat{D}_c$ is estimated by (15)]. Fig. 7 also shows a dotted $y = x$ line in every plot. We see that the total distortion can be closely estimated by the proposed estimation method after identifying coded 4×4 blocks.

We compare the performance of different distortion models, i.e., the models given by (14) and (15) and the ρ-domain distortion model [13] in Table III in terms of the estimation errors and the R^2 values for P-frames. We see from Table III that, for all test sequences, the estimation errors are significantly reduced by the proposed estimation method with the proposed distortion models of coded blocks. In our rate control algorithms, we adopt the simplified model in (15) in estimating the distortion of coded blocks for its simplicity and good performance.

D. Block Type Identification

To apply the proposed source rate and distortion models properly, we have to identify whether a 4×4 block is coded or not at different QPs after DCT and quantization. In H.264, the integer DCT and the corresponding scalar quantization are applied to a 4×4 block to avoid multiplications and divisions [30]. Let $X(i,j)$ and $X_q(i,j)$ denote the DCT coefficients at position (i,j) before and after the quantization operation, respectively. Coefficient $X(i,j)$ is quantized via

$$X_q(i,j) = \mathrm{sign}\{X(i,j)\} \cdot [(|X(i,j)| \cdot A(Q_M,i,j) \\ + f \cdot 2^{17+Q_E}) \gg (17+Q_E)] \tag{16}$$

where $Q_M = \mathrm{QP} \bmod 6$, $Q_E = \mathrm{QP}/6$, $A(Q_M,i,j)$ is a function specified by the quantization table that gives the integer quantization scaling factor at (i,j), and f controls the quantization width near the origin. As indicated above, Q_M and Q_E

TABLE II
PERFORMANCE COMPARISON OF DIFFERENT SOURCE RATE MODELS FOR P-FRAMES

| Sequences	Eq. (11)		Eq. (12)		q-domain [23]		p-domain [13]	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
News	118.4	0.9987	204.5	0.9962	764.9	0.9425	286.1	0.9933
Carphone	185.4	0.9990	445.4	0.9945	735.1	0.9835	427.9	0.9947
Table Tennis	276.3	0.9989	278.3	0.9984	1,253.8	0.9726	778.7	0.9930
Mot & Dau	160.4	0.9964	177.8	0.9956	668.6	0.9357	432.8	0.9832
Foreman	220.3	0.9993	324.2	0.9985	801.03	0.9908	717.2	0.9934
Silent	126.6	0.9983	188.1	0.9964	985.0	0.9160	369.8	0.9898
Coastguard	354.5	0.9994	362.2	0.9993	1,936.4	0.9724	994.8	0.9940
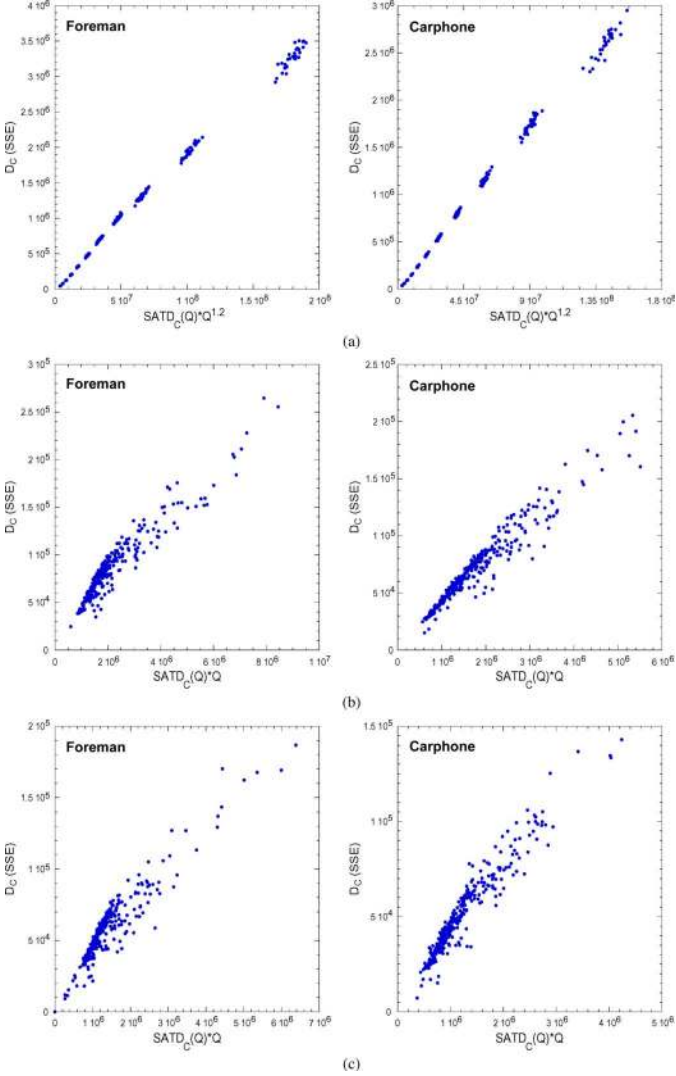
Fig. 6. Relationship between the distortion of coded blocks and $\mathrm{SATD}_c(Q) \cdot Q^p$ for test sequences. (a) I-frames. (b) P-frames. (c) B-frames.
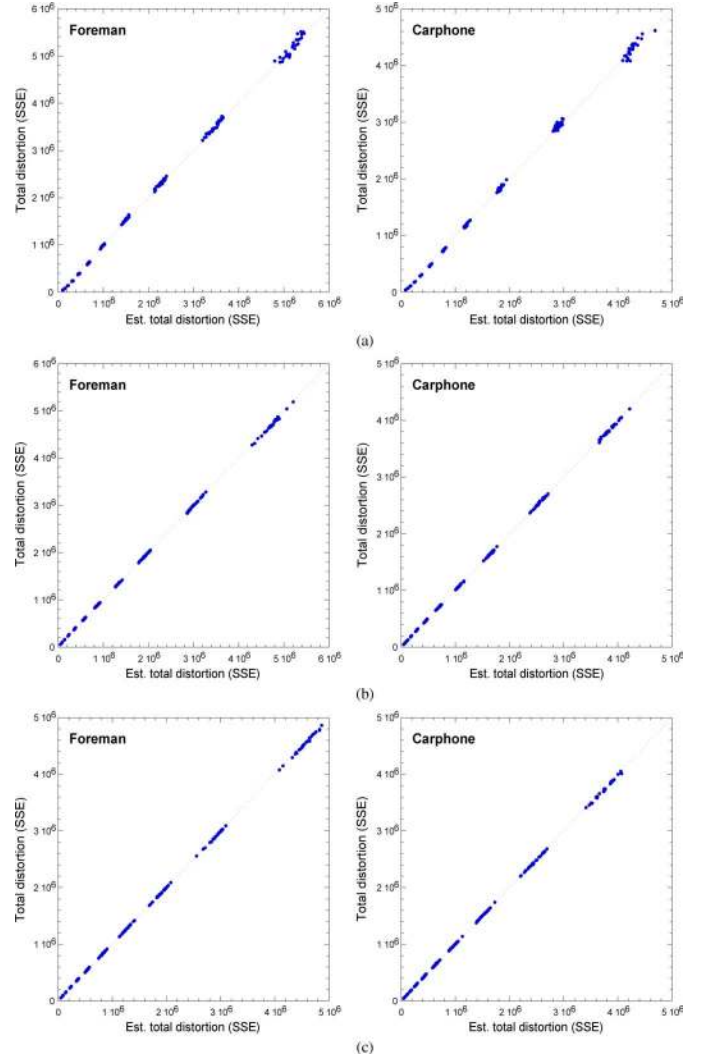

Fig. 7. Relationship between the actual total distortion and the estimated total distortion for test sequences. (a) I-frames. (b) P-frames. (c) B-frames.

depend on the QP and, for that reason, the quantization process is performed differently.

Coded 4×4 blocks are identified during the DCT and quantization module at the first stage of our proposed encoder structure shown in Fig. 2. Since the difference between QP_1 and QP_2 is restricted to Δ (which is set to 3 in this work) as mentioned before, we only need to identify block types at the QP values from $\mathrm{QP}_1 - \Delta$ to $\mathrm{QP}_1 + \Delta$. To determine whether a 4 $\times$ 4 block is coded or not at these admissible values, we first

find DCT coefficient $X_{\max}$ that gives the maximum value of $|X(i,j)| \cdot A(Q_M, i, j)$ with QP_1

$$X_{\max} = \operatorname*{arg\,max}_{1 \leq i,j \leq 4} |X(i,j)| \cdot A(Q_M, i, j). \quad (17)$$

After that, $X_{\max}$ is requantized using all admissible values. Please note that, while $A(Q_M, i, j)$ depends on the QP, their rank order remains the same regardless of it. In other words, $X_{\max}$ has the maximum value of $|X(i,j)| \cdot A(Q_m, i, j)$ at other QPs, too. Thus, by quantizing $X_{\max}(i,j)$ with all admissible

TABLE III
PERFORMANCE COMPARISON OF DIFFERENT DISTORTION MODELS FOR P-FRAMES

Sequences	Eq. (14)		Eq. (15)		ρ-domain [13]	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
News	4,621	0.9999	5,945	0.9999	39,744	0.9980
Carphone	4,889	0.9999	5,940	0.9999	68,862	0.9887
Table Tennis	4,496	0.9999	7,550	0.9999	51,460	0.9931
Mot & Dau	1,824	0.9999	2,987	0.9999	15,635	0.9989
Foreman	5,514	0.9999	8,997	0.9998	61,874	0.9929
Silent	9,046	0.9998	11,437	0.9997	80,449	0.9927
Coastguard	10,837	0.9999	12,313	0.9998	94,390	0.9911

values, we can identify whether a 4×4 block is coded or not for this range of QPs. That is, if $X_{\max}$ is zero after quantization at a particular QP, the 4×4 block is a skipped block at this value. Otherwise, it is a coded block.

IV. RATE CONTROL FOR H.264 BASELINE-PROFILE ENCODER

Although the ρ-domain rate and distortion models have been employed successfully for various video encoders, they are difficult to apply with respect to the H.264 encoder. The main problem is that it is not easy to find the one-to-one correspondence between ρ and Q in H.264 due to the complicated expression given in (16). To apply the ρ-domain rate and distortion models, we may quantize all DCT coefficients using all the admissible QPs to obtain (ρ, Q) pairs. It demands more computations than the coded 4×4 block identification method as described above. Generally speaking, the ρ-domain approach is less direct than the q-domain approach with the enhanced rate and distortion models discussed in Sections III-A–III-C. In this section, in order to show the efficiency of the two-stage encoder structure and the enhanced rate and distortion models, we propose two constant bit rate control algorithms for the H.264 baseline-profile encoder, which can be used in the real-time conversational video coding applications such as video conferencing.

Based on the two-stage encoder structure and the enhanced rate model, we propose a straightforward H.264 rate control algorithm for the baseline-profile encoder as follows.

Rate Control Algorithm Without Bit Allocation:

1) Initial bit allocation to a frame.
   Allocate a certain bit budget to the current frame, R_T. A simple frame-level bit allocation method is described in Section V.

2) First-stage encoding.
   Perform the first stage of encoding using QP_1 to get the residual signal and the header information of the current frame, where QP_1 is chosen to be the average QP_2 of the previous frame. The coded 4×4 blocks are identified using the method described in Section III-D and the corresponding $\mathrm{SATD}_c(Q)$ of all MBs are computed for $QP \in (QP_1 - \Delta, QP_1 + \Delta)$. Estimate the required number of header bits using (9) so that the available source bits R_{src} can be obtained.

3) Second-stage encoding.
   Suppose that the current MB number is m. Let R_{src} be the available source bits before encoding the mth MB. We can determine the QP QP_2 for the mth MB, denoted by $QP_{2,m}$, as follows. First, compute the sum of $\mathrm{SATD}_c(Q)$ of remaining MBs for $QP \in (QP_1 - \Delta, QP_1 + \Delta)$. Using

(12) and $\mathrm{SATD}_c(Q)$ values, we can estimate source bits, $\hat{R}_{\mathrm{src}}(Q)$ for the QPs ranging from $QP_1 - \Delta$ to $QP_1 + \Delta$, and choose the value that minimizes the distance between the available source bits and the estimated source bits, i.e.,

$$QP_{2,m} = \arg\min_{QP_1 - \Delta \leq q \leq QP_1 + \Delta} \left| \hat{R}_{\mathrm{src}}(Q(q)) - R_{\mathrm{src}} \right|$$

where $Q(q)$ is the actual quantization stepsize with respect to a QP q. Then, we can encode the residual signal of the mth MB using $QP_{2,m}$ and update R_{src} by subtracting the actual source bits of the mth MB from it. The above procedure is repeated for all MBs in the current frame.

4) Model update.
   Update model parameters γ and α and the buffer fullness. The model parameters are updated by the least square approximation [23] using data from the previous five frames. Then, we proceed to the next frame until the last frame is reached.

Please note that there are $(2\Delta + 1)$ values of $\mathrm{SATD}_c(Q)$ for each MB and $(2\Delta + 1)$ estimated source bits $\hat{R}_{\mathrm{src}}(Q)$ corresponding to each QP between $QP_1 - \Delta$ and $QP_1 + \Delta$. Therefore, $QP_{2,m}$ is determined from a discrete set of values by minimizing the distance to the available source bits rather than being computed from a closed-form source rate model directly.

The rate control algorithm without bit allocation does not take distortion into account in determining $QP_{2,m}$. To further improve picture quality, a R-D optimized bit allocation method is proposed among MB classes. Let N_c be the number of classes, and $R_{\mathrm{src},l}$ and D_l be the number of source bits and the distortion of the lth class, respectively. Given available source bits R_{src} for a frame, our objective is to determine $R_{\mathrm{src},l}$ for all $1 \leq l \leq N_c$ such that the total distortion of the frame is minimized, i.e.,

$$\text{minimize} \quad \sum_{l=1}^{N_c} D_l \quad \text{subject to} \quad \sum_{l=1}^{N_c} R_{\mathrm{src},l} = R_{\mathrm{src}}. \quad (18)$$

Several MB classification methods were proposed for various purposes including rate control before [31]. A good MB classification method can improve the coding performance. In this work, a simple grouping scheme is considered, where MBs in the same row, i.e., group of s (GOB) are classified into the same class, to demonstrate the idea.

To perform R-D optimized bit allocation among MB classes, a discrete set of R-D data for the lth class is estimated after the first stage of encoding for all $1 \leq l \leq N_c$. For example, $\mathrm{SATD}_c(Q)$ of each class is computed by the sum of $\mathrm{SATD}_c(Q)$ of all MBs in that class for the QPs from $QP_1 - \Delta$ to $QP_1 + \Delta$.

Sequences	Target bitrate	JM8.1a		Proposed w/o Bit Alloc.		Proposed w/ Bit Alloc.	
		Bitrate	PSNR	Bitrate	PSNR	Bitrate	PSNR
Mot & Dau	48.00	48.06	36.30	48.01	36.62 (+0.32)	48.00	37.05 (+0.75)
	64.00	64.04	37.73	64.00	38.01 (+0.28)	64.00	38.40 (+0.67)
	96.00	96.12	40.14	96.00	40.38 (+0.24)	95.99	40.55 (+0.41)
Silent	48.00	48.03	32.18	48.01	32.19 (+0.01)	47.99	32.57 (+0.39)
	64.00	64.03	33.54	64.00	33.56 (+0.02)	64.00	33.95 (+0.41)
	96.00	96.05	36.65	96.01	36.82 (+0.17)	96.00	37.02 (+0.37)
Salesman	48.00	48.01	33.78	48.00	33.99 (+0.21)	47.99	34.66 (+0.88)
	64.00	64.04	35.36	64.01	35.59 (+0.23)	63.98	36.26 (+0.90)
	96.00	96.09	38.51	96.00	38.78 (+0.27)	95.99	39.18 (+0.67)
Carphone	48.00	48.06	31.07	48.01	31.44 (+0.37)	48.00	31.66 (+0.59)
	64.00	64.09	32.43	64.01	32.74 (+0.31)	63.99	32.94 (+0.51)
	96.00	96.11	34.41	96.01	34.67 (+0.26)	95.99	34.77 (+0.36)
Foreman	48.00	48.11	29.99	48.00	30.25 (+0.26)	48.00	30.52 (+0.53)
	64.00	64.12	31.53	64.00	31.79 (+0.26)	64.00	31.93 (+0.40)
	96.00	96.23	33.52	96.00	33.70 (+0.18)	96.00	33.83 (+0.31)
News	48.00	48.10	32.89	48.02	33.24 (+0.35)	47.99	33.78 (+0.89)
	64.00	64.15	34.48	63.99	34.87 (+0.39)	63.99	35.50 (+1.02)
	96.00	96.21	37.79	95.99	38.02 (+0.23)	95.99	38.38 (+0.59)

Then, the number of source bits and the corresponding distortion of each class can be estimated for each admissible QP. That is, the source bits and distortion are estimated by the simplified source rate model in (12) and the simplified distortion model in (15), respectively. Using the estimated sets of R-D data, the following algorithm can be conducted, which performs R-D optimized bit allocation among MB classes using the gradient descent method [32].

Bit Allocation Among MB Classes:

1) Initialization with $q_l = \text{QP}_1 - \Delta$.
   We initialize the rate allocation with the finest QP and compute $\hat{R}_{\text{src},l} = \hat{R}_{\text{src},l}(Q(q_l))$ for all $1 \leq l \leq N_c$, where $Q(q_l)$ is the quantization stepsize corresponding to q_l. Let $\hat{R}_{\text{src}} = \sum_{l=1}^{N_c} \hat{R}_{\text{src},l}$ be the total number of allocated bits. If $\hat{R}_{\text{src}} \leq R_{\text{src}}$, we stop the bit allocation procedure since the maximum bit rate under our scheme is still less than the available source rate. Then, we simply allocate bits to the lth class via $R_{\text{src},l} = \hat{R}_{\text{src},l}$ for all $1 \leq l \leq N_c$. Otherwise, if $\hat{R}_{\text{src}} > R_{\text{src}}$, we need to enlarge the quantization paramter by going to the next step.

2) R-D optimization.
   Based on the simplified rate and distortion models, we can compute

$$\lambda(l,k) = -\frac{\hat{D}_l(Q(q_l)) - \hat{D}_l(Q(k))}{\hat{R}_{\text{src},l}(Q(q_l)) - \hat{R}_{\text{src},l}(Q(k))}$$

   where $1 \leq l \leq N_c$ and $q_l + 1 \leq k \leq \text{QP}_1 + \Delta$, and choose l' and k' that give the minimum value of $\lambda(l,k)$. This means that l' and k' yield the least quality degradation under a unit rate reduction. Then, we can set $q_{l'} = k'$ and $\hat{R}_{\text{src},l'} = \hat{R}_{\text{src},l'}(Q(k'))$ and update $\hat{R}_{\text{src}} = \sum_{l=1}^{N_c} \hat{R}_{\text{src},l}$. We repeat this step until the following stopping criterion is met.

3) Stopping Criterion.
   If $\hat{R}_{\text{src}} = R_{\text{src}}$ or $q_l = \text{QP}_1 + \Delta$ for all $1 \leq l \leq N_c$, stop Step 2 and the final allocated bits to the lth class, $R_{\text{src},l}$ is set to $\hat{R}_{\text{src},l}$ for all $1 \leq l \leq N_c$.

The basic idea of the above bit allocation method can be simply stated as follows. First, it allocates the maximum bits to all classes using the smallest QPs. Usually, this rate will be higher than the available bit rate. Then, we have to reduce the allocated bits from some classes. In Step 2, we choose the class that gives the minimum distortion increase per bit rate reduction. This process repeats until the target bit is met. With MB-class bit allocation, the proposed rate control algorithm can be performed on MBs within the same class.

V. EXPERIMENTAL RESULTS

The proposed rate control algorithms with and without bit allocation are compared with the rate control algorithm in JM8.1a [33] for the H.264 baseline-profile encoder. In our experiments, all frames are encoded as P-frames except for the first I-frame. Generally speaking, the coded video quality largely depends on the frame-level bit allocation method. More bits are allocated to the I-frame and the preceding P-frames in many rate control algorithms to improve average quality according to the monotonicity property [6]. However, since the main purpose of our experiments is to show the advantage of the new encoder structure with the enhanced rate and distortion models, we adopt a simple frame-based bit allocation as stated below.

Because we use the rate control scheme in JM8.1a [17] as the benchmark, the first I-frame is encoded using the QP determined by the same rule as in JM8.1a for fair comparison. The target bits for subsequent P-frames are allocated by the following rule in our scheme. Let $N_r(n)$ and $T_r(n)$ be the number of remaining frames and the number of remaining bits before encoding the nth frame, respectively. Then, the target bits for the nth frame is simply determined as $T_r(n)/N_r(n)$. The Δ value is set to 3 and the initial values of model parameters α, β, and γ in (5), (12), and (15) are set to 6.0, 0.4 and 0.04, respectively, for the first P-frame in our algorithm. For all experiments, the buffer size is set to two times of the channel rate and the initial buffer fullness is set to half of the buffer size.

Several QCIF sequences are encoded with a single reference frames at 30 fps. As mentioned in Section IV, the MBs in the
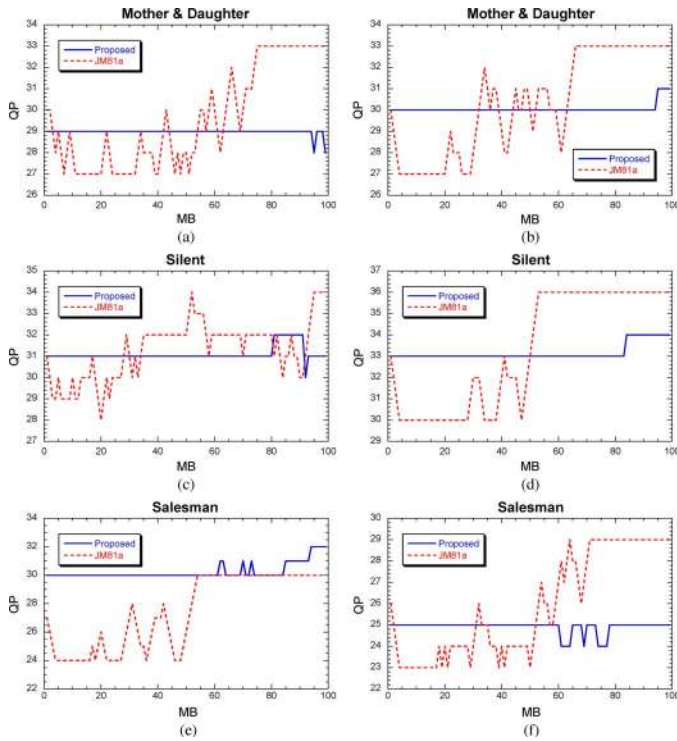
Fig. 8. Variation of QPs in a frame by the proposed rate control scheme without bit allocation (solid line) and the rate control in JM8.1a (dashed line). (a)–(b) 10th and 50th frames of "Mother & Daughter" at 48 kbps. (c)–(d) 10th and 50th frames of "Silent" at 64 kbps. (e)–(f) 10th and 50th frames of "Salesman" at 96 kbps.

Fig. 9. Distributions of $QP_1 - QP_2$ by the proposed MB layer rate control without bit allocation for the QCIF sequences. (a) "Foreman" and (b) "News" at 64 kbps.

Fig. 10. Performance comparison of the proposed rate control with bit allocation (solid line) and the rate control in JM8.1a (dashed line) for "Mother & daughter" at 48 kbps. (a) Number of allocated bits per frame. (b) PSNR value per frame.

same GOB are classified into the same class for bit allocation. Table IV shows the experimental results by three different rate control algorithms. Note that the coding gains in this table are all calculated with respect to the rate control scheme in JM81a. The results show the proposed algorithm with or without bit allocation can meet the target bit rate more accurately for all test sequences at all bit rates. The picture quality is also improved. While the average coding gain and the maximal coding gain are 0.25 and 0.39 dB, respectively, by the proposed rate control algorithm without bit allocation, they can go up to 0.60 and 1.02 dB, respectively, by the proposed rate control with bit allocation.

In simplified rate and distortion models, there are three parameters to be tuned: ω in the header rate model (5), p in the source rate model (12) and p in the distortion model (15). They are employed to simplify models given by (4), (11), and (14), respectively. These parameters as specified in Section III work well for all test sequences. In other words, they are robust with respect to test sequences shown in Table IV.

The improved coding gain by the proposed rate control scheme without bit allocation mainly comes from the saving of bits required to encode the difference of the QPs between two successive MBs. Fig. 8 illustrates the variation of QPs in a frame for several sequences at different bit rates. We see that even though each MB can have its own QP, the variation is small due to the accurate rate models. The difference between the largest value and the smallest value is at most 2 for most of the frames. As a result, fewer bits are required to encode the difference of them. On the contrary, the inaccurate source rate
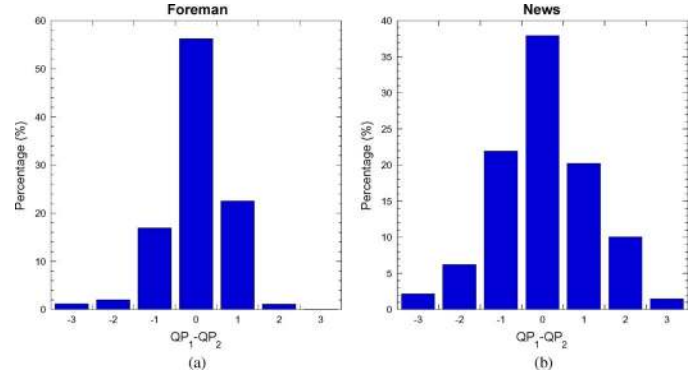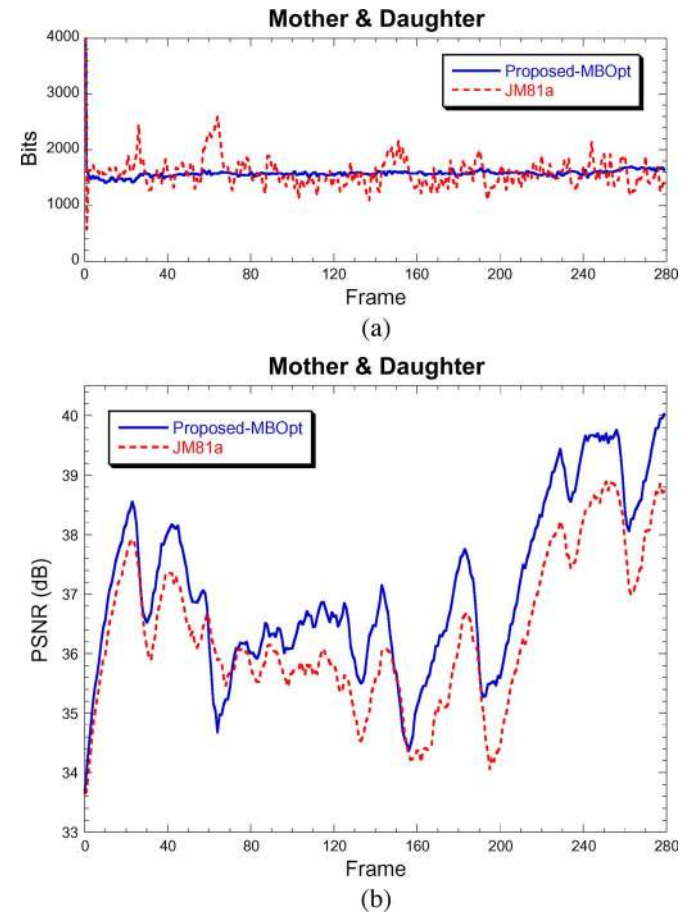
and the header bit estimation method in JM8.1a causes a large variation in QPs. Consequently, it demands more bits to encode their difference. It also causes significant quality fluctuations within a frame. As a evidence, we can easily conjecture from Fig. 8 that JM8.1a rate control algorithm produces higher quality MBs in the beginning part of a frame and poorer quality MBs in the ending part of a frame. Furthermore, by employing bit allocation, we can further improve the PSNR value by
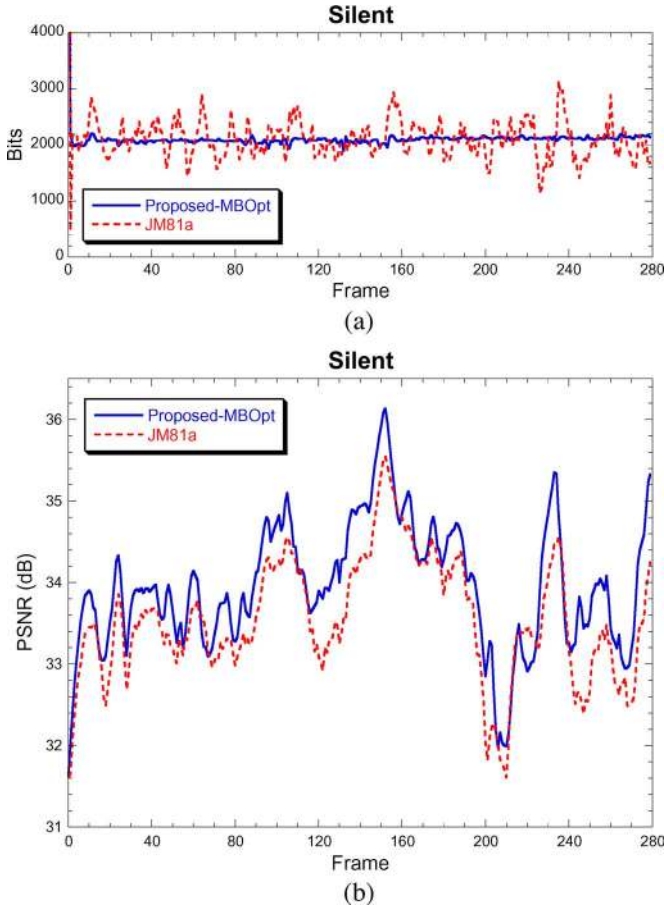
Fig. 11. Performance comparison of the proposed rate control with bit allocation (solid line) and the rate control in JM8.1a (dashed line) for "Silent" at 64 kbps. (a) Number of allocated bits per frame. (b) PSNR value per frame.

Fig. 12. Performance comparison of the proposed rate control with bit allocation (solid line) and the rate control in JM8.1a (dashed line) for "Salesman" at 96 kbps. (a) Number of allocated bits per frame. (b) PSNR value per frame.

0.35 dB on average over the one without bit allocation. The amount of improvement depends on the characteristics of video sequences. When MBs in a frame have more different R-D characteristics, the additional coding gain is larger.

Fig. 9 shows the distribution of $QP_1 - QP_2$ for the "Foreman" and "News" sequences, when they are coded at 64 kbps by the proposed rate control without bit allocation. In case of the "Foreman" sequence, the probability of $|QP_1 - QP_2| \leq 1$ is larger than 95% and the probability of $|QP_1 - QP_2| \geq 3$ is less than 2%. In case of the "News" sequence, these probabilities are larger than 80% and less than 4%, respectively. According to the discussion in Section II, the high probability of $|QP_1 - QP_2| \leq 1$ implies that the coding gain loss by the two-stage encoding is very small. The low probability of $|QP_1 - QP_2| \geq 3$ implies that 3 is a reasonable choice for the value of Δ. Furthermore, it is observed that the difference between these two QP values is small for most frames in a sequence whether the sequence is of high activity or not. The large difference between them (i.e., ≥ 3) is observed in the scene change frames. In the "News" sequences, the probability of $|QP_1 - QP_2| \geq 3$ is larger due to the background scene changes at the 90th, 150th, and 240th frames, and we can reduce the sudden quality change at these frames by setting $\Delta = 3$. Generally speaking, we conclude from Fig. 9
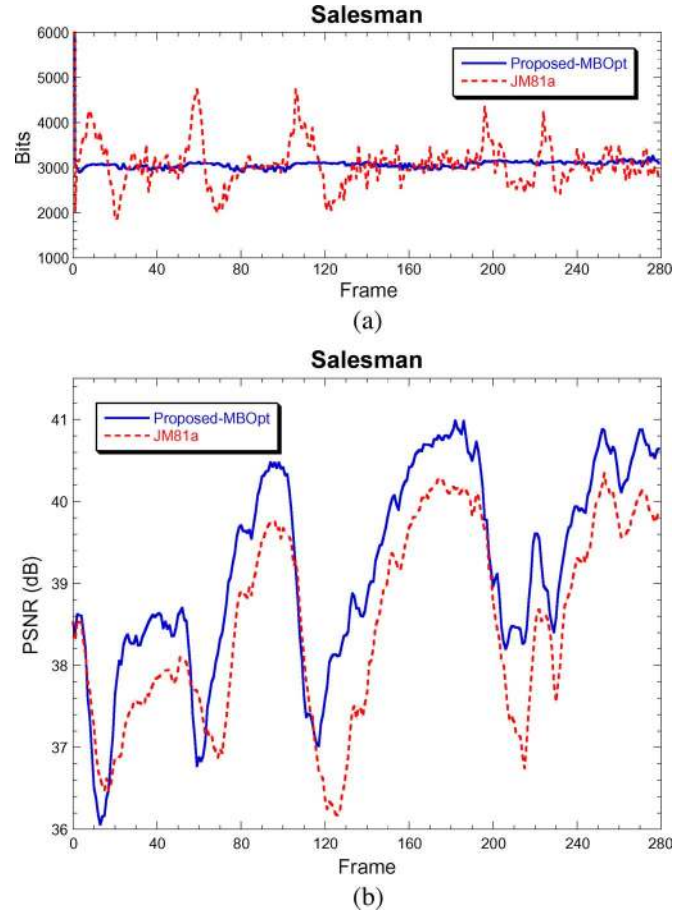
that the average QP_2 of the previous frame provides a good estimate of QP_1 of the current frame in this application with the baseline profile.

We show the number of allocated bits and the PSNR value as a function of the frame number by the proposed rate control algorithm with bit allocation and the rate control algorithm in JM8.1a for the three test sequences in Figs. 10–12. As shown in these figures, the target bit of each frame can be achieved very precisely by the proposed rate control algorithm. This is important when the decoder buffer size is limited, where frame skipping may occur often if the number of bits cannot be well controlled. Furthermore, the PSNR improvement of the proposed rate control scheme over the rate control algorithm in JM8.1a is evident.

VI. CONCLUSION AND FUTURE WORK

A novel model-based rate control algorithm based on the two-stage encoding was proposed for the H.264 baseline-profile encoder in this research. The inter-dependency of RDO and rate control is resolved by the two-stage encoding scheme at the cost of acceptable extra encoding complexity. An enhanced header rate model was established to estimate the number of header

bits more accurately due to the increased importance of header bits in H.264. Enhanced source rate and distortion models were also proposed based on coded 4×4 block identification. It was shown by experimental results that both rate and distortion can be well estimated by the proposed models. In addition, a MB-based bit allocation method was proposed to improve the picture quality. The proposed rate control algorithm can achieve the target bit rate of a frame more accurately with improved R-D performance as compared with the rate control algorithm in JM8.1a.

It is worthwhile to point out that we may need a more sophisticated estimation method for QP_2 to improve the performance for different video coding applications, for example, non-conversational applications that require a finite GOP structure with B-frames. Furthermore, the coding gain could be further improved if a more sophisticated classification method is employed instead of the GOB-based classification. These open issues are under our current study.

REFERENCES

[1] G. M. Schuster and A. K. Katsaggelos, "Fast and efficient model and quantizer selection in the rate distortion sense for H.263," in *Proc. SPIE, Conf. Vis. Commun. Image Process.*, Mar. 1996, pp. 784–795.

[2] D. Mukherjee and S. K. Mitra, "Combined mode selection and macroblock quantization step adaptation for the H.263 video encoder," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1997, pp. 26–29.

[3] J. Lee and D. W. Dickinson, "Rate-distortion optimized frame type selection for MPEG encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 3, pp. 501–510, Jun. 1997.

[4] S. Liu and C.-C. J. Kuo, "Joint temporal-spatial bit rate control for video coding with dependency," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 15–26, Jan. 2005.

[5] H. Song and C.-C. J. Kuo, "Rate control for low-bit-rate video via variable-encoding frame rates," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 4, pp. 512–521, Apr. 2001.

[6] K. Ramchandran, A. Ortega, and M. Vetterli, "Bit allocation for dependent quantization with applications to multiresolution and MPEG video coders," *IEEE Trans. Image Process.*, vol. 3, no. 9, pp. 533–545, Sep. 1993.

[7] A. Ortega, K. Ramchandran, and M. Vetterli, "Optimal trellis-based buffered compression and fast approximation," *IEEE Trans. Image Process.*, vol. 3, no. 1, pp. 26–40, Jan. 1994.

[8] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 1, pp. 12–20, Feb. 1996.

[9] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 446–459, Aug. 1998.

[10] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, Feb. 1999.

[11] P.-Y. Cheng, J. Li, and C.-C. J. Kuo, "Rate control for an embedded wavelet video coder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 3, pp. 696–702, Aug. 1997.

[12] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, ser. 11, vol. 12, pp. 970–982, Nov. 2002.

[13] Z. He and S. K. Mitra, "Optimum bit allocation and accurate rate control for video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840–849, Oct. 2002.

[14] I. Ahmad and J. Luo, "On using game theory for perceptually tuned rate control algorithm for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 2, pp. 202–208, Feb. 2006.

[15] 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 Syst. Video Technol.*, ser. 7, vol. 7, pp. 688–703, Jul. 2003.

[16] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 7, pp. 1–19, Jul. 2003.

[17] Z. G. Li, F. Pan, K. P. Lim, and S. Rahardja, "Adaptive rate control for H.264," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 745–748.

[18] P. Yin and J. Boyce, "A new rate control scheme for H.264 video coding," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 449–452.

[19] M. Jiang and N. Ling, "On enhancing H.264/AVC video rate control by PSNR-based frame complexity estimation," *IEEE Trans. Consum. Electron.*, vol. 51, no. 2, pp. 281–286, Feb. 2005.

[20] N. Wang and Y. He, "A new bit rate control strategy for H.264," in *Proc. IEEE Int. Conf. Inf. Commun. Signal Process.*, Dec. 2003, pp. 1370–1373.

[21] J. Xu and Y. He, "A novel rate control for H.264," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2004, pp. III-809–III-812.

[22] S. Kim and Y. S. Ho, "Rate control algorithm for H.264/AVC video coding standard based on rate-quantization model," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2004, pp. 165–168.

[23] 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.

[24] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2001, pp. 535–542.

[25] H. Chung, D. Romancho, and A. Ortega, "Fast long-term motion estimation for H.264 using multiresolution search," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2003, pp. 905–908.

[26] C. Kim, H.-H. Shih, and C.-C. J. Kuo, "Feature-based intra-prediction mode decision for H.264," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2004, pp. 769–772.

[27] J. L. Devore and N. R. Farnum, *Applied Statistics for Engineers and Scientists*. New York: Duxbury, 1999.

[28] A. Vetro, H. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 186–199, Feb. 1999.

[29] 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, Sep. 2000.

[30] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 7, pp. 599–603, Jul. 2003.

[31] H. M. Kim, "Adaptive rate control using nonlinear regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 5, pp. 432–439, May 2003.

[32] E. A. Riskin, "Optimal bit allocation via the generalized BFOS algorithm," *IEEE Trans. Inf. Theory*, vol. 37, no. 3, pp. 400–402, Mar. 1991.

[33] Joint Video Team, JVT Reference Software Encoder, version 8.1a. [Online]. Available: http://www.bs.hhi.de/seuhring/tml/download

Do-Kyoung Kwon received the M.S. and Ph.D. degrees from the University of Southern California (USC), Los Angeles, in 2002 and 2006, respectively, both in electrical engineering.

He was a Research Assistant in the Signal and Image Processing Institute at USC from August 2003 to December 2006. His research interests include a wide range of issues related with video compression and communication, including rate control, video streaming, and QoS provision over wired and wireless networks.

Mei-Yin Shen received the B.S. degree from the National Cheng-Kung University, Taiwan, R.O.C., in 1990, and the M.S. and Ph.D. degrees from the University of Southern California (USC), Los Angeles, in 1993 and 1999, respectively, all in electrical engineering.

Since 2002, she has been a Postdoctoral Researcher at USC. Her research interests covers a broad range of multimedia technologies, including image/video compression and processing, rate control, image/video mosaicing, and 3-D mesh compression.

C.-C. Jay Kuo (S'83–M'86–SM'92–F'99) received the B.S. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, R.O.C., in 1980 and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively.

He is the Director of the Signal and Image Processing Institute (SIPI) and Professor of Electrical Engineering, Computer Science, and Mathematics, University of Southern California (USC), Los Angeles. His research interests are in the areas of digital image/video analysis and modeling, multimedia data compression, communication and networking, and biological signal/image processing. He is the coauthor of approximately 120 journal papers, 680 conference papers, and seven books.

Dr. Kuo is a Fellow of SPIE and a member of the Association of Computing Machinery. He is Editor-in-Chief of the *Journal of Visual Communication and Image Representation* and Editor of the *Journal of Information Science and Engineering*, *LNCS Transactions on Data Hiding and Multimedia Security*, and the *EURASIP Journal of Applied Signal Processing*. He was on the Editorial Board of the *IEEE Signal Processing Magazine* in 2003–2004. He served as Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1995 to 1998, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 1995 to 1997, and the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING from 2001 to 2003. He was the recipient of the National Science Foundation Young Investigator Award (NYI) and the Presidential Faculty Fellow (PFF) Award in 1992 and 1993, respectively. He was the recipient of the Northrop Junior Faculty Research Award from the USC Viterbi School of Engineering in 1994. He received the Best Paper Award from the Multimedia Communication Technical Committee of the IEEE Communication Society in 2005. He was an IEEE Signal Processing Society Distinguished Lecturer in 2006. He is also Advisor to the Society of Motion Picture Television Engineers-USC Student Chapter.