Novel Rate Control Scheme for Low Delay Video Coding of HEVC

Wei Wu, Jiong Liu, and Lei Feng

In this paper, a novel rate control scheme for low delay video coding of High Efficiency Video Coding (HEVC) is proposed. The proposed scheme is developed by considering a new temporal prediction structure of HEVC. In the proposed scheme, the relationship between bit rate and quantization step is exploited firstly to formulate an accurate quadratic rate-quantization (R-Q) model. Secondly, a method of determining the quantization parameters (QPs) for the first frames within a group of pictures is proposed. Thirdly, an accurate frame-level bit allocation method is proposed for HEVC. Finally, based on the proposed R-Q model and the target bit allocated for the frame, the QPs are predicted for coding tree units by using rate-distortion (R-D) optimization. We compare our scheme against that of three other state-of-the-art rate control schemes. Experimental results show that the proposed rate control scheme can increase the Bjøntegaard delta peak signal-to-noise ratio by 0.65 dB and 0.09 dB on average compared with the JCTVC-I0094 and JCTVC-M0036 schemes, respectively, both of which have been implemented in an HEVC test model encoder; furthermore, the proposed scheme achieves a similar R-D performance to Wang's scheme, as well as obtaining the smallest bit rate mismatch error of all the schemes.

Keywords: High Efficiency Video Coding, rate control, low delay video coding, R-Q model, bit allocation.

Manuscript received Mar. 25, 2014; revised Apr. 8, 2015, accepted Sept. 10, 2015.

This work was supported in part by the National Natural Science Foundation of China under Grant 61471277, the Ningbo Natural Science Foundation under Grant 2015A610129, the 111 Project under Grant B08038, also supported by ISN State Key Laboratory.

Wei Wu (corresponding author, wwu@xidian.edu.cn), Jiong Liu (liujiong@mail.xidian.edu.cn), and Lei Feng (fenglei@mail.xidian.edu.cn) are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China.

I. Introduction

To meet the rapid increasing demand of video content, a new video coding standard called "High Efficiency Video Coding (HEVC)" [1] was established by the Joint Collaborative Team on Video Coding (JCT-VC) in January 2013. In contrast to previous video coding standards, HEVC not only employs a flexible quad-tree coding block partitioning structure but also improved intra-prediction and coding and adaptive motion parameter prediction and coding, both of which significantly improve the coding efficiency. Apart from the coding efficiency, rate control is also an important issue in video services [2], particularly for real-time video communications. The objective of rate control is to achieve good video quality by adjusting encoding parameters to prevent a buffer from overflowing and underflowing under the constraint of transmission bandwidth.

Although rate control is not a normative part of any video coding standard, every video coding standard has its own recommendation on rate control for informative purposes [3], such as reference model [4] for H.261, adaptive quantization algorithm [5] for MPEG-1, Test Model 5 [6] for MPEG-2, Test Model Near-term 8 [7] for H.263, Verification Model 18 [8] for MPEG-4, and JVT-W042 [9], developed based on JVT-G012 [10], for H.264/AVC. In addition to JVT-W042 adopted in Joint Model, many rate control schemes [11]–[15] have also been designed for H.264/AVC.

In HEVC, a flexible quad-tree coding block partitioning structure is adopted that enables the efficient use of multiple sizes of coding units (CUs), prediction units (PUs), and transform units (TUs). The CU, PU, and TU define the regions sharing the same prediction mode, the same prediction information, and the same transformation. These new features

differentiate the rate control method for HEVC from those adopted by the previous video coding standard.

For HEVC, several rate control schemes have been proposed in [16]–[18] and [19]. A rate control scheme for HEVC and its improvement have been proposed in [16] and [19], respectively, and are founded on a pixel-based unified rate quantization model. This scheme allocates the target bit for a group of picture (GOP), a frame, and a coding tree unit (CTU), respectively, and then predicts the quantization parameter (QP) value for a CTU. However, the achieved rate-distortion (R-D) performance is not perfect. In [17], a rate control scheme based on an R-λ model is proposed for HEVC. The scheme adopts a bit allocation method to estimate the target bit for a CTU. It then uses the R-λ model and the target bit to compute a value of λ, and finally determines the QP value according to the relationship between QP and λ [20]. In [18] a ρ-domain Rate-GOP-based frame-level rate control scheme is proposed for HEVC. In this scheme, a reference picture set–based hierarchical rate control structure is designed, and then the distortion and rate of the coding frame are represented by the distortion and rate of its reference frame; finally, based on the models, the QPs of the frames are predicted. In the above three schemes, three rate models are used, including q-domain, λ-domain, and ρ-domain, respectively. The λ-domain and ρ-domain rate control schemes in [17] and [18] have better R-D performances than the q-domain scheme in [19].

In this paper, a new q-domain rate control scheme is proposed for HEVC. The proposed scheme is developed with consideration for a new temporal prediction structure adopted into HEVC. The main contributions of the proposed scheme can be summarized as follows. Firstly, the relationship between bit rate and quantization step (QS) for frames is exploited to propose an accurate rate model for HEVC. Secondly, according to the temporal prediction structure in HEVC, a method of determining the QPs for the first frames within GOPs is proposed. Thirdly, an accurate frame-level bit allocation for HEVC is proposed. Finally, based on the proposed rate model and bit allocation, a method for optimizing the R-D performance is used to predict QPs for CTUs.

An HEVC video encoder works with three kinds of temporal prediction structure — intra-only configuration, low delay configuration, and random access configuration [21] — resulting in different rate controls for different structures. Of the three aforementioned temporal prediction structures, low delay configuration is designed for low-delay video coding, which can be widely used in real-time video communications. Thus, in this paper, a novel rate control scheme is developed particularly for the low delay configuration.

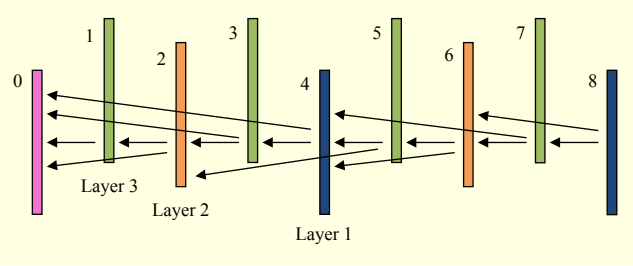

Fig. 1. Graphical presentation of low delay configuration.

II. Rate-Quantization (R-Q) Model in HEVC

1. Low Delay Configuration

Figure 1 shows a graphical presentation of the low delay configuration [21]. The number associated with each frame represents the encoding order. The frame with an index of 0 is an instantaneous decoding refresh (IDR) frame; all subsequent frames are encoded to be interframes. Every group of four successive frames following on from the IDR constitutes a GOP. The interframes are divided into three layers as shown in the figure; in total, there is one frame in layer 1, one frame in layer 2, and two frames in layer 3 within every GOP.

2. Proposed R-Q Model

In video coding, QS is used to compress the discrete cosine transform coefficients of prediction residual. Upon compression, a corresponding texture bit is obtained; the number associated with the texture bit changes with the QS value. Usually, the larger the QS value is, the lower the number associated with the texture bit becomes. Besides a texture bit, a non-texture bit is also included in the total bit required for encoding a current CTU. An R-Q model is often adopted in rate control to represent the relationship between the texture bit or the total bit and QS or QP. Before the proposed R-Q model is introduced, the difference between QP and QS is described below. In scalar quantization, QS is the actual step size used by a quantizer, while QP indicates the index of QS. There is a nonlinear relationship between QP and QS in HEVC.

$$\mathrm{QS} = 2^{\mathrm{QP}/6} \times v\,(\mathrm{QP} \bmod 6), \tag{1}$$

where $v(0) = 0.625$, $v(1) = 0.703$, $v(2) = 0.797$, $v(3) = 0.891$, $v(4) = 1.000$, and $v(5) = 1.125$.

An HEVC adopts new coding tools and a new temporal prediction structure in the low delay configuration, which affects the relationship between bit rate and QS. In Fig. 2, the relationships among the total bit; the encoding complexity; the width and height of the area containing samples; and QS are illustrated for the frame in level 1, the frame in level 2, and the

ETRI Journal, Volume 38, Number 1, February 2016
http://dx.doi.org/10.4218/etrij.16.0114.0366

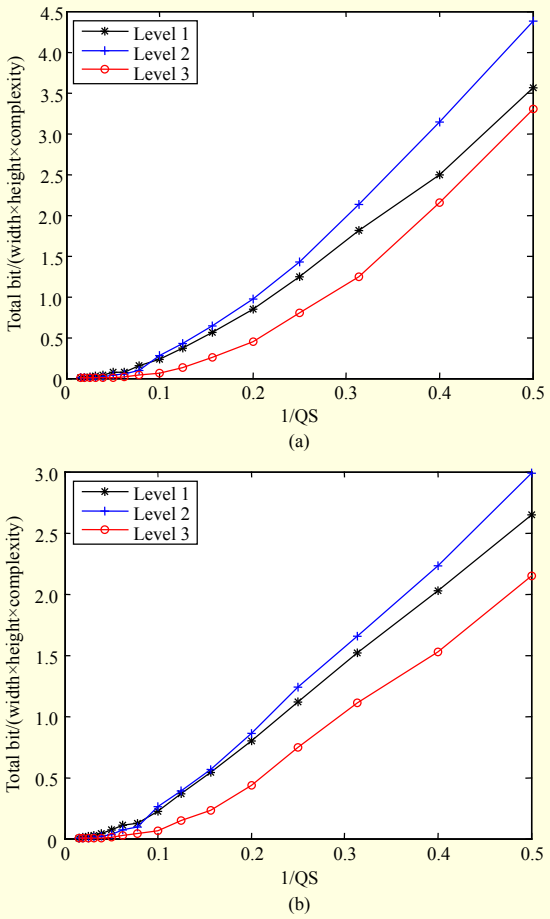

Fig. 2. Relationships for frames in levels 1 and 2, and first frame in level 3 within same GOP: (a) second GOP in "BQSquare" and (b) third GOP in "BlowingBubbles."

first frame in level 3 within the same GOP, respectively. From the figure, it can be seen that the relationship can be represented by a quadratic model and that the frames in different levels have different relationships.

Therefore, in this paper, an accurate R-Q model is proposed for HEVC as follows:

$$\frac{T_{\text{total},l}}{W \cdot H} = \frac{a_l \cdot m_l}{\text{QS}} + \frac{b_l \cdot m_l}{\text{QS}^2}, \tag{2}$$

where l is the layer index and $T_{\text{total},l}$, m_l, a_l, and b_l are the total bit; the mean absolute difference between the original block and the prediction block (used to indicate the encoding complexity); and two model parameters for a CTU in the layer l, respectively. Moreover, W and H are the actual width and height of the area containing samples in the CTU, respectively.

In previous literatures, a number of R-Q models have been developed based on observations and analyses. In [22], the source statistics are assumed to be Laplacian distributed and a well-known quadratic R-Q model is proposed as

$$T_{\text{texture}} = \frac{a \cdot m}{\text{QS}} + \frac{b \cdot m}{\text{QS}^2}, \tag{3}$$

where T_{texture} represents the texture bit; m indicates the complexity; and a and b are the model parameters. Another form of quadratic R-Q model [19] is proposed as

$$\frac{T_{\text{total}}}{N_{\text{pixel}}} = \frac{a \cdot m}{\text{QS}} + \frac{b \cdot m}{\text{QS}^2}, \tag{4}$$

where N_{pixel} is the number of pixels. Besides the two quadratic models, a linear relationship between T and 1/QS is denoted in [23] and [24] as

$$T_{\text{total}} = \frac{a \cdot m}{\text{QS}} + b. \tag{5}$$

Based on an analyses of some experimental results, both a quadratic model and a linear R-Q model using QP are proposed in [25] as

$$T_{\text{total}} = \frac{a \cdot m}{\text{QP}} + \frac{b \cdot m}{\text{QP}^2}, \tag{6}$$

$$T_{\text{total}} = \frac{a \cdot m}{\text{QP}} + b. \tag{7}$$

To evaluate the performance of the proposed R-Q model, extensive experiments are performed in this paper. Essentially, four test video sequences — BQSquare, RaceHorses, BasketballDrill, and BlowingBubbles — are encoded with QP values of 10 to 40 corresponding to QS values of 2 to 64. The texture bit, the non-texture bit, and the actual complexity values are obtained.

The accuracies of the R-Q models described above are specified by an F-statistic [22], which is a measurement for aptness of the fit and is expressed as

$$F = \frac{\sum_i (Y_i - \bar{Y})^2}{k-1} \Big/ \frac{\sum_i (Y_i - \hat{Y}_i)^2}{n-k}, \tag{8}$$

where Y_i corresponds to the ith data point, $\bar{Y}$ is the mean of all data points, $\hat{Y}_i$ is the estimated value of the ith data point, n is the number of data points, and k is the number of model parameters. The larger the F ratio is, the more accurate the model is. The F ratio results of all the six models are shown in Table 1.

From the results in the table, for all the four sequences, model (2) has the largest F-ratio values among the experimental models. Based on this observation, of the six R-Q models, model (2) is thus the more suitable for HEVC.

In (2), m_l is predicted as follows:

$$m_l = c_l \times m_l^{\text{actu}} + d_l, \tag{9}$$

Table 1. F ratio values of six R-Q models.

	(2)	(3)	(4)	(5)	(6)	(7)
BQSquare	102.3	78.3	84.6	96.5	42.3	41.3
RaceHorses	88.2	55.3	74.0	84.0	15.4	71.0
BasketballDrill	441.5	331.8	405.2	16.9	145.8	15.6
BlowingBubbles	67.5	55.3	62.3	65.0	36.3	34.0
Average	174.9	130.2	156.5	65.6	59.9	40.5

where m_l^{actu} is the actual complexity of the collocated CTU in the previous frame in the same layer, and c_l and d_l are two model parameters.

III. Proposed Rate Control Scheme

1. Method of Determining QPs for First Frames within GOPs

Many GOPs exist in a video sequence, and a GOP consists of several frames. In the traditional rate control schemes [10], [19], the QPs of the first frames within GOPs are usually defined by deterministic values, not rate control; thus, they may not be accurate enough to obtain good R-D performances.

In the low delay configuration, only the first frame within the first GOP in a video sequence is encoded as an IDR frame, and the other first frames within GOPs are inter-coded frames for which rate control can be used. Therefore, to achieve accurate rate control, in the proposed rate control scheme, the QPs of the first frames within all the GOPs except the first and second are computed using rate control. For the first frames within the first and second GOPs, to obtain the coding information used to predict the information for the subsequent frames, their QP values are set to be defined by deterministic values, which is the same as in previous rate control schemes.

2. GOP-Level Rate Control

The total number of bits in a GOP is managed in the GOP-level rate control. When the jth frame within the ith GOP is encoded, the bit used for the remaining frames in the GOP is calculated as

$$B_{i,j}=\begin{cases}\dfrac{R_{i,j}}{f}\times N_{\text{GOP}}-V_{i,j} & j=1,\\ B_{i,j-1}+\dfrac{R_{i,j}-R_{i,j-1}}{f}\times(N_{\text{GOP}}-j+1)-b_{i,j-1} & j=2,3,\ldots,N_{\text{GOP}},\end{cases} \tag{10}$$

where $R_{i,j}$ is the available bit rate and f represents the predefined frame rate; N_{GOP} indicates the number of frames within a GOP; $V_{i,j}$ is the virtual buffer occupancy computed by using (11) and (12); and $b_{i,j-1}$ is the generated bit of the (j–1)th frame. Thus, $V_{i,j}$ is computed as

$$V_{i,1}=\begin{cases}0 & i=2,\\ V_{i-1,N_{\text{GOP}}}+b_{i-1,N_{\text{GOP}}}-\dfrac{R_{i-1,N_{\text{GOP}}}}{f}-A_{i-1,N_{\text{GOP}}} & \text{otherwise,}\end{cases} \tag{11}$$

$$V_{i,j}=V_{i,j-1}+b_{i,j-1}-\frac{R_{i,j-1}}{f}-A_{i,j-1}\quad j=2,3,\ldots,N_{\text{GOP}}, \tag{12}$$

where $A_{i,j}$ represents the adjustment bit for the jth frame within the ith GOP, which is also considered in [19]. The generated bit of the IDR frame is usually much larger than the average available bit used for encoding one frame due to the constrained channel bandwidth, which is represented by $R_{i,j}/f$. When the IDR frame has been encoded, the excess bit should be offset. If the generated bit of an interframe is less than the average available bit used for encoding one frame, then a proportion of the difference between the generated bit and the average available bit is adopted to be the adjustment bit for offsetting the excess bit. The adjustment bit $A_{i,j}$ is defined as follows:

$$A_{i,j}=\begin{cases}\eta\times O_{i,j} & \text{if } j\neq 1, I_{i,j-1}>0, O_{i,j}<0, (I_{i,j-1}+\eta\times O_{i,j})\geq 0,\\ -I_{i,j-1} & \text{if } j\neq 1, I_{i,j-1}>0, O_{i,j}<0, (I_{i,j-1}+\eta\times O_{i,j})<0,\\ \eta\times O_{i,j} & \text{if } j\neq 1, I_{i,j-1}<0, O_{i,j}>0, (I_{i,j-1}+\eta\times O_{i,j})\leq 0,\\ -I_{i,j-1} & \text{if } j\neq 1, I_{i,j-1}<0, O_{i,j}>0, (I_{i,j-1}+\eta\times O_{i,j})>0,\\ \eta\times O_{i,1} & \text{if } I_{i-1,N_{\text{GOP}}}>0, O_{i,1}<0, (I_{i-1,N_{\text{GOP}}}+\eta\times O_{i,1})\geq 0,\\ -I_{i-1,N_{\text{GOP}}} & \text{if } I_{i-1,N_{\text{GOP}}}>0, O_{i,1}<0, (I_{i-1,N_{\text{GOP}}}+\eta\times O_{i,1})<0,\\ \eta\times O_{i,1} & \text{if } I_{i-1,N_{\text{GOP}}}<0, O_{i,1}>0, (I_{i-1,N_{\text{GOP}}}+\eta\times O_{i,1})\leq 0,\\ -I_{i-1,N_{\text{GOP}}} & \text{if } I_{i-1,N_{\text{GOP}}}<0, O_{i,1}>0, (I_{i-1,N_{\text{GOP}}}+\eta\times O_{i,1})>0,\\ 0 & \text{otherwise,}\end{cases} \tag{13}$$

$$I_{i,j}=\begin{cases}O_{i,j} & i=1,\ j=1,\\ I_{i-1,N_{\text{GOP}}}+A_{i,j} & i=2,3,\ldots,\ j=1,\\ I_{i,j-1}+A_{i,j} & \text{otherwise,}\end{cases} \tag{14}$$

$$O_{i,j}=b_{i,j}-\frac{R_{i,j}}{f}, \tag{15}$$

where $I_{i,j}$ represents the total bit needed to offset the excess bit, η is a parameter, and $O_{i,j}$ indicates the difference between the generated bit and the average available bit.

3. Frame-Level Bit Allocation

The scheme in [19] uses the target bit budgets based on the bit used for the remaining frames and based on the target buffer level to propose a frame-level bit allocation. However, the layers in the low delay configuration are not taken into account

ETRI Journal, Volume 38, Number 1, February 2016
http://dx.doi.org/10.4218/etrij.16.0114.0366

in [19]. In this subsection, an accurate bit allocation method for the low delay configuration is proposed considering the layers.

There are three layers of interframes in the low delay configuration. For the frames in different layers, even if they are encoded using an equivalent QS value, then the average generated bit may be different. The target bit budget based on the bit used for the remaining frames is computed as

$$\hat{T}_{i,j} = \frac{\overline{W}_{l_{\text{curr}}, p_{l_{\text{curr}}}-1} \times B_{i,j}}{\sum_{l=1}^{m} \overline{W}_{l,p_l-1} \times N_{l,r,i}}, \tag{16}$$

where $\overline{W}_{l,p_l-1}$ indicates the average weighting factor of the $(p_l - 1)$th frame in the lth layer because the average weighting factors of the frames in different layers may be different; l_{curr} is the layer index of the jth frame within the ith GOP, which can be also represented as the $p_{l_{\text{curr}}}$ th frame in the l_{curr}th layer; $N_{l,r,i}$ is the number of remaining frames in the lth layer within the ith GOP; and m is the number of layers. The average weighting factor $\overline{W}_{l_{\text{curr}}, p_{l_{\text{curr}}}}$ is computed as

$$\overline{W}_{l_{\text{curr}}, p_{l_{\text{curr}}}} = \frac{\text{QP}_{i,j} \times b_{i,j}}{8} + \frac{7 \times \overline{W}_{l_{\text{curr}}, p_{l_{\text{curr}}}-1}}{8}. \tag{17}$$

For the jth frame within the ith GOP, the target buffer level is determined using

$$S_{i,j} = \begin{cases} V_{i,j} & j = 1, \\ S_{i,j-1} - \frac{S_{i,1}}{N_{\text{GOP}} - 1} + \phi_{l_{\text{curr}}, p_{l_{\text{curr}}}-1} \times \frac{R_{i,j}}{f} & j = 2,3,\ldots,N_{\text{GOP}}, \end{cases} \tag{18}$$

$$\phi_{l_{\text{curr}}, p_{l_{\text{curr}}}-1} = \frac{\overline{W}_{l_{\text{curr}}, p_{l_{\text{curr}}}-1} \times N_{\text{GOP}}}{\sum_{l=1}^{m} \overline{W}_{l,p_l-1} \times N_{l,i}} - 1, \tag{19}$$

where $N_{l,i}$ denotes the number of frames in the lth layer within the ith GOP, and $S_{i,1}/(N_{\text{GOP}} - 1)$ indicates that the target buffer level is expected to be zero after all frames within the GOP have been encoded. Therefore, the target bit budget based on the target buffer level is calculated as follows:

$$\tilde{T}_{i,j} = \frac{\overline{W}_{l_{\text{curr}}, p_{l_{\text{curr}}}-1} \times N_{\text{GOP}}}{\sum_{l=1}^{m} \overline{W}_{l,p_l-1} \times N_{l,i}} \times \frac{R_{i,j}}{f} + \gamma \times (S_{i,j} - V_{i,j}), \tag{20}$$

where γ is a parameter.

After $\hat{T}_{i,j}$ and $\tilde{T}_{i,j}$ are obtained, the target bit allocated for the jth frame within the ith GOP is obtained as

$$T_{i,j} = \beta \times \hat{T}_{i,j} + (1 - \beta) \times \tilde{T}_{i,j}, \tag{21}$$

where β is a parameter. The final target bit is then restricted by using $U_{i,j}$ and $L_{i,j}$.

$$T_{i,j} = \min\left(U_{i,j}, \max\left(L_{i,j}, T_{i,j}\right)\right), \tag{22}$$

$$U_{i,j} = \begin{cases} R_{i,j} \times \varpi - V_{i,j} & j = 1, \\ U_{i,j-1} - b_{i,j-1} + \frac{R_{i,j-1}}{f} & \text{otherwise}, \end{cases} \tag{23}$$

$$L_{i,j} = \begin{cases} \frac{R_{i,j}}{f} - V_{i,j} - \frac{I_{i-1,N_{\text{GOP}}}}{f} & j = 1, \\ L_{i,j-1} - b_{i,j-1} + \frac{R_{i,j-1}}{f} & \text{otherwise}, \end{cases} \tag{24}$$

where ϖ is a parameter.

4. CTU-Level QP Prediction

After the proposed R-Q model and the proposed bit allocation method are described, an efficient QP prediction method is necessary to determine the QP value for a CTU. In this paper, a rate distortion–optimized QP prediction method is proposed.

For the first CTU in the jth frame, its QP is computed as

$$\text{QP}_{i,j,1} = \begin{cases} \overline{\text{QP}}_{i,j-1} & i = 3,4,\ldots, j \neq 1, \\ \overline{\text{QP}}_{i-1,N_{\text{GOP}}} & i = 3,4,\ldots, j = 1, \end{cases} \tag{25}$$

where $\text{QP}_{i,j,1}$ is the QP value for the first CTU in the jth frame within the ith GOP, and $\overline{\text{QP}}_{i,j}$ is the average QP value for the jth frame within the ith GOP.

In [15], the sum of the inverses of the distortions of all the macroblocks in a frame is maximized to compute the QPs for the macroblocks. In this paper, a similar method is used to predict the QS values for the CTUs. When the kth CTU in the jth frame is encoded, its QS value is found as follows:

$$\text{maximize} \quad \frac{1}{N_{\text{CTU}} - k + 1} \sum_{g=k}^{N_{\text{CTU}}} (D_{i,j,g})^{-1}, \tag{26}$$

$$\text{s.t.} \quad \sum_{g=k}^{N_{\text{CTU}}} T_{i,j,g} - T_{r,i,j} = 0, \tag{27}$$

where N_{CTU} is the total number of CTUs in a frame; $D_{i,j,g}$ and $T_{i,j,g}$ are the distortion of and the bit required for the gth CTU in the jth frame within the ith GOP, respectively; and $T_{r,i,j}$ is the target bit for the remaining CTUs.

In this paper, a linear D-Q model is used to represent the relationship between the distortion and the QS, which is described as follows:

$$D_l = \rho_l \times \text{QS}, \tag{28}$$

where D_l is the distortion of a CTU in the lth layer, and ρ_l is a model parameter.

Substituting the proposed R-Q model and the linear D-Q model into (26) and (27), respectively, (26) and (27) become

$$\text{maximize} \quad \frac{1}{N_{\text{CTU}}-k+1}\sum_{g=k}^{N_{\text{CTU}}}\left(\rho_{l_{\text{curr}},p_{l_{\text{curr}}},g}\times \text{QS}_{i,j,g}\right)^{-1}, \tag{29}$$

s.t.

$$\sum_{g=k}^{N_{\text{CTU}}}\left(W_g\cdot H_g\cdot m_{l_{\text{curr}},p_{l_{\text{curr}}},g}\cdot\left(\frac{a_{l_{\text{curr}},p_{l_{\text{curr}}},g}}{\text{QS}_{i,j,g}}+\frac{b_{l_{\text{curr}},p_{l_{\text{curr}}},g}}{\text{QS}_{i,j,g}^2}\right)\right)-T_{r,i,j}=0. \tag{30}$$

The QS for the kth CTU in the jth frame is computed by using the Lagrange multiplier method as follows:

$$\text{QS}_{i,j,k}=-\frac{a_{l_{\text{curr}},p_{l_{\text{curr}}},k}}{2b_{l_{\text{curr}},p_{l_{\text{curr}}},k}}+\frac{\rho^{-1}_{l_{\text{curr}},p_{l_{\text{curr}}},k}}{b_{l_{\text{curr}},p_{l_{\text{curr}}},k}\cdot m_{l_{\text{curr}},p_{l_{\text{curr}}},k}\cdot W_k\cdot H_k}\times\sqrt{\frac{T_{r,i,j}+\sum_{g=k}^{N_{\text{CTU}}}\frac{a^2_{l_{\text{curr}},p_{l_{\text{curr}}},g}\cdot m_{l_{\text{curr}},p_{l_{\text{curr}}},g}\cdot W_g\cdot H_g}{4b_{l_{\text{curr}},p_{l_{\text{curr}}},g}}}{\sum_{g=k}^{N_{\text{CTU}}}\frac{\rho^{-2}_{l_{\text{curr}},p_{l_{\text{curr}}},g}}{b_{l_{\text{curr}},p_{l_{\text{curr}}},g}\cdot m_{l_{\text{curr}},p_{l_{\text{curr}}},g}\cdot W_g\cdot H_g}}}. \tag{31}$$

When the kth CTU in the jth frame is encoded, the parameters $a_{l_{\text{curr}},p_{l_{\text{curr}}},g}$, $b_{l_{\text{curr}},p_{l_{\text{curr}}},g}$, and $\rho_{l_{\text{curr}},p_{l_{\text{curr}}},g}$ are unavailable for $k<g\le N_{\text{CTU}}$. Therefore, $a_{l_{\text{curr}},p_{l_{\text{curr}}},k}$, $b_{l_{\text{curr}},p_{l_{\text{curr}}},k}$, and $\rho_{l_{\text{curr}},p_{l_{\text{curr}}},k}$ are used to approximate the corresponding parameters with g satisfying $k<g\le N_{\text{CTU}}$, respectively. Then, (31) becomes

$$\text{QS}_{i,j,k}=-\frac{a_{l_{\text{curr}},p_{l_{\text{curr}}},k}}{2b_{l_{\text{curr}},p_{l_{\text{curr}}},k}}+\frac{1}{m_{l_{\text{curr}},p_{l_{\text{curr}}},k}\cdot W_k\cdot H_k}\times\sqrt{\frac{T_{r,i,j}+\frac{a^2_{l_{\text{curr}},p_{l_{\text{curr}}},k}}{4b_{l_{\text{curr}},p_{l_{\text{curr}}},k}}\left(\sum_{g=k}^{N_{\text{CTU}}}m_{l_{\text{curr}},p_{l_{\text{curr}}},g}\cdot W_g\cdot H_g\right)}{b_{l_{\text{curr}},p_{l_{\text{curr}}},k}\sum_{g=k}^{N_{\text{CTU}}}\left(m_{l_{\text{curr}},p_{l_{\text{curr}}},g}\cdot W_g\cdot H_g\right)^{-1}}}. \tag{32}$$

5. Steps of Proposed Rate Control Scheme

The proposed rate control scheme is summarized in Fig. 3. The detailed steps of the proposed scheme are described as follows:

1) The QP values for the first frames within the first and second GOPs are set to be QP_{init} and QP_{init}+3, respectively, where QP_{init} is the initial QP. The QP values of the other frames within the second GOP are set to be QP_{init}+2, QP_{init}+3, and QP_{init}+1, respectively. Encode all five frames using these QP values. Obtain the generated bits, the buffer occupancies, and the average weighting factors of the frames. Let $i=3, j=1$, and go to step 2).

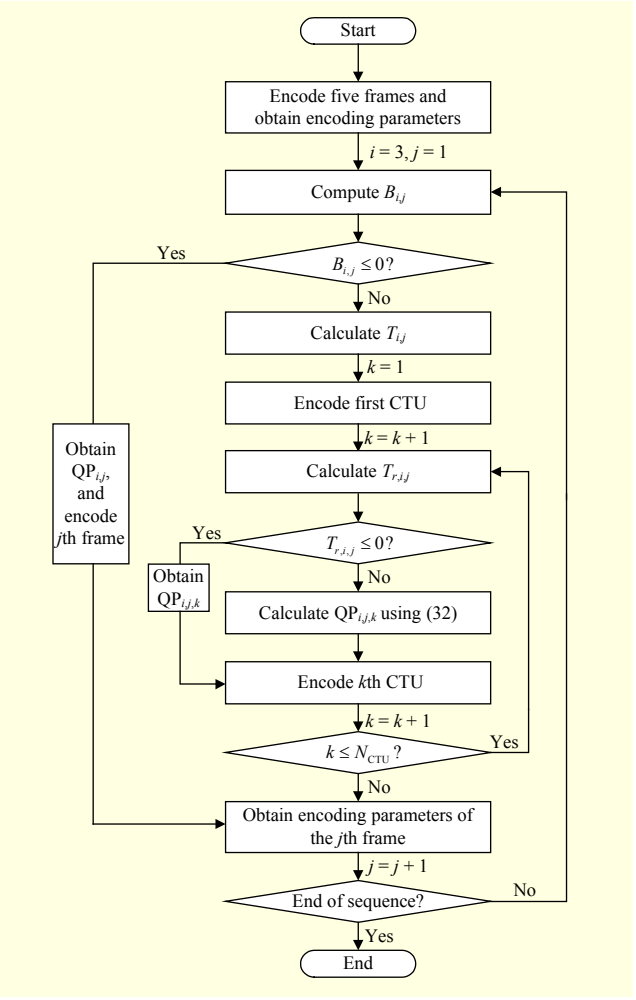

Fig. 3. Flowchart of proposed rate control scheme.

2) Compute the virtual buffer occupancies $V_{i,j}$, and the bits $B_{i,j}$ for the remaining frames within the ith GOP. If $B_{i,j}\le 0$, then the value of $\text{QP}_{i,j}$ is set to be $\overline{\text{QP}}_{i,j-1}+2$ and $\text{QP}_{i,j}$ is further bounded by $\text{QP}_{i,j}=\max\{1,\min\{51,\text{QP}_{i,j}\}\}$. Then, the CTUs in the jth frame are encoded by using $\text{QP}_{i,j}$. Then go to step 8). Otherwise, go to step 3).
3) Calculate $\hat{T}_{i,j}$ by using (16). Estimate $\tilde{T}_{i,j}$ and compute $T_{i,j}$ for the jth frame. Let $k=1$ and go to step 4).
4) Obtain $\text{QP}_{i,j,1}$ according to (25) and then encode the CTU. The number of the generated bit for the CTU is recorded. Let $k=k+1$, and go to step 5).
5) Calculate the target bits $T_{r,i,j}$ for the remaining CTUs in the jth frame. If $T_{r,i,j}\le 0$, then let $\text{QP}_{i,j,k}=\overline{\text{QP}}_{i,j-1}+2$ and bound $\text{QP}_{i,j,k}$ by using $\text{QP}_{i,j,k}=\max\{1,\min\{51,\text{QP}_{i,j,k}\}\}$, and then go to step 7). Otherwise, go to step 6).
6) Update parameters $a_{l_{\text{curr}},p_{l_{\text{curr}}},k}$, $b_{l_{\text{curr}},p_{l_{\text{curr}}},k}$, $c_{l_{\text{curr}},p_{l_{\text{curr}}},k}$, and $d_{l_{\text{curr}},p_{l_{\text{curr}}},k}$ according to the actual encoding data of the previous CTUs in the same layer. Predict $m_{l_{\text{curr}},p_{l_{\text{curr}}},g}$

($k \le g \le N_{CTU}$) for the remaining CTUs. If the value of the expression that lies beneath the square root symbol in (32) is negative, then let $QP_{i,j,k} = \overline{QP}_{i,j-1} - 1$. Otherwise, calculate $QS_{i,j,k}$ from (32) and convert $QS_{i,j,k}$ into $QP_{i,j,k}$. Bound $QP_{i,j,k}$ by using $QP_{i,j,k} = \max\{\overline{QP}_{i,j-1} - 2, \min\{\overline{QP}_{i,j-1} + 2, QP_{i,j,k}\}\}$ and $QP_{i,j,k} = \max\{1, \min\{51, QP_{i,j,k}\}\}$. Go to step 7).

7) Encode the kth CTU using $QP_{i,j,k}$. Then, obtain the generated bit and the actual complexity value of the CTU. Let $k = k + 1$. If $k \le N_{CTU}$, then go back to step 5) and encode the next CTU. Otherwise, go to step 8).
8) Obtain the average QP value, the generated bit, the buffer occupancy, and the average weighting factor of the jth frame. Then, encode the next frame until the last frame in the video sequence.

IV. Experimental Results

The performances of the proposed rate control scheme for HEVC are evaluated in this section. The experiment is implemented on an HEVC test model encoder HM-13.0. To compare the proposed scheme with the three state-of-the-art rate control schemes in [17], [18], and [19], all the video sequences in classes B, C, D, and E are tested with low delay configuration. In the three schemes, JCTVC-I0094 [19] and JCTVC-M0036 [17] have been adopted by JCTVC and implemented in HM. The tested sequences and testing configurations are detailed in JCTVC-I1100 [26]. In the experiment, QP_{init} is set to be 32, and the low complexity setting is used. The parameters η, γ, β, and ϖ are set to be 0.2, 0.25, 0.9, and 0.9, respectively.

The R-D performance is evaluated in terms of the Bjøntegaard delta (BD-PSNR) [27], which is used to represent the average and bit rate differences. A positive value for BD-PSNR indicates that the corresponding scheme achieves better R-D performance.

The BD-PSNR values obtained for all the video sequences in each class are averaged, and the results are shown in Tables 2 and 3. From the results in Table 2, it can be seen that for each of the four classes the proposed scheme can achieve better R-D performances than JCTVC-I0094 and JCTVC-M0036, and the average BD-PSNR values between the proposed scheme and JCTVC-I0094 and between the proposed scheme and JCTVC-M0036 are 0.69 dB and 0.10 dB, respectively. Compared to Wang's scheme [18], the proposed scheme can achieve better R-D performances for classes B and E, and the average BD-PSNR of all the classes is 0.02 dB, which shows that the R-D performance of the proposed scheme is similar to that of Wang's scheme. Furthermore, the same conclusions can be observed from the results in Table 3.

Figure 4 shows the R-D curves of the four schemes for

Table 2. R-D performance comparisons with LB-main.

Video sequences	Proposed vs. JCTVC-I0094 (dB)	Proposed vs. JCTVC-M0036 (dB)	Proposed vs. Wang's scheme (dB)
Class B	0.91	0.05	0.08
Class C	0.46	0.10	–0.05
Class D	0.36	0.07	–0.13
Class E	1.01	0.18	0.18
Average	0.69	0.10	0.02

Table 3. R-D performance comparisons with LP-main.

Video sequences	Proposed vs. JCTVC-I0094 (dB)	Proposed vs. JCTVC-M0036 (dB)	Proposed vs. Wang's scheme (dB)
Class B	0.85	0.02	0.08
Class C	0.44	0.04	0.15
Class D	0.35	0.02	0.08
Class E	0.94	0.04	0.15
Average	0.61	0.08	–0.01

Table 4. Bit rate mismatch comparisons with LB-main.

Video sequences	Average ΔR (%)			
	JCTVC-I0094	JCTVC-M0036	Wang's scheme	Proposed
Class B	2.30	0.01	0.05	0.04
Class C	2.58	0.03	0.06	0.16
Class D	2.41	0.08	0.42	0.06
Class E	2.03	0.22	0.16	0.06
Average	2.33	0.09	0.17	0.08

Table 5. Bit rate mismatch comparisons with LP-main.

Video sequences	Average ΔR (%)			
	JCTVC-I0094	JCTVC-M0036	Wang's scheme	Proposed
Class B	2.33	0.01	0.05	0.04
Class C	2.47	0.03	0.07	0.16
Class D	2.41	0.08	0.41	0.06
Class E	2.03	0.21	0.15	0.07
Average	2.31	0.08	0.17	0.08

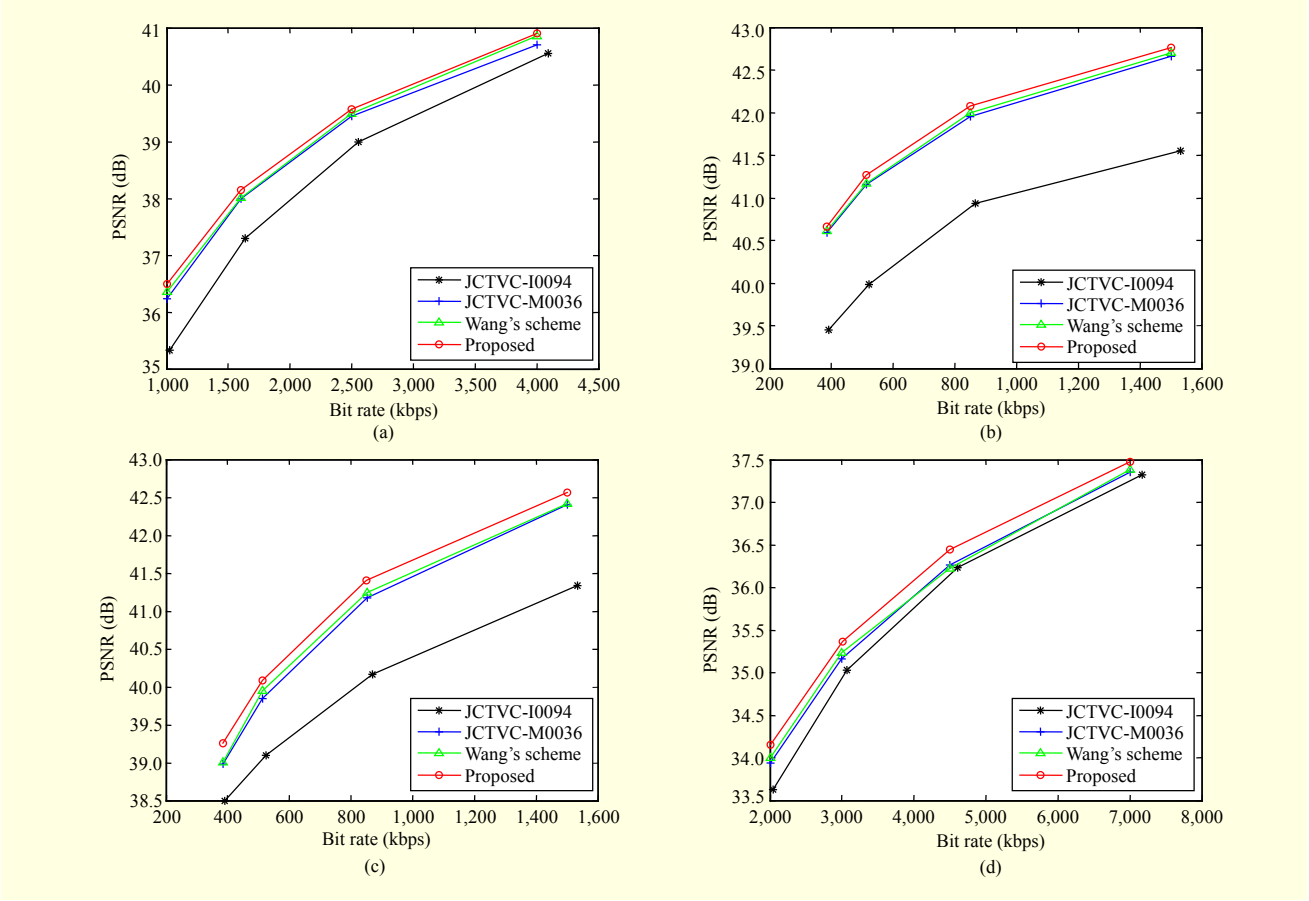

Fig. 4. R-D curves of four schemes with LB-main: (a) "Kimono," (b) "Johnny," (c) "KristenAndSara," and (d) "BasketballDrive."

the "Kimono," "Johnny," "KristenAndSara," and "BasketballDrive" sequences. From the figure, it can be observed that the proposed scheme has better R-D performances than the other three schemes for the four sequences.

In addition, the accuracy of the bit rate mismatch is defined for rate control in terms of mismatch error as follows:

$$\Delta R = \frac{|R_t - R_b|}{R_b} \times 100\%, \tag{33}$$

where R_t is the bit rate performance resulting from the test scheme, and R_b is the target bit rate.

Tables 4 and 5 show the average bit rate mismatch errors of all the four classes for the four schemes with LB-main and LP-main configurations, respectively. From the results in the tables, it can be seen that the proposed scheme can achieve the smallest bit rate mismatch error between the target bit rate and the actual bit rate.

V. Conclusion

In this paper, a novel q-domain rate control scheme for low delay video coding of HEVC was proposed. In the proposed scheme, an accurate R-Q model and a method for determining the QPs of the first frames within GOPs was proposed. Subsequently, a frame-level bit allocation method was also presented. Finally, based on the proposed R-Q model and the target bit allocated, QPs were predicted for CTUs by using rate-distortion optimization. The proposed scheme can be applied to real-time video communications. Experimental results show that the proposed scheme can achieve better R-D performances than JCTVC-I0094 and JCTVC-M0036, and a similar R-D performance to Wang's scheme; furthermore, it obtained the smallest bit rate mismatch error among all the four schemes.

The proposed scheme is developed for a low delay configuration. For a random access configuration, the proposed R-Q model and a CTU-level QP prediction can be used; the method of determining QPs for the first frames within GOPs, the GOP-level rate control, and the frame-level bit allocation should all be correspondingly adjusted, which is also our future work.

References

[1] ITU-T H.265 | ISO/IEC 23008-2, *High Efficiency Video Coding*,

Jan. 2013.

[2] W. Wu and H.K. Kim, "A Novel Rate Control Initialization Algorithm for H.264," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, May 2009, pp. 665–669.

[3] Y. Pitrey and M. Babel, "ρ-Domain Based Rate Control Scheme for Spatial, Temporal, and Quality Scalable Video Coding," *Proc. SPIE 7257 Visual Commun. Image Process.*, San Jose, CA, USA, Jan. 18–19, 2009, pp. 5–8.

[4] CCITT SG XV WP/1/Q4, *Description of Reference Model 8 (RM8)*, June 1989.

[5] E. Viscito and C. Gonzales, "A Video Compression Algorithm with Adaptive Bit Allocation and Quantization," *Proc. SPIE 1605 Visual Commun. Image Process.*, Boston, MA, USA, Nov. 1–2, 1991, pp. 58–72.

[6] ISO/IEC JTC/SC29/WG11, *MPEG Test Model 5 (TM5)*, Apr. 1993.

[7] 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, Feb. 1999, pp. 172–185.

[8] ISO/IEC JTC1/SC29/WG11, *MPEG-4 Video Verification Model Version 18.0: Coding of Moving Pictures and Audio*, Jan. 2001.

[9] A. Leontaris and A.M. Tourapis, "Rate Control Reorganization in the Joint Model (JM) Reference Software," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 23rd Meeting, San Jose, CA, USA, Doc. JVT-W042, Apr. 2007.

[10] Z. Li et al., "Adaptive Basic Unit Layer Rate Control for JVT," ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6 7th Meeting, Pattaya, Thailand, Doc. JVT-G012, Mar. 2003.

[11] M.-J. Kim and M.-C. Hong, "Fast Rate Control Algorithm in Frame-Layer for H.264/AVC Video Coding," *IEEE Trans. Consum. Electron.*, vol. 58, no. 3, Aug. 2012, pp. 872–879.

[12] M. Li et al., "Frame Layer Rate Control for H.264/AVC with Hierarchical B-frames," *J. Image Commun.*, vol. 24, no. 3, Mar. 2009, pp. 177–199.

[13] Y. Liu, Z.G. Li, and Y.C. Soh, "A Novel Rate Control Scheme for Low Delay Video Communication of H.264/AVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, Jan. 2007, pp. 68–78.

[14] S. Hu et al., "Rate Control Optimization for Temporal-Layer Scalable Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, Aug. 2011, pp. 1152–1162.

[15] H. Wang and S. Kwong, "A Rate-Distortion Optimization Algorithm for Rate Control in H.264," *IEEE Int. Conf. Acoust., Speech Signal Process.*, Honolulu, HI, USA, Apr. 15–20, 2007, pp. 1149–1152.

[16] H. Choi et al., "Rate Control Based on Unified RQ Model for HEVC," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 8th Meeting, San Jose, CA, USA, Doc. JCTVC-H0213, Feb. 2012.

[17] B. Li, H. Li, and L. Li, "Adaptive Bit Allocation for R-lambda Model Rate Control in HM," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 13th Meeting, Incheon, Rep. of Korea, Doc. JCTVC-M0036, Apr. 2013.

[18] S. Wang et al., "Rate-GOP Based Rate Control for High Efficiency Video Coding," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, Sept. 2013, pp. 1101–1111.

[19] H. Choi et al., "Improvement of the Rate Control Based on Pixel-Based URQ Model for HEVC," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 9th Meeting, Geneva, Switzerland, Doc. JCTVC-I0094, Apr. 2012.

[20] B. Li et al., "QP Determination by Lambda Value," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 9th Meeting, Geneva, Switzerland, Doc. JCTVC-I0426, Apr. 2012.

[21] I. Kim et al., "High Efficiency Video Coding (HEVC) Test Model 12 (HM12) Encoder Description," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 14th Meeting, Vienna, Austria, Doc. JCTVC-N1002, July 2013.

[22] 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, Feb. 1997, pp. 246–250.

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

[24] J. Dong and N. Ling, "A Context-Adaptive Prediction Scheme for Parameter Estimation in H.264/AVC Macroblock Layer Rate Control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, Aug. 2009, pp. 1108–1117.

[25] Y. Liu, Z.G. Li, and Y.C. Soh, "A Novel Rate Control Scheme for Low Delay Video Communication of H.264/AVC Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, Jan. 2007, pp. 68–78.

[26] F. Bossen, "Common Test Conditions and Software Reference Configurations," JCT-VC of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 9th Meeting, Geneva, Switzerland, Doc. JCTVC-I1100, Apr. 2012.

[27] G. Bjøntegaard, "Calculation of Average PSNR Differences between RD-Curves," ITU-T SG16 Q.6 VCEG 13rd Meeting, Austin, TX, USA, Doc. VCEG-M33, Apr. 2001.

Wei Wu received his BS degree in electronic materials and elements, and his MS and PhD degrees in communication and information systems from Xidian University, Xi'an, China, in 1998, 2001, and 2005, respectively. He is currently an associate professor with the School of Telecommunication Engineering, Xidian University. From 2007 to 2008, he was a postdoctoral researcher at Sejong University, Seoul, Rep. of Korea. His research interests include video coding and video signal processing.

Jiong Liu received his BS and MS degrees in communication and information systems from Xidian University, Xi'an, China, in 1995 and 2001, respectively. He is currently a lecturer with the School of Telecommunication Engineering, Xidian University. His research interests include video coding and video signal processing.

Lei Feng received his BS and MS degrees in communication and information systems from Xidian University, Xi'an, China, in 1999 and 2002, respectively. He is currently a lecturer with the School of Telecommunication Engineering, Xidian University. His research interests include video signal processing.

ETRI Journal, Volume 38, Number 1, February 2016
http://dx.doi.org/10.4218/etrij.16.0114.0366