A VBR Video Encoding for Locally Consistent Picture Quality With Small Buffering Delay Under Limited Bandwidth

Hoonjae Lee and Sanghoon Sull, *Member, IEEE*

***Abstract*—During consecutive large picture information changes under limited bandwidth, the use of constant bit rate (CBR) encoding for high definition television (HDTV) broadcasting systems often causes serious picture quality degradation due to the tight constraint on small buffering delay, whereas a variable allocation of bits depending on the picture information changes results in large buffering delay. In this paper, we propose a frame-layer variable bit rate (VBR) video encoding method for maintaining locally consistent picture quality of an encoded video with small buffering delay, which can be used for a video on demand (VOD) system which streams an encoded video file to a client device under limited bandwidth or a real-time streaming application allowing a small delay. An objective function defined as a sum of the local distortion variation of the encoded pictures and the underflow levels at the decoder buffer is minimized with respect to the discrete frame-layer quantization step sizes subject to a constraint on the target number of bits within a temporal sliding window. Then, the constrained discrete minimization problem is converted to an unconstrained continuous minimization problem by introducing a Lagrange multiplier and is solved by applying the first-order necessary condition for local minima with respect to the quantization step sizes. Experimental results demonstrate that the proposed method provides a lower local standard deviation of PSNR comparing with the recent VBR method by 29.2% on average with small buffering delay.**

***Index Terms*—Buffering delay, consistent picture quality, rate control, variable bit rate (VBR), video encoding.**

I. Introduction

DIGITAL television broadcasting standards impose a tight constraint on small buffering delay under limited bandwidth. For example, the ATSC (Advanced Television systems Committee) HDTV standard [1] limits the buffering delay to 0.5 seconds under 19.39 Mbps. For this reason, HDTV broadcasting systems use a CBR encoding, but the problem is that serious picture quality degradations often appear during consecutive large picture information changes due to such as abrupt brightness change, fast motion and scene change. To maintain consistent picture quality of an encoded video without such degradation, bits can be allocated variably during encoding, depending on the picture information changes. For example, an encoder used in a VOD system that streams a pre-encoded video under limited bandwidth can variably allocate bits for consistent picture quality, but a long buffering delay could occur.

A typical VBR encoding method which variably allocates bits depending on the picture information changes is basically interested in maintaining consistent picture quality subject to a target number of bits for the digital video storage applications such as digital versatile disk (DVD) or blu-ray disk. Jagmohan *et al.* [2] proposed a VBR encoding method for MPEG-4 [3] under the assumption that adjusting the resolution of a frame cause less distortion than adjusting the quantization parameter, and Kamaci *et al.* [4] proposed a VBR encoding method which estimates the distortion for an encoded picture by analyzing the distribution of the coefficients obtained by applying the discrete cosine transform. Bagni *et al.* [5] proposed a VBR encoding approach which adjusts the reference frame-layer quantization parameter based on a current frame, and Wang *et al.* [6] proposed a VBR encoding which limits the distortion of an encoded frame in a given range. These typical VBR encoding methods can maintain consistent picture quality, but they allocate bits without much considering the buffering delay or limited bandwidth and thus a long buffering delay could occur when the encoded video is delivered under limited bandwidth.

On the other hands, the use of the CBR encoding and the buffer-constrained VBR yields smaller or negligible buffering delay when an encoded video is delivered under limited bandwidth. Xie *et al.* [7] proposed a CBR encoding method which adjusts the quantization parameter of a current frame to prevent the underflow and overflow of the decoder buffer when the buffer level is close to zero and the maximum buffer level, respectively. Rezaei *et al.* [8] proposed a VBR encoding which determines the quantization parameter of a current frame by adding or subtracting several control values from the quantization parameter of the previous frame to maintain consistent picture quality of the encoded video while preventing the underflow and overflow of the decoder buffer given a buffer size and bandwidth. The buffer-constrained VBR encoding methods proposed in [9]–[11] determine the quantization parameter of a current frame by estimating the distortion of the encoded current frame and the number of bits to encode the current frame while preventing the underflow and overflow. Although these methods attempt to maintain consistent picture quality, there is an inherent limitation on the reduction of the abrupt picture quality fluctuation during consecutive large picture information changes since they determine the quantization param-

Manuscript received February 01, 2011; revised May 19, 2011; accepted July 27, 2011. Date of publication September 22, 2011; date of current version February 23, 2012. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2010-0028117).

The authors are with the School of Electrical Engineering, Korea University, Sungbuk-ku, Seoul 136-701, Korea (e-mail: hoonjae@mpeg.korea.ac.kr; sull@korea.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2011.2164308

eter for a single current frame without considering its neighboring frames. Such quality fluctuation in short duration becomes more annoying to viewers especially when a high-resolution video is presented on a HDTV. The buffer-constrained VBR encoding methods proposed in [12] and [13] reduce the distortion variation of the encoded frames for consistent picture quality of the encoded video. However, the performance of those methods could be limited, since the encoding method in [12] minimizes only the distortion difference between the encoded current frame and the encoded past frame, and the encoding method in [13] reduces the distortion variation by estimating the rate-distortion for the inter-frames without considering the picture information changes. Several two-pass VBR encoding methods [14]–[18] analyzed the entire frames of a video to maintain consistent picture quality before encoding, but due to its two-pass encodings, these methods cannot be used for real-time streaming applications even though a small delay is allowed.

In this paper, we propose a VBR video encoding method that maintains locally consistent picture quality of an encoded video with small buffering delay under limited bandwidth. The proposed VBR encoding method can be used to encode videos for a VOD system which streams an encoded video file to a client device under limited bandwidth or a real-time streaming application allowing a small delay. Our idea is to allocate more bits to the frames with the consecutive large picture information changes while assigning less bits to other frames with less picture information changes within a temporal sliding window. Thus, the quantization step size for a current frame is determined by considering the current frame and its neighboring frames while reducing the decoder buffer underflows. We formulate our encoding method as a constrained minimization problem and minimize an objective function which is defined as a sum of the local distortion variation of the encoded pictures and the buffer underflow levels subject to a constraint on the target number of bits within the temporal sliding window centered at a current frame. Since our formulation considers the temporal sliding window consisting of a given number of past, current and future frames, the encoding delay corresponding to the number of the future frames to be considered to encode the current frame occurs although the encoding delay does not appear during decoding.

In the constrained minimization formulation, the distortion of the encoded pictures and the number of bits needed to encode pictures are modeled as simple functions of the frame-layer quantization step sizes for the current and future frames within the temporal sliding window, and we search for a set of the optimal quantization step sizes for the current and future frames that minimize the objective function while satisfying the constraint on the target number of bits where the actual values of the quantization step sizes, the distortion and the number of bits to encode the past frames obtained from the previous temporal sliding windows are used as the values for the past frames during the minimization. To accomplish this, we convert the constrained minimization problem to an unconstrained minimization problem by introducing a Lagrange multiplier, and solve it by applying the first-order necessary condition for local minima with respect to the frame-layer quantization step sizes. After solving the unconstrained minimization, the part of the minimizing solution corresponding to the current frame is selected as a quantization step size for the current frame and the temporal sliding window moves forward by one frame to determine the quantization step size for the next frame. Note that, even though we minimize the objective function which includes only the local distortion variation and the decoder buffer underflow levels without considering the distortion itself, the overall picture quality degradation due to allocating insufficient number of bits does not occur because a set of quantization step sizes is selected satisfying the constraint on the target number of bits within the temporal sliding window. Experimental results demonstrate that the proposed method provides a lower local standard deviation of PSNR comparing to the buffer-constrained VBR method [8] by 29.2% on average at the expense of a small increase of buffering delay.

The paper is organized as follows. In Section II, we describe a VBR encoding method for maintaining locally consistent picture quality of the encoded video with small buffering delay under limited bandwidth. In Section III, we demonstrate the performance of the proposed method through experiments. Section IV concludes the paper.

II. Proposed VBR Encoding Method

In this section, we describe the proposed VBR encoding method. We first describe our mathematical formulation to maintain locally consistent picture quality with small buffering delay based on a frame-layer rate control scheme. Then, we present a rate-quantization (R-Q) model and a distortion- quantization (D-Q) model to estimate the rate and distortion of an encoded frame for a given quantization step size. Finally, we describe how to determine the quantization step size for each frame based on our mathematical formulation, R-Q and D-Q models.

A. Mathematical Formulation

The goal of the proposed VBR encoding is to encode frames in a video so that the encoded frames have locally consistent picture quality with small buffering delay under limited bandwidth, assuming that the encoded bits are delivered to a decoder from a VOD server or a real-time streaming server in a constant bit-rate. An objective function to be minimized is defined as a sum of the local distortion variation of the encoded pictures and the underflow levels at the decoder buffer subject to a constraint on the target number of bits within a temporal sliding window.

For a temporal sliding window centered at a current frame f_k with the length of $2N$ frames, the proposed encoding is formulated as a constrained minimization problem with respect to the N variables, $q_k, q_{k+1}, \cdots, q_{k+N-1}$ which denote quantization step sizes corresponding to the current and $N-1$ future frames, respectively:

$$\begin{aligned}\underset{q_k,\cdots,q_{k+N-1}}{\arg} \min \frac{1}{2N} \sum_{n=k-N}^{k+N-1} \left(d_n(q_n) - \overline{d}(q_{k-N},\cdots,q_{k+N-1})\right)^2 \\ + w \cdot \frac{1}{N} \sum_{n=k}^{k+N-1} \sigma\left(l_n(q_n)\right)\end{aligned}$$

$$\text{subject to} \sum_{n=k-N}^{k+N-1} b_n(q_n) = B_{\text{local}}, \tag{1}$$

where $q_{k-N}, \cdots, q_{k-1}$ that are the actual values of the quantization step sizes used to encode the past frames are considered as constants, $d_n(q_n)$ represents the distortion resulting from encoding f_n with q_n, $b_n(q_n)$ is the number of bits to encode f_n with q_n, $\overline{d}(q_{k-N}, \cdots, q_{k+N-1})$ is the average distortion for the encoded frames, and w is a weight value. The constant B_{local} denotes the local target number of bits to encode the frames in the temporal sliding window and calculated as follows:

$$B_{\text{local}} = 2N \cdot (R_{\text{target}}/F_{\text{rate}}), \tag{2}$$

where R_{target} and F_{rate} denote a given target bit-rate and frame-rate, respectively.

The number of bits residing in the decoder buffer, $u_n(q_n)$, after the encoded f_n is decoded is computed by adding $R_{\text{target}}/F_{\text{rate}}$ (the number of bits of the encoded video data that the decoder receives during the time interval of $1/F_{\text{rate}}$) to $u_{n-1}(q_{n-1})$ and subtracting the number of bits needed to encode f_n from it as follows:

$$u_n(q_n) = u_{n-1}(q_{n-1}) + \frac{R_{\text{target}}}{F_{\text{rate}}} - b_n(q_n). \tag{3}$$

Defining the decoder buffer level $l_n(q_n)$ as the value of $u_n(q_n)$ normalized by R_{target}, we have

$$l_n(q_n) \triangleq u_n(q_n)/R_{\text{target}}. \tag{4}$$

The value of $l_n(q_n)$ can be interpreted as the time required to receive the amount of the encoded video data residing in the decoder buffer. Thus, the negative value of $l_n(q_n)$ implies that the decoder should wait for $l_n(q_n)$ seconds to receive the amount of the encoded video data. In our experiments, the typical values of $l_n(q_n)$ range from -2 to 2 seconds. Since only the negative buffer level corresponding to the decoder buffer underflow is to be minimized, the sigmoid function $\sigma(x)$ is used to make the objective function smooth during the minimization.

To find a solution of (1), a discrete space which includes all combinations of the discrete frame quantization step sizes within the temporal sliding window could be searched, but the search space is too huge since there exist $(N_{\text{QP}})^N$ possible sets of quantization step sizes if N_{QP} denotes the number of the available quantization step sizes defined in the video coding standard such as MPEG-2/4 [3], [19] and H.264/AVC [20]. In this paper, the discrete minimization problem (1) is converted to a continuous minimization problem by assuming the quantization step size for each frame is continuous. Then, the constrained minimization problem is converted to an unconstrained minimization problem by introducing a Lagrange multiplier for the constraint on the target number of bits within the temporal sliding window, and is solved by applying the first-order necessary conditions for local minima with respect to the set of quantization step sizes. Then, only the quantization step size q_k corresponding to the current frame from the resulting solution set is selected and used to actually encode the current frame f_k. The temporal sliding window moves forward by one frame to determine the quantization step size for the next current frame. It is noted that our formulation maintains the overall picture quality although only the sum of the local distortion variation and the buffer underflow levels is minimized without including the term for the average distortion since a set of quantization step sizes $\{q_k, q_{k+1}, \cdots, q_{k+N-1}\}$ should satisfy the constraint on the target number of bits through $b_n(q_n)$s during the minimization.

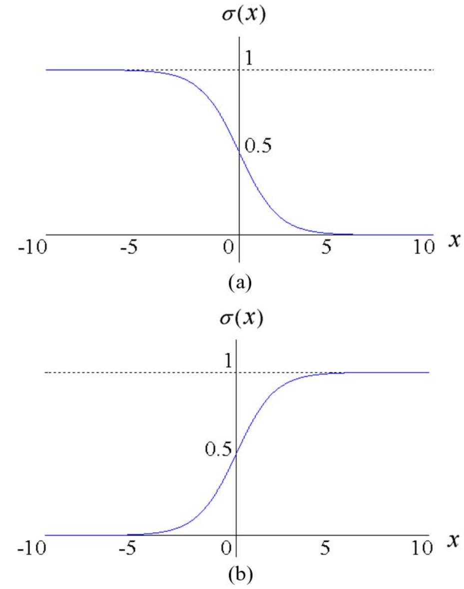

Fig. 1. Sigmoid function. (a) $s = -1$. (b) $s = 1$.

The underflow at the decoder buffer means that no video data remains in the decoder buffer, causing a buffering delay in order to receive more video data. Even though the buffer level cannot have a negative value physically, we can interpret the negative value of buffer level as the buffering delay as described previously. Thus, the buffering delay for f_n is obtained as follows:

$$\text{buffering delay} = -\min(l_n(q_n), 0). \tag{5}$$

Since only the levels of the underflows corresponding to the negative values of $l_n(q_n)$ is to be minimized to reduce the buffering delay, the sigmoid function of $l_n(q_n)$ is minimized with respect to the quantization step sizes to make the objective function smooth. The well-known sigmoid function $\sigma(x)$ and the first derivative of the sigmoid function are expressed as follows:

$$\sigma(x) = \frac{1}{1+e^{sx}},$$
$$\frac{\partial \sigma(x)}{\partial x} = s \cdot \sigma(x) \cdot \{1 - \sigma(x)\}, \tag{6}$$

where s is a constant that determines the steepness of the sigmoid function. The sigmoid function is plotted in Fig. 1. The sigmoid function acts like a unit-step function, but the sigmoid function is differentiable whereas the unit-step function is not. When the buffer level $l_n(q_n)$ is input to the sigmoid function with a negative value of s, the value of $\sigma(l_n(q_n))$ is close to zero if $l_n(q_n)$ is positive, and the value of $\sigma(l_n(q_n))$ is close to one if $l_n(q_n)$ is negative. Therefore, if we minimize the sum of

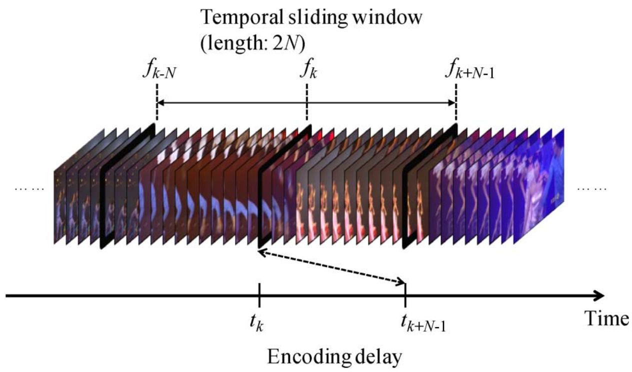

Fig. 2. Relation between a temporal sliding window and the encoding delay.

the sigmoid functions of $l_n(q_n)$s with a negative s, the levels of the underflows can be minimized.

Note that the decoder buffer level can rise to a higher level if the numbers of bits smaller than $R_{\rm target}/F_{\rm rate}$ are used to encode consecutive frames in (3) and (4). However, the average number of bits allocated to encode a frame becomes $R_{\rm target}/F_{\rm rate}$ with the proposed method since the target number of bits within the temporal sliding window corresponding to $2N$ frames is constrained to $2N \cdot (R_{\rm target}/F_{\rm rate})$ in (1) and (2). Therefore, the decoder buffer level does not rise to a higher level even though we minimize the objective function which includes only the sum of the local distortion variation and the decoder buffer underflow levels without considering the overflow. In our experiments, we observed that the decoder buffer level stayed under $2R_{\rm target}$ which is acceptable for client devices.

The proposed method requires $N-1$ future frames within a temporal sliding window to encode a current frame, resulting in the encoding delay in real-time encoding applications. Fig. 2 illustrates the encoding delay related to a temporal sliding window. To determine the quantization step size for the current frame f_k, $N-1$ future frames within the temporal sliding window need to be considered and thus f_k is encoded at t_{k+N-1}. Note that the encoding delay does not need to be taken into account in our application because the encoding delay does not appear during decoding.

B. Models for Rate-Quantization and Distortion-Quantization

The actual values of $b_n(q_n)$ and $d_n(q_n)$ in (1) can be obtained only after encoding f_n with q_n, but the actual encoding of a frame to obtain $b_n(q_n)$ and $d_n(q_n)$ for a large number of values of q_n during the minimization is impractical because of high computational complexity. Therefore, we need to estimate the values of $b_n(q_n)$ and $d_n(q_n)$ using the appropriate rate-quantization (R-Q) model and the distortion-quantization (D-Q) model, respectively. In this paper, we use a simple R-Q model and a D-Q model which are experimentally found out to yield satisfactory results to estimate the values of $b_n(q_n)$ and $d_n(q_n)$ for current and future frames in a temporal window.

1) Rate-Quantization Model: The R-Q models proposed in [21], [22] use MAD_n defined as the mean absolute difference of the original image f_n and its reconstructed image based on the observation that more bits are needed to encode more complex frames. Since MAD_n is available only after f_n is encoded, it is estimated from MAD_{n-1} in the conventional encoding schemes under the restrictive assumption that consecutive frames are similar [21]–[27]. However, the assumption often does not hold when there are large picture information changes.

In this paper, we use a modified linear R-Q model by using the square root of the sum of absolute difference (SAD) between the frame f_n and its reference frame as the measure for the picture information change during inter-frame coding of f_n. Assuming one reference frame f_{n-L} for simplicity where L is the distance between f_n and f_{n-L}, the picture information change at f_n is simply estimated as

$$X_n = \sqrt{SAD_n} = \sqrt{\sum_{i=1}^{I} |f_{n-L}(i) - f_n(i)|}, \tag{7}$$

where $f_n(i)$ is the pixel value in the original image f_n, and I is the number of pixels in f_n. It is experimentally found out that X_n well reflects the large picture information changes when L is equal to one for the IPPP structure. Thus, the modified linear R-Q model for inter-frame f_n using SAD_n is given as follows:

$$b_n(q_n) = \frac{K_n \cdot X_n}{q_n}, \tag{8}$$

where K_n is the model parameter to be determined by using the actual values of $X_{n-M}, \cdots, X_{n-1}$, which are the picture information changes for M recent inter-frames $f_{n-M}, \ldots, f_{n-1}$, and the actual values of $q_{n-M}, \cdots, q_{n-1}$ and $b_{n-M}(q_{n-M}), \cdots, b_{n-1}(q_{n-1})$ that were used to encode $f_{n-M}, \ldots, f_{n-1}$. Denoting $[X_{n-M}/q_{n-M}, X_{n-M+1}/q_{n-M+1}, \cdots, X_{n-1}/q_{n-1}]^{\rm T}$ and $[b_{n-M}(q_{n-M}), b_{n-M+1}(q_{n-M+1}), \cdots, b_{n-1}(q_{n-1})]^{\rm T}$ by $\mathbf{a}$ and $\mathbf{b}$, respectively, and applying the least square method to (8) for M recent inter-frames in a similar way in [22], we arrive at

$$\mathbf{a}^{\rm T}\mathbf{a} \cdot K_n = \mathbf{a}^{\rm T}\mathbf{b}. \tag{9}$$

Letting a_m and b_m be the m-th element of $\mathbf{a}$ and $\mathbf{b}$, respectively, K_n is calculated with the following equation:

$$K_n = \left(\sum_{m=1}^{M} a_m b_m\right) \Big/ \left(\sum_{m=1}^{M} a_m^2\right). \tag{10}$$

Then, $b_n(q_n)$ is estimated from (8).

If the frame f_n is an intra-frame, K_n is calculated based on $q_{n-M}, \cdots, q_{n-1}$ and $b_{n-M}(q_{n-M}), \cdots, b_{n-1}(q_{n-1})$ which are the actual values used to encode M recent intra-frames. However, the picture information changes $X_{n-M}, \cdots, X_n$ corresponding to M recent intra-frames and the current frame does not have to be taken into account because the number of bits to encode an intra-frame does not depend on other frames. Thus, the modified linear R-Q model for the intra-frame f_n is given as follows:

$$b_n(q_n) = \frac{K_n}{q_n}. \tag{11}$$

Denoting $[1/q_{n-M}, 1/q_{n-M+1}, \cdots, 1/q_{n-1}]^{\rm T}$ and $[b_{n-M}(q_{n-M}), b_{n-M+1}(q_{n-M+1}), \cdots, b_{n-1}(q_{n-1})]^{\rm T}$ by $\mathbf{a}$ and $\mathbf{b}$, respectively, K_n is calculated by applying the least square method to (11) for M recent intra-frames and using (9) and (10). Then, $b_n(q_n)$ is estimated from (11).

Note that the numerous macroblocks in the inter-frame could be encoded with the intra coding modes if there are large picture information changes such as abrupt scene changes. However, it is common to allocate the smaller number of bits to encode the inter-frames than the intra-frames. Therefore, we can use two simple R-Q models in (8) and (11) for the inter-frames and intra-frames, respectively, even in such special cases, to allocate the adequate number of bits according to the frame type.

2) Distortion-Quantization Model: The relation between $d_n(q_n)$ and q_n is modeled as a quadratic function in [28] as $d_n(q_n) = (2 \cdot q_n)^2/12$. By assuming q_n vary within a narrow range over the consecutive frames in the temporal sliding window to maintain locally consistent picture quality of the encoded video, the quadratic function can be locally approximated as the linear function as follows:

$$d_n(q_n) = c_n \cdot q_n, \tag{12}$$

where c_n is a model parameter which is calculated by using the distortions of M recent frames, $d_{n-M}(q_{n-M}), \ldots, d_{n-1}(q_{n-1})$, and the quantization step sizes used to encode M recent frames, $d_{n-M}, \ldots, d_{n-1}$, as follows:

$$c_n = \frac{1}{M} \sum_{m=n-M}^{n-1} \frac{d_m(q_m)}{q_m}. \tag{13}$$

Note that since the temporal sliding window contains the past frames which were already encoded, the estimated values of the distortion and rate are used only for the current and future frames based on (8), (11) and (12) while the actual values of the distortion and rate are used for the past frames.

To check the validity of the R-Q and D-Q models we used, the estimated and actual values of $b_n(q_n)$ and $d_n(q_n)$ when the 1800 frames from three HD video sequences are encoded with various target bit-rate by using the proposed R-Q model are plotted in Fig. 3(a) and (b), respectively, and the mismatch (computed by subtracting the actual value from the estimated value) for the number of bits and the distortion are plotted in Fig. 3(c) and (d), respectively. In Fig. 3(c), the mismatch becomes larger than 5×10^6 bits for the 13 inter-frames noted in the dashed box when the picture information change X_n is large and the quantization step size q_n is small in (8) under the target bit-rate of 10^8 bits/s, resulting in the larger estimated number of bits compared with the actual number of bits. In this case, the actual numbers of bits to encode all frames within the temporal sliding window become smaller than the local target number of bits B_{local} in (2). Therefore, the overall visual quality of the encoded frames could be slightly reduced, and the decoder buffer level could increase, implying that such relatively large mismatch does not result in the underflow at the decoder. Note that the mismatch for the number of bits to encode the intra-frames estimated by using (11) is smaller compared with that of the inter-frames where (8) used. In most cases, we can see that the values of rate and distortion are well estimated by using our R-Q model and D-Q model so that the models could be used to estimated the rate and distortion of current and future frames in a temporal sliding window.

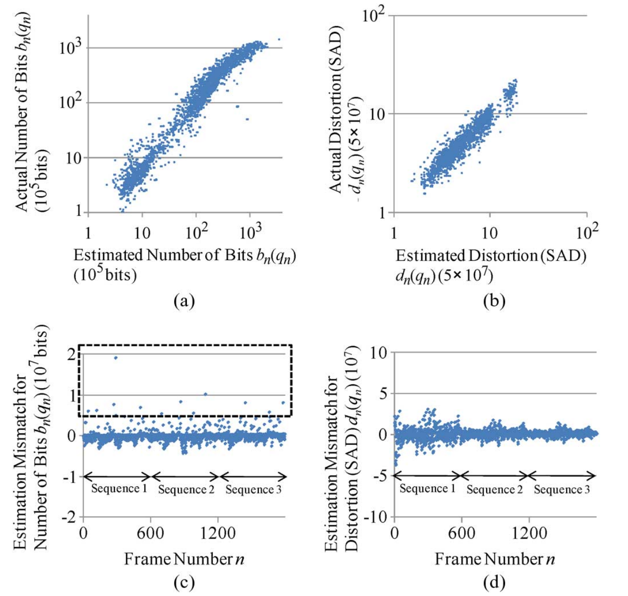

Fig. 3. Performance of our R-Q and D-Q models for 1800 frames from three HD video sequences: (a) Actual and estimated numbers of bits. (b) Actual and estimated distortions. (c) Mismatches between the actual and estimated numbers of bits. (d) Mismatches between the actual and estimated distortions.

C. Determination of Quantization Step Size

Rewriting (1) by using (2), (8) and (12), we obtain the following:

$$\begin{aligned} \arg\min_{q_k,\cdots,q_{k+N-1}} \; & \frac{1}{2N} \sum_{n=k-N}^{k+N-1} \left(c_n \cdot q_n - \overline{d}(q_{k-N},\cdots,q_{k+N-1})\right)^2 \\ & + w \cdot \frac{1}{N} \sum_{n=k}^{k+N-1} \sigma\left(l_n(q_n)\right) \\ \text{subject to} \; & \sum_{n=k-N}^{k+N-1} \frac{K_n \cdot X_n}{q_n} = 2N \cdot \frac{R_{\text{target}}}{F_{\text{rate}}}. \end{aligned} \tag{14}$$

To solve (14), the constrained minimization is converted to an unconstrained minimization by using a Lagrange multiplier as follows:

$$\arg\min_{q_k,\cdots,q_{k+N-1},\lambda} J(q_k,\cdots,q_{k+N-1},\lambda)$$

with

$$\begin{aligned} & J(q_k,\cdots,q_{k+N-1},\lambda) \\ & \quad = f(q_k,\cdots,q_{k+N-1}) + \lambda \cdot g(q_k,\cdots,q_{k+N-1}), \\ & f(q_k,\cdots,q_{k+N-1}) \\ & \quad = \frac{1}{2N} \sum_{n=k-N}^{k+N-1} \left(c_n \cdot q_n - \overline{d}(q_{k-N},\cdots,q_{k+N-1})\right)^2 \\ & \qquad + w \cdot \frac{1}{N} \sum_{n=k}^{k+N-1} \sigma\left(l_n(q_n)\right), \\ & g(q_k,\cdots,q_{k+N-1}) \\ & \quad = \sum_{n=k-N}^{k+N-1} \frac{K_n \cdot X_n}{q_n} - 2N \cdot \frac{R_{\text{target}}}{F_{\text{rate}}}, \end{aligned} \tag{15}$$

where $J(q_k, \cdots, q_{k+N-1}, \lambda)$ is the cost function, and λ is the Lagrange multiplier. We solve (15) by applying the first-order necessary conditions for local minima:

$$\begin{aligned}
&\frac{\partial J(q_k, \cdots, q_{k+N-1}, \lambda)}{\partial q_k} \\
&= \frac{2 \cdot c_k}{N} \cdot \left\{ c_k \cdot q_k - \frac{1}{N} \sum_{n=k-N}^{k+N-1} (c_n \cdot q_n) \right\} \\
&\quad + \frac{w}{N} \sum_{n=k}^{k+N-1} \left\{ \frac{\partial \sigma\left(l_n(q_n)\right)}{\partial l_n(q_n)} \cdot \frac{\partial l_n(q_n)}{\partial q_k} \right\} - \lambda \cdot \frac{K_k \cdot X_k}{q_k^2} \\
&= 0, \\
&\quad \vdots \\
&\quad \frac{\partial J(q_k, \cdots, q_{k+N-1}, \lambda)}{\partial q_{k+N-1}} \\
&= \frac{2 \cdot c_{k+N-1}}{N} \cdot \left\{ c_{k+N-1} \cdot q_{k+N-1} - \frac{1}{N} \sum_{n=k-N}^{k+N-1} (c_n \cdot q_n) \right\} \\
&\quad + \frac{w}{N} \sum_{n=k}^{k+N-1} \left\{ \frac{\partial \sigma\left(l_n(q_n)\right)}{\partial l_n(q_n)} \cdot \frac{\partial l_n(q_n)}{\partial q_{k+N-1}} \right\} \\
&\quad - \lambda \cdot \frac{K_{k+N-1} \cdot X_{k+N-1}}{q_{k+N-1}^2} \\
&= 0, \\
&\quad \frac{\partial J(q_k, \cdots, q_{k+N-1}, \lambda)}{\partial \lambda} \\
&= \sum_{n=k}^{k+N-1} \left(\frac{K_n \cdot X_n}{q_n} \right) - 2N \cdot \frac{R_{\text{target}}}{F_{\text{rate}}} \\
&= 0. \qquad (16)
\end{aligned}$$

Then, to obtain the solution of (14), Newton's method is applied to find the roots of (16): We first rewrite (16) as follows:

$$\mathbf{F}(\mathbf{x}) = 0,$$

where

$$\begin{aligned}
\mathbf{F}(\mathbf{x}) &= [F_1 \quad \ldots \quad F_N \quad F_{N+1}]^{\mathrm{T}} \\
&= \left[\frac{\partial J(\mathbf{x})}{\partial x_1} \quad \ldots \quad \frac{\partial J(\mathbf{x})}{\partial x_N} \quad \frac{\partial J(\mathbf{x})}{\partial x_{N+1}} \right]^{\mathrm{T}}, \\
\mathbf{x} &= [x_1, \cdots, x_N, x_{N+1}]^{\mathrm{T}} \\
&= [q_k, \cdots, q_{k+N-1}, \lambda]^{\mathrm{T}}. \qquad (17)
\end{aligned}$$

To find the root of (17), we update $\mathbf{x}$ iteratively so that $\mathbf{F}(\mathbf{x})$ converges to zero by obtaining the updating value $\delta\mathbf{x}$ from the following equation for each step:

$$\mathbf{F}(\mathbf{x} + \delta\mathbf{x}) = 0. \qquad (18)$$

Equation (18) can be expanded by using Taylor series as follows:

$$\mathbf{F}(\mathbf{x} + \delta\mathbf{x}) = \mathbf{F}(\mathbf{x}) + \mathbf{J} \cdot \delta\mathbf{x} + \mathbf{R}, \qquad (19)$$

where $\mathbf{R}$ denotes the high-order remainder, and $\mathbf{J}$ is the Jacobian matrix where the element of $\mathbf{J}$ is defined as

$$J_{ij} \triangleq \frac{\partial F_i}{\partial x_j}. \qquad (20)$$

From (18) and (19), we have the following linear equation by neglecting the high-order remainder $\mathbf{R}$:

$$\mathbf{J} \cdot \delta\mathbf{x} = -\mathbf{F}(\mathbf{x}). \qquad (21)$$

Then, we can obtain $\delta\mathbf{x}$ by solving (21) and $\mathbf{x}$ is updated to $\mathbf{x} + \delta\mathbf{x}$. If we update $\mathbf{x}$ iteratively until $\mathbf{F}(\mathbf{x})$ converges to zero, we can find the solution. By using the average values of the previous quantization step sizes as the initial values of q_ks, Newton's iteration method generally converged well in our experiments.

The part of the solution of (14) corresponding to the current frame f_k is selected and converted to the nearest quantization parameter, and used to encode the current frame. Then, the temporal sliding window moves forward by one frame and the quantization step size for the next frame is calculated in a similar way.

III. Experimental Results

We implemented the proposed method to evaluate its performance by using the H.264/AVC reference software (JM13.2) on a PC with an Intel Core 2 Duo 2.4GHz CPU and 2GB RAM. For comparison, we also implemented the buffer-constrained VBR method [8]. Experiments are performed on several HD (1920×1080) video sequences. Five video sequences (*HD Music Show 1*, $\cdots$, *HD Music Show 5*) were generated from a terrestrial HDTV music show program containing the latest most popular dance music and songs which have consecutive large picture information changes such as abrupt brightness change, fast motion and scene change. In addition, three HD sequences (*Crow run*, *Duck Take Off*, and *Park Joy)* are also used in the experiments. The original raw video sequences are encoded with the proposed method, the existing VBR method [8], and the H.264/AVC reference software with a constant quantization parameter, respectively. The video sequences are encoded with the frame rate of 30 fps and IPPP structure with GOP length of 15 frames. The target bit-rate for each video sequence is set to the average bit-rate of the constant quantization parameter (QP) encoding. For example, the target bit-rate for *HD Music Show 1* is set to 6,394,395 bits/s.

In our experiments, the length of the temporal sliding window $2N$ is set to 60 frames corresponding to four GOPs, and w in (1) is set to 3.0×10^6 by considering the range of the distortion and buffer level in (14). Newton's method is implemented based on the routine in [29] where the maximum iteration number is set to 30. The decoder buffer size of the existing VBR method [8] is set to the R_{target}. When we simulate the situation that a pre-encoded video file is delivered from its beginning under limited bandwidth, the initial buffer level is assumed to zero in our experiments.

Fig. 4 shows the comparison of the PSNRs and the decoder buffer levels when *HD Music Show 1* and *HD Music Show 4* are encoded with the CBR encoding method [7], constant QP, existing VBR method [8] and the proposed method, respectively, where the buffer size for the CBR encoding method is set to the half of R_{target}. Severe quality fluctuations in the encoded picture appear with the CBR encoding method and the severe underflow and large buffering delay occur with the constant QP

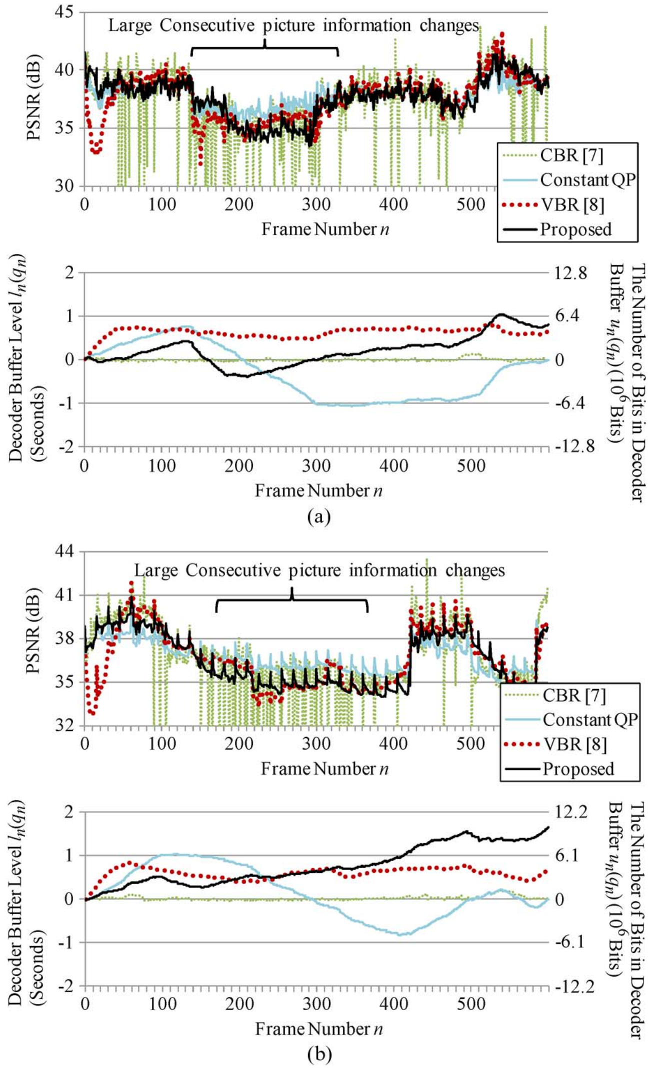

Fig. 4. Comparison of the PSNRs and the buffer levels: (a) *HD Music Show 1*. (b) *HD Music Show 4*.

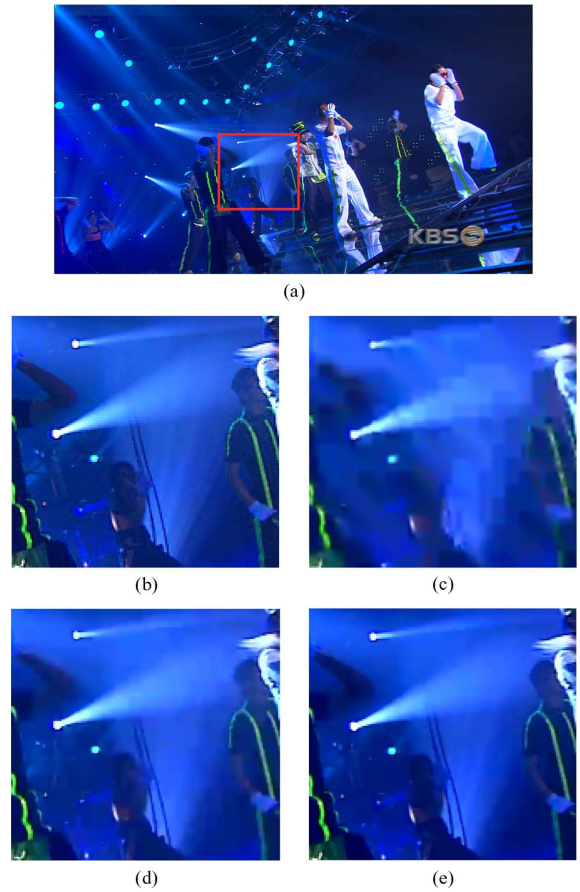

Fig. 5. Comparison of the encoding result for the 180st frame of the *HD Music Show 1*. (a) An original input image. (b) A magnified original input image corresponding to the red box in (a). (c) Result of the CBR encoding [7] (PSNR: 27.16dB). (d) Result of the existing VBR method [8] (PSNR: 33.76dB). (e) Result of the proposed method (PSNR: 37.82dB).

TABLE I
COMPARISON OF THE SUBJECTIVE VISUAL QUALITY

Encoding Method	*HD Music Show 1*	*HD Music Show 4*
CBR [7]	2.76	2.68
VBR [8]	3.16	3.00
Proposed	3.78	3.80

encoding especially when there are consecutive large picture information changes. The existing VBR method [8] prevents the underflow by abruptly increasing the quantization parameter for a current frame when the decoder buffer level is low and thus we can see the large picture quality fluctuations both at the consecutive large picture information changes and at the beginning of the video delivery where the buffer level is relatively low. In the case of the proposed method, the picture quality of the encoded video is more locally consistent because the distortion variation is minimized within the temporal sliding window centered at a current frame. Moreover, the underflow levels at the decoder buffer are also minimized, yielding the small buffering delay. Comparison of the encoding results for the 180st frame of the *HD Music Show 1* is shown in Fig. 5. The CBR encoding method [7] results in the severe picture quality degradation and the existing VBR method [8] shows some picture quality degradation whereas the proposed method maintains the picture quality of the encoded video.

To compare the subjective visual quality of the encoded video for the different rate control schemes, ten viewers evaluated the visual quality of the *HD Music Show 1* and *HD Music Show 4* encoded with the CBR encoding method [7], the existing VBR method [8], and the proposed method, respectively, while the encoded videos are displayed on a 24 inch HD monitor. Table I shows the evaluation results where the score of 0 and 5.0 denote the worst and the best visual quality, respectively, and we can see that the proposed method gives the better visual subjective quality of the encoded videos compared with the existing VBR method [8] and the existing CBR method [7] as the proposed method reduces the local standard deviation of PSNR of the encoded video.

Table II shows the performance comparison for the proposed method with the existing VBR method [8] and the constant QP

TABLE II
PERFORMANCE COMPARISON OF THE RATE CONTROL METHODS FOR VARIOUS VIDEO SEQUENCES.

Sequence	Encoding Method	Bit-rate (bits/s)	PSNR Average (dB)	PSNR Average Local Standard Deviation* (dB)	PSNR Maximum Local Standard Deviation* (dB)	Buffering Delay (s)
HD Music Show 1	Constant QP (34)	6,394,395	37.83	0.67	1.33	1.07
	VBR [8]	6,185,331	37.53	0.99	2.31	0.00
	Proposed	**6,136,922**	**37.60**	**0.82**	**1.73**	**0.39**
HD Music Show 2	Constant QP (34)	3,375,072	37.16	0.26	0.51	1.69
	VBR [8]	3,248,592	36.93	0.51	1.46	0.10
	Proposed	**2,710,402**	**36.32**	**0.34**	**0.81**	**0.12**
HD Music Show 3	Constant QP (34)	4,434,666	37.81	0.86	1.40	2.27
	VBR [8]	4,285,363	37.43	1.20	2.73	0.02
	Proposed	**4,030,200**	**37.39**	**0.96**	**1.62**	**0.10**
HD Music Show 4	Constant QP (34)	6,101,974	36.74	0.52	1.22	0.82
	VBR [8]	5,892,586	36.47	0.89	2.75	0.01
	Proposed	**5,600,630**	**36.41**	**0.40**	**1.88**	**0.02**
HD Music Show 5	Constant QP (34)	6,964,461	37.14	0.76	1.39	0.87
	VBR [8]	6,734,470	36.87	1.13	2.40	0.02
	Proposed	**6,507,514**	**36.81**	**0.90**	**1.78**	**0.04**
Crow Run	Constant QP (40)	6,460,208	28.42	0.26	0.31	0.17
	VBR [8]	5,869,935	27.89	0.55	1.79	0.04
	Proposed	**6,397,262**	**28.39**	**0.33**	**0.42**	**0.12**
Duck Take Off	Constant QP (40)	8,803,955	28.14	0.27	0.54	0.02
	VBR [8]	8,123,674	27.64	0.48	1.53	0.01
	Proposed	**8,940,092**	**28.24**	**0.45**	**0.83**	**0.14**
Park Joy	Constant QP (40)	8,006,259	27.35	0.46	1.05	0.62
	VBR [8]	7,326,228	26.93	0.62	1.18	0.02
	Proposed	**7,799,707**	**27.33**	**0.33**	**1.42**	**0.20**
Average	Constant QP	6,317,624	33.82	0.51	0.97	0.94
	VBR [8]	5,958,272	33.46	0.80	2.02	0.03
	Proposed	**6,116,024**	**33.56**	**0.57**	**1.31**	**0.14**

* The local standard deviation is calculated within the sliding window of length $2N$.

encoding. The proposed method provides a lower average local standard deviation of the PSNR by 29.2% and maximum local standard deviation of the PSNR by 35.1% in comparison with the existing VBR method [8], while the average buffering delay is 0.14 second which is very small in comparison with the constant QP encoding. These results show that the proposed method

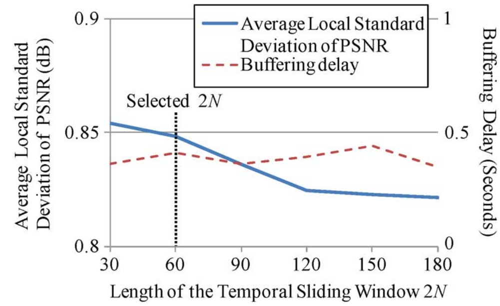

Fig. 6. Average local standard deviation of PSNR and the buffering delay when the *HD Music show 1* is encoded with the various lengths of the temporal sliding window $2N$.

provides more locally consistent picture quality of the encoded video compared with existing VBR method [8] at the expense of a small increase of the buffering delay. The constant QP encoding usually gives good quality consistency for the encoded pictures, but the decoder buffer underflow cannot be considered. Especially when the *HD Music Show 3* is encoded with the constant QP encoding method under the target bit-rate of 4.4×10^6 bits/s, the value of the number of bits residing in the decoder buffer, $u_n(q_n)$, in (3) decreases to -10^7 bits since it contains the severe consecutive picture information changes, resulting in the large buffering delay of 2.27 seconds.

The existing VBR method [8] results in small buffering delay at the expense of the large picture quality fluctuation because it assigns a large quantization parameter to the frame to prevent the underflow when the decoder buffer level is low, resulting in the sudden picture quality degradation. On the other hands, the proposed method minimizes both the local distortion variation of the encoded pictures and the underflows of decoder buffers, and thus can achieve locally consistent picture quality of the encoded video with small buffering delay.

To observe the effect of choosing different value of the length of the temporal sliding window $2N$ in (1), the *HD Music Show 1* is encoded with the various $2N$ where w is set to 3.0×10^6. Fig. 6 shows the average local standard deviation of the PSNR and the buffering delay. As $2N$ increases, the average local standard deviation of the PSNR slightly decreases, but the encoding time rapidly increases. In our experiments, $2N$ is set to 60 frames for acceptable encoding time.

Fig. 7 shows the effect of the various weight values of w in (1) on the average local standard deviation of PSNR and the buffering delay for the *HD Music Show 1* is encoded where the length of the temporal sliding window $2N$ is set to 60 frames. We can observe that the buffering delay becomes lower while the average local standard deviation of PSNR slightly increases as the weight becomes larger. Therefore, the values of w can be adjusted depending on various applications: for example, large w can be used if the buffering delay should be small. In our experiments, w is set to 3.0×10^6. Note that, referring to the objective function in (1), even though the size of the temporal sliding window $2N$ is varied, the underflow level term (the second term in the objective function) remains near the local extrema so that the margin for the further improvement is small, because the underflow level term is multiplied by a large weighting factor

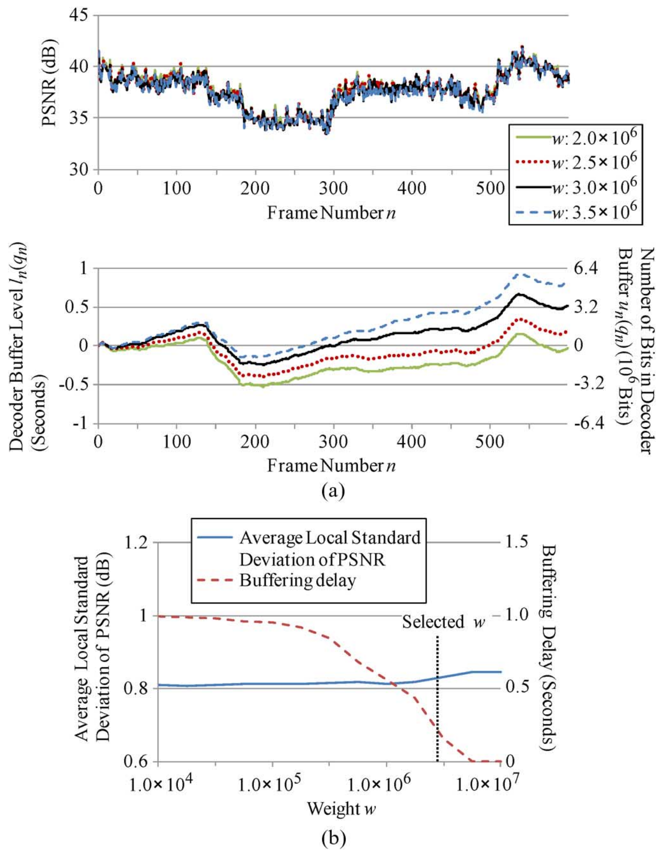

Fig. 7. Encoding results of the proposed method with various values of w for *HD Music Show 1*. (a) PSNR and decoder buffer level. (b) Average local standard deviation of PSNR and buffering delay.

TABLE III
COMPARISON OF THE AVERAGE ENCODING TIME (1920 × 1080, 600 FRAMES)

Encoding Method	Encoding Time
Constant QP	2 h 4 m 28 s
CBR [7]	2 h 37 m 3 s
VBR [8]	2 h 7 m 5 s
Proposed	2 h 25 m 36 s

w (3.0×10^6) in the proposed method. In the meantime, the local distortion variation term (the first term in the objective function) decreases if $2N$ is increased as can be seen in Fig. 6. Thus, the large weighting factor w needs not be changed as $2N$ is increased, but it should be increased as $2N$ is decreased to maintain the local distortion variation. For example, w should be increased by 3% when $2N$ decreases from 60 to 30.

To compare the encoding time of each encoding method, we measure the average time to encode 600 frames of HD (1920×1080) video sequences. Table III shows that the encoding time of the proposed method is longer than the existing VBR method [8] by 14.6%, but this could not be problem for encoding videos for VOD applications. The encoding time of the CBR encoding [7] is larger than other methods because some frames are encoded more than one.

IV. CONCLUSION

In this paper, we have proposed a VBR video encoding method to maintain locally consistent picture quality of the encoded video with small buffering delay under limited bandwidth. For locally consistent picture quality of the encoded video, the frame-layer bit allocation problem is formulated as a constrained minimization problem so that an objective function which is defined as a sum of the local distortion variation of the encoded pictures and the underflow levels at the decoder buffer is minimized subject to a constraint on the local target number of bits within a temporal sliding window. Experimental results demonstrate that the proposed method provides a lower local standard deviation of PSNR by 29.2% comparing to the existing buffer-constrained VBR method [8] at the expense of a small increase of buffering delay. We plan to extend our method to the macroblock level rate control by utilizing object segmentation and tracking, for example.

ACKNOWLEDGMENT

The authors would like to thank Korean Broadcasting System (KBS) for providing HD video sequences used in the experiments.

REFERENCES

[1] *Program and System Information Protocol for Terrestrial Broadcast and Cable (Revision B)*, ATSC Standard A/65B, Advanced Television Systems Committee, 2003.
[2] A. Jagmohan and K. Ratakonda, "MPEG-4 One-pass VBR rate control for digital storage," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 5, pp. 447–452, May 2003.
[3] *Coding of Audio-Visual Objects—Part 2: Visual*, International Standard 14496-2, ISO/IEC, 2004, ISO/IEC JTC1/SC29/WG11 (MPEG).
[4] N. Kamaci, Y. Altunbasak, and R. M. Mersereau, "Frame bit allocation for the H.264/AVC video coder via cauchy-density-based rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 994–1006, Aug. 2005.
[5] D. Bagni, B. Biffi, and R. Ramalho, "A constant-quality, single-pass VBR control for DVD recorders," *IEEE Trans. Consum. Electron.*, vol. 49, no. 3, pp. 653–662, Aug. 2003.
[6] K. Wang and J. W. Woods, "MPEG motion picture coding with long-term constraint on distortion variation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 3, pp. 294–304, Mar. 2008.
[7] B. Xie and W. Zeng, "A sequence-based rate control framework for consistent quality real-time video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 56–71, Jan. 2006.
[8] M. Rezaei, M. Hannuksela, and M. Gabbouj, "Semi-fuzzy rate controller for variable bit rate video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 633–645, May 2008.
[9] Z. G. Li, L. Xiao, C. Zhu, and P. Feng, "A novel rate control scheme for video over the internet," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2002, vol. 2, pp. 2065–2068.
[10] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.
[11] J. Y. Lee and H. W. Park, "A rate control algorithm for DCT-based video coding using simple rate estimation and linear source model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 9, pp. 1077–1085, Sep. 2005.
[12] Z. Chen and K. N. Ngan, "Distortion variation minimization in real-time video coding," *Signal Process.: Image Commun.*, vol. 21, no. 4, pp. 273–279, Apr. 2006.
[13] L. J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 4, pp. 446–459, Aug. 1998.
[14] Y. Yu, J. Zhou, Y. Wang, and C. Chen, "A novel two-pass VBR coding algorithm for fixed-size storage application," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 345–356, Mar. 2001.

[15] H. Yin, X. Fang, L. Chen, and J. Hou, "A practical consistent-quality two-pass VBR video coding algorithm for digital storage application," *IEEE Trans. Consum. Electron.*, vol. 50, no. 4, pp. 1142–1150, Nov. 2004.

[16] B. Han and B. Zhou, "VBR rate control for perceptually consistent video quality," *IEEE Trans. Consum. Electron.*, vol. 54, no. 4, pp. 1912–1919, Nov. 2008.

[17] H. K. Arachchi and W. A. C. Fernando, "An intelligent rate control algorithm to improve the video quality at scene transitions for off-line MPEG-1/2 encoders," *IEEE Trans. Consum. Electron.*, vol. 49, no. 1, pp. 210–219, Feb. 2003.

[18] D. Zhang, Z. Chen, and K. N. Ngan, "Constant distortion rate control for H.264/AVC high definition videos with scene change," in *IEEE Int. Symp. Circuits Syst.*, May 2008, pp. 3198–3501.

[19] *Information Technology—Generic Coding of Moving Pictures and Associated Audio Information: Video*, International Standard 13818-2, ISO/IEC, 2000, ISO/IEC JTC1/SC29/WG11 (MPEG).

[20] *Coding of Audio-Visual Objects—Part 10: Advanced Video Coding*, International Standard 14496-10, ISO/IEC, 2004, ISO/IEC JTC1/SC29/WG11 (MPEG).

[21] 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.

[22] J. Dong and N. Ling, "On model parameter estimation for H.264/AVC rate control," in *IEEE Int. Symp. Circuits Syst.*, May 2007, vol. 1, pp. 289–292.

[23] W. Yuan, S. Lin, Y. Zhang, W. Yuan, and H. Luo, "Optimum bit allocation and rate control for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 6, pp. 705–715, Jun. 2006.

[24] C. H. Lee, Y. H. Jung, S. J. Lee, Y. J. Oh, and J. S. Kim, "Real-time frame-layer H.264 rate control for scene-transition video at low bit rate," *IEEE Trans. Consum. Electron.*, vol. 53, no. 3, pp. 1084–1092, Aug. 2007.

[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, pp. 68–78, Jan. 2007.

[26] C. Yu, J. Lu, X. Xie, and J. Zheng, "A novel content-based rate control algorithm for H.264," in *Proc. IEEE Int. Conf. Commun., Circuits Syst.*, Jul. 2007, vol. 1, pp. 611–615.

[27] H. Xiong, J. Sun, S. Yu, J. Zhou, and C. Chen, "Rate control for real-time video network transmission on end-to-end rate-distortion and application-oriented QoS," *IEEE Trans. Broadcast.*, vol. 51, no. 1, pp. 122–132, Mar. 2005.

[28] G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[29] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge University Press, 1992.

Hoonjae Lee received the B.S. degree in electronic engineering from the Korea University, Seoul, Korea, in 2006. He is currently working toward the M.S. and Ph.D. joint degree in electrical engineering at Korea University.

His research interests are image and video signal processing, digital broadcasting, and other issues on image and video technologies.

Sanghoon Sull (S'79–M'81) received the B.S. degree (with honors) in electronics engineering from the Seoul National University, Korea, in 1981, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology in 1983, and the Ph.D. degree in electrical and computer engineering from the University of Illinois, Urbana-Champaign, in 1993.

In 1983-1986, he was with the Korea Broadcasting Systems, working on the development of the teletext system. In 1994-1996, he conducted research on motion analysis at the NASA Ames Research Center. In 1996-1997, he conducted research on video indexing/browsing and was involved in the development of the IBM DB2 Video Extender at the IBM Almaden Research Center. He joined the School of Electrical Engineering at the Korea University as an Assistant Professor in 1997 and is currently a Professor. His current research interests include image/video processing, computer vision, video search, and their applications to 3D TV and smart phone. Dr. Sull is currently an Associate Editor for IEEE Transactions on Circuit and System for Video Technology.