Window-Level Rate Control for Smooth Picture Quality and Smooth Buffer Occupancy

Long Xu, Debin Zhao, Xiangyang Ji, Lei Deng, Sam Kwong, *Senior Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*—In rate control, smooth picture quality and smooth buffer occupancy are both important but contrary to each other at a given bit rate. How to get a good tradeoff between them was not devoted much attention previously. To deal with this problem, a theoretical window model is proposed in this paper, in which several adjacent frames grouped as a window are considered together. The smoothness of both picture quality and buffer occupancy can be gracefully achieved by regulating the size of the window. To illustrate the usage of window model, a window-level rate control algorithm cooperated with the traditional ρ-domain rate-distortion model is further introduced. In experiments, we first show how the proposed window model achieves the tradeoff between picture quality smoothness and buffer smoothness, and then demonstrate the significant PSNR improvement, accuracy of bit control and consistency of visual quality of the proposed window-level rate control algorithm.**

***Index Terms*—Rate control, smooth buffer occupancy, smooth picture quality, video coding, window-level rate control, window model.**

I. Introduction

With the boom of multimedia entertainment, various digital videos from the Internet and digital storage media have been beyond the capacity of transmission and storage. Video coding has become increasingly popular and important. Different video coding standards such as MPEG-1,2,4 [1], [2], [5], H.261,2,3,4 [2]–[5], and AVS [6] have been promoted for various applications, such as video communications, standard and high definition TV, digital storage and video streaming over Internet.

Rate control plays an important role in video coding for streaming application. It aims to achieve the best perceptual video quality under certain constraints, such as bandwidth, decoding delay, buffer capacity and computational complexity. Rate control is usually classified into constant bit rate (CBR) control and variable bit rate (VBR) control. To maintain a constant short term average bit rate, CBR adopts the uniform bit allocation among different coding units irrespective of individual picture's characteristic, which usually leads to the fluctuation of perceptual video quality and low coding efficiency. On the contrary, VBR is designed to fluctuate its output bit rate according to the varying characteristics of video content in order to achieve consistent video quality in a long term. In VBR, we need to preanalyze video sequence to capture the variation of video content. The target bits will be distributed to various segments, allocating more bits to complex scene and less bits to simple scene. Traditionally, VBR is generally implemented by two-pass or multipass encoding processes, which require long time delay and high computational complexity, and could not be used in real-time encoding.

Typical CBR algorithms include TM5 [7] for MPEG-2, TMN8 for H.263 [8]–[10] and VM8 for MPEG-4 [11]. They all employ the R-D models for QP decision. The key of them is to find the relation between the rate and the distortion, namely rate-distortion (R-D) model, which is the function with respect to quantization parameter (QP). Usually the quantized DCT coefficients are assumed to be Laplacian distribution [13]–[17]. As a result, the classical LOG R-D model can be derived, whose Taylor expansion is usually used as its approximation in linear and quadratic R-D models. For example, in VM8 of MPEG-4 and H.264/AVC reference codec, the quadratic R-D model [12] is employed. In TMN8 and TMN12 of H.263, the MB-level R-D model [10] is used. Besides, in ρ-domain R-D model [18], the rate is directly represented by the percentage of zero coefficients after quantization. Conventionally, bit allocation should be performed before QP decision for allocating the target bits quota for each frame/MB. Then QP of each frame/MB can be decided according to the R-D model. In TM5 of MPEG-2, the same bits quota is assigned to each GOP, and then frame-level bit allocation is performed according to the anticipated complexity and characteristic of each picture. Such a bit allocation strategy is also used in the rate control [19]–[21] of H.264/AVC. Moreover, the hypothetical reference decoder (HRD) is additionally considered in the bit allocation for further regulating the bits quota of each frame to avoid the buffer overflow and underflow.

For QP decision of H.264/AVC, there is an inherent dilemma when rate control and RDO [24], [25] are both enabled. The mean absolute distortion (MAD) of each MB/frame should be provided beforehand. However, it is only available after the MB/frame has been coded. Therefore, the predicted MAD is used instead of the actual MAD. When scene changes or high motion occurs, the predicted MAD will be much more different

Manuscript received January 21, 2010; revised May 08, 2010; accepted July 20, 2010. Date of publication August 05, 2010; date of current version February 18, 2011. This work was supported in part by the Major State Basic Research Development Program of China under 973 Program 2009CB320905, the National Science Foundation of China under Grant 60736043, and the Hong Kong RGC General Research Fund (GRF) 9041353 (CityU 115408). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhou Wang.

L. Xu and S. Kwong are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail: lxu222@cityu.edu.hk; CSSAMK@cityu.edu.hk).

D. Zhao is with the Department of Computer Science, Harbin Institute of Technology, Harbin 150001, China (e-mail: dbzhao@vilab.hit.edu.cn).

X. Ji is with the Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: xyji@tsinghua.edu.cn).

L. Deng and W. Gao are with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: ldeng@jdl.ac.cn; wgao@jdl.ac.cn).

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/TIP.2010.2063708

from the actual one. When it is used to calculate QP by the given R-D model, it will result in a big QP variation or significant fluctuations of picture quality. In [26]–[28], the improvements on the smooth picture quality are achieved by introducing the efficient spatio-temporal MAD prediction. In [29], a low-pass filter is employed to smooth the distortion of frame encoded for MPEG4 video encoder. In [30], the authors propose a sequence-based frame-level bit allocation model to track the nonstationary characteristics in the video source. In [31], the picture quality variation and buffer fluctuation are formulated into an optimization problem to get as smooth as possible picture quality while maintaining smooth buffer. Besides, in JM codec, a series of "bound" and "clip" techniques are provided for controlling picture quality and buffer occupancy.

The methods mentioned previously cannot provide an explicit and analytic way to obtain the smooth picture quality while meeting the buffer constraint. Meanwhile, the complexity of future frames and global statistics is not available before encoding, so the flaw of traditional bit allocation with the assumption of stationary scene cannot be overcome. Furthermore, the existing buffer control mechanism cannot make full use of the available buffer for smoothing picture quality of coded pictures. Therefore, a theoretical model which can tackle the tradeoff between picture quality and buffer occupancy, and efficient bit allocation scheme are expected. In our previous work [32], a window model was proposed to tackle such a tradeoff, where the effect of buffer constraint is expressed as an exponential formula empirically. However, the theoretical analysis and experiments are not enough. In addition, the tradeoff is not well investigated and validated. The work in this paper is an extension of [32].

This paper proposed a theoretical window model, in which several adjacent frames grouped as a window are considered together. The smoothness of both picture quality and buffer occupancy can be gracefully achieved by regulating window size. In addition, a window-level rate control algorithm cooperating with the ρ-domain R-D model [18] is provided. As the proposed rate control scheme operates on window level, in which the bits quota of a window is determined according to the average bit rate, the traditional bit allocation which performs frame by frame on the predicted complexity of a frame is not required anymore.

The rest of this paper is organized as follows. In Section II, we formulate a window model to reflect the relation among window size, QP variation and bits variation. The window-level ρ-domain rate control algorithm is detailed in Section III. Section IV provides the experimental results and discussions. A brief conclusion is given in the last section.

II. Window Model

Rate control consists of bit allocation and QP decision. Via considering buffer status in bit allocation, the coded bitstream conforms to both the target bit rate and buffer constraint. Rate control should guarantee a small gap between the target bit rate and the actual coded bit rate, and achieve the average PSNR improvement. On the other hand, the picture quality smoothness and buffer smoothness are also very important in evaluating the efficiency of a rate control algorithm. Generally, the smooth picture quality can give the spectators temporally consistent visual experience, which usually requires two-pass or multipass encoding [32]–[38]. The smooth buffer occupancy makes the fluent transmission and decoding of bitstream under the constraint of the application requirements, or the capacities of transmission channel and decoding devices. In video streaming applications, violating buffer constraint may cause annoying jitter.

In the traditional rate control algorithms, the anticipated complexity and anticipated MAD of future frames are used for bit allocation and QP calculation, respectively [19], [20]. These methods work well for the low motion object and similar scenes among the adjacent frames. However, they may result in the significant fluctuations of picture quality and buffer occupancy when scene changes or high motion object occurs. In such situations, the abrupt variation of buffer occupancy usually occurs due to big bits fluctuation. To avoid the buffer overflow and underflow, we need to further adjust the bits quota of a frame, which may also arouse the significant deterioration of picture quality. Moreover, in the traditional rate control schemes, there is no theoretical model on how to properly make a good tradeoff between picture quality and buffer occupancy in the encoding process to meet the given requirements.

To tackle the problem mentioned previously, a theoretical window model is proposed, in which the bounds on QP variation (ΔQ) and bits variation (ΔR) are considered for efficiently controlling the picture quality smoothness and buffer smoothness. Conventionally, there is usually a big QP variation between frames in the traditional CBR rate control on frame-level. As an alternative, within a given buffer delay, several frames can be grouped together for bit allocation and QP calculation at each time. As a result, the proposed window model should find the optimal relation between the number of the adjacent frames in a window, denoted as L, ΔQ and ΔR to keep the good tradeoff between picture quality smoothness and buffer smoothness in rate control. For simplicity, ΔQ is only used to represent the variation of picture quality under the assumption of a single scene. For the multiple scenes, ΔQ can be extended according to [33], [34]. The buffer controlling can be realized by regulating the bits of each frame. Generally, the less the bits variation is, the less the buffer size is. Thus, it is reasonable to use the bits variation (ΔR) to represent buffer fluctuation.

The relation between L and ΔQ can be investigated from the Law of Large Numbers (LLN) [41], [42], where the number of samples and error bound are corresponding to L and ΔQ, respectively. The classical Chebyshev LLN assumes $\xi_1, \xi_2, \ldots, \xi_n$ are random variables independent from each other, and the variances of them are finite, i.e., $\exists C > 0$, $\neg D\xi_i \leq C,\ i = 1, 2, \ldots, n$. Then, $\forall \varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \frac{1}{n}\sum_{i=1}^{n}E\xi_i\right| < \varepsilon\right) = 1. \tag{1}$$

C is a constant, and E and D represent mathematical expectation and variance operators, respectively. Assuming that QP variation conforms to a normal random variable ξ, the QP variation of each frame is a sample from ξ denoted as $\{\xi\}$ $(i = 1, 2, \ldots, n)$ which are independent from each other. Applying LLN to $\{\xi\}, 1/n\sum_{i=1}^{n}\xi_i$ converges at $E\xi$. The measure of such convergence can be further explained explicitly by the Center Limit Theory (CLT). CLT assumes $\xi_1, \xi_2 \ldots, \xi_n$ are a group of identically distributed random variables, $E\xi_k = a$, $D\xi_k = \sigma^2(\sigma^2 > 0)$, $k = 1, 2, \ldots, n$, then

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n}\xi_k - na}{\sigma\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt. \quad (2)$$

That is to say, the distribution of $\sum_{i=1}^{n}\xi_k$ converges at a normal variable $N(na, \sigma\sqrt{n})$. From (2), we get (3), shown at the bottom of the page. Obviously, it shows both the convergence and how to converge, where the sample size n and error bound ε correspond to the window size L and ΔQ, respectively.

A. L-ΔQ Model

Suppose the bounds of QP variation and bits variation required in encoding are ΔQ and ΔR, respectively, represented by two Gaussian random variables $\xi(\omega)(\mu_\xi = 0, \sigma_\xi^2)$ and $\eta(\omega)(\mu_\eta = 0, \sigma_\eta^2)$. According to CLT, the average of $\{\xi_k(\omega)\}(k = 1, 2, \ldots, n)$ converges at a Gaussian random variable. From (3), we can get

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \mu_\xi\right| < \Delta Q\right) = \frac{1}{\sqrt{2\pi}}\int_{-\Delta Q(\sqrt{n}/\sigma_\xi)}^{\Delta Q(\sqrt{n}/\sigma_\xi)} e^{-t^2/2}dt. \quad (4)$$

Let the right part of (4) equals p, the integral limit of (4) is obtained by looking up the standardized normal table [42]. Assuming $\Delta Q(\sqrt{n}/\sigma_\xi) = \sqrt{\alpha}$, replacing n by L, then the L-ΔQ relation can be obtained as

$$L = \alpha\frac{\sigma_\xi^2}{\Delta Q^2} \quad (5)$$

where L is the minimum window size that conforms to ΔQ.

B. $\Delta R - \Delta Q$ Model

The relation between L and ΔQ is formulated by L-ΔQ model. Another criterion ΔR can be introduced with the help of $R - Q$ model. In our consideration, performing differential operation on the classical LOG $R - Q$ model

$$R = \log\left(\frac{2e}{\lambda Q}\right) + C \quad (6)$$

where R is the bit rate, C is a constant, then we can get the relation between the differential $Q(dQ)$ and differential $R(dR)$, i.e., $dR = -\beta/QdQ$. Thus, the bits variation caused by ΔQ is calculated by

$$\delta R = -\frac{\beta}{Q}\Delta Q \quad (7)$$

where β is a model parameter updated for each window, the negative means the inverse relation between δR and ΔQ. If δR does not exceed the given bound of bits variation ΔR, the window size L from $L - \Delta Q$ model will be the final decision which conforms to both ΔQ and ΔR; otherwise, such L should be regulated to make sure δR is less than ΔR. Equation (7) is applicable to not only the same basic coding unit but also a group of different basic coding units. With (5) and (7), the window model can be deduced finally in the following subsection.

C. Window Model

Obviously, as small as possible QP variation and bits variation which represent the smooth picture quality and small buffer size/delay, respectively, are preferred. However, they cannot be achieved concurrently due to the contradictions between them. In our consideration, the bounds of QP variation and bits variation ΔQ and ΔR are given as two requirements of encoding. They are contradictory to each other, i.e., the big bits fluctuation will arise generally if a small ΔQ is required, and vice versa. Thus, the minimum bits fluctuation: δR can be calculated as $\delta R = \beta/Q\Delta Q$ from (7) under given ΔQ.

We first get the minimum window size L from $L - \Delta Q$ model (5). Then if $\delta R \le \Delta R$, the window size L from $L - \Delta Q$ model will be the final decision which conforms to both $\Delta Q > \Delta R$ and ΔR; otherwise, such L should be regulated to make sure δR is less than ΔR. Conventionally, the video content of a sequence varies along time. Thus, for a window if the smooth QPs are assigned, the larger L is, the larger δR is. Conversely, the smaller L is, the smaller δR is. Therefore, L should be decreased if $\delta R > \Delta R$. Using ΔR instead of δR in (7), then we can obtain $\Delta Q = Q/\beta\Delta R$. Substituting it into (5), the window model is finally derived as

$$L = \begin{cases} \alpha\frac{\sigma_\eta^2}{\Delta Q^2}, & \text{if } \frac{\beta}{Q}\Delta Q \le \Delta R \\ \frac{\alpha\sigma_\eta^2}{\left(\frac{Q}{\beta}\Delta R\right)^2}, & \text{otherwise.} \end{cases} \quad (8a)$$

$$\begin{aligned} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - a\right| < \varepsilon\right) &= P\left(\frac{\sum_{k=1}^{n}\xi_k - na}{\sigma\sqrt{n}} < \varepsilon\left(\frac{\sqrt{n}}{\sigma}\right)\right) - P\left(\frac{\sum_{k=1}^{n}\xi_k - na}{\sigma\sqrt{n}} < -\varepsilon\left(\frac{\sqrt{n}}{\sigma}\right)\right) \\ &\approx \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\varepsilon(\sqrt{n}/\sigma)} e^{-t^2/2}dt - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\varepsilon(\sqrt{n}/\sigma)} e^{-t^2/2}dt \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\varepsilon(\sqrt{n}/\sigma)}^{\varepsilon(\sqrt{n}/\sigma)} e^{-t^2/2}dt \xrightarrow{n\to\infty} 1 \end{aligned} \quad (3)$$

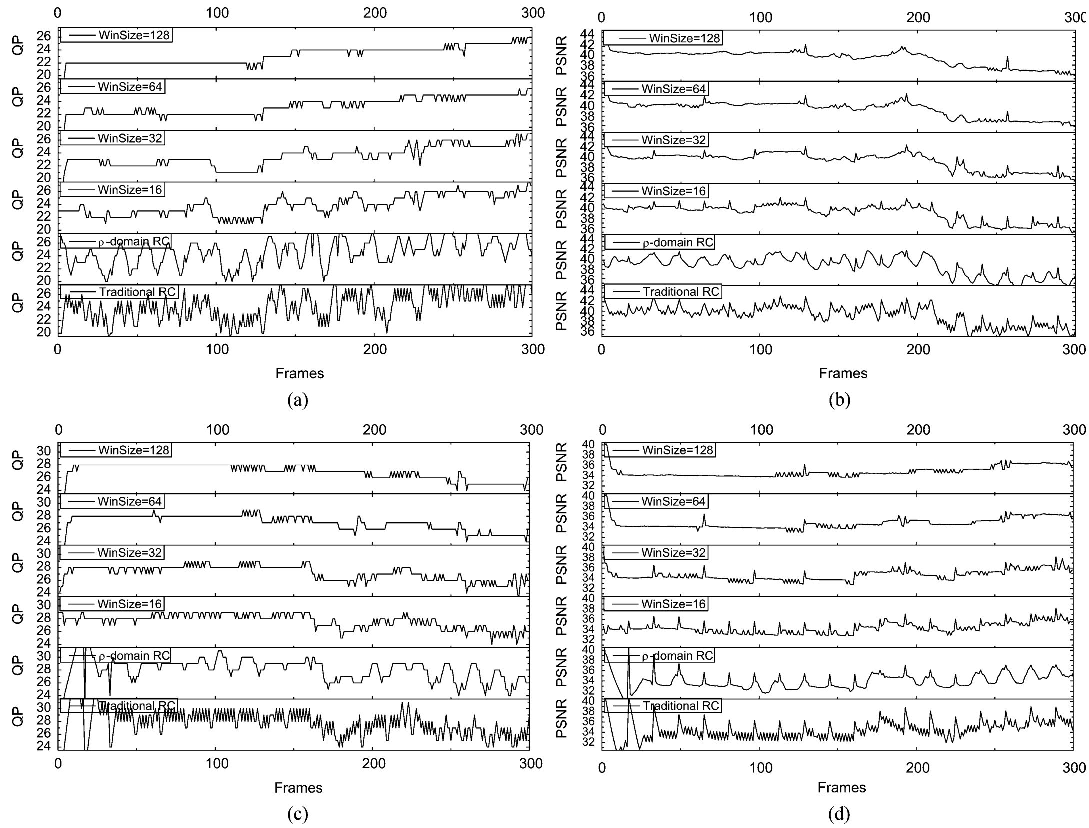


Fig. 1. Comparisons of frame QP/PSNR among the proposed and benchmarks (Foreman: 1000 kb/s, Mobile: 2000 kb/s, 30 Hz). (a) Frame QP for "Foreman." (b) Frame PSNR for "Foreman." (c) Frame QP for "Mobile." (d) Frame PSNR for "Mobile."

Equation (8a) can be expressed by a simple format as

$$L = \alpha \frac{\sigma_\eta^2}{\Delta Q^2} \min \left\{ 1, \left(\frac{\left(\frac{\frac{\beta}{Q}}{\Delta Q} \right)}{\Delta R} \right)^2 \right\} \tag{8b}$$

where the model parameters which α and β are manually initialized and adaptively updated for each window.

From (8), the window model is built on the tri-parameters $(L,\ \Delta Q,\ \Delta R)$. If ΔQ and ΔR are predefined for controlling picture quality smoothness and buffer smoothness, the minimum window size L can be obtained from (8) resulting in the minimum encoding delay. Then, the encoding process on the window is implemented so that the encoding results are subject to the given ΔQ and ΔR. Another typical application of window model is to get as smooth as possible picture quality if ΔR and L are predefined for buffer smoothness and startup delay of encoding. From (5), the QP variation can be expected as $\delta Q = \sqrt{\alpha \sigma_\eta^2 / L}$ if the bits variation bound (ΔR) is not considered. From (7), $\delta Q = -Q/\beta \Delta R$ if bits variation bound (ΔR) is required. Thus, the QP variation is finally deduced from (5) and (7) given L and ΔR as follow:

$$\begin{cases} \delta Q = \sqrt{\alpha \frac{\sigma_\eta^2}{L}}, & \text{if } \Delta R \geq \frac{\beta}{Q} \delta Q \\ \delta Q = \frac{Q}{\beta} \Delta R, & \text{otherwise.} \end{cases} \tag{9}$$

From (9), given L and ΔR, we firstly predict the QP variation (δQ) from the upper equation of (9). According to (7), the bits variation (δR) corresponding to δQ is $\delta R = \beta / Q \delta Q$. If $\Delta R \geq \delta R$, then, the final QP variation is decided by the upper equation of (9). Otherwise, the final QP variation is predicted from the lower equation of (9).

The large enough buffer is usually offered to VBR applications, so L can be set to the total number frames of a sequence. Thus, a constant QP for all frames can be deduced. If only one frame is in a window, the traditional CBR case is considered. In video coding applications, the smoothness of picture quality can be represented by QP smoothness under the assumption of a single scene. As for buffer smoothness, it must be under the constraint of HRD for the conformance of bitstream in H.264/AVC, where the minimum buffer size and initial buffer fullness are provided for the bitstream compressed.

III. Window-Level ρ-Domain Rate Control

The window model is a theoretical model mapping the requirements on picture quality smoothness and buffer smoothness to the window size which can be dynamically regulated in the encoding process. It tells us that the best efficiency subject to these requirements can be obtained in theory. To achieve such efficiency, we should further provide the corresponding rate control algorithm on the proposed window model. In fact, any frame-level R-D model can be employed in the proposed window model. In this paper, we propose an efficient rate control algorithm by employing the ρ-domain R-D model [18]. In ρ-domain model, the percentage of zero coefficients quantized

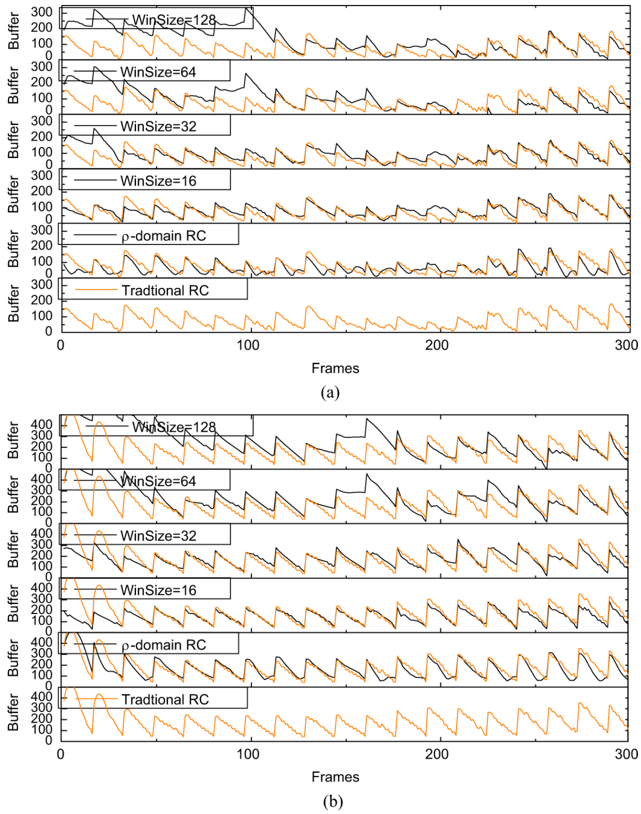


Fig. 2. Comparisons of buffer fluctuation among the proposed and benchmarks (Foreman: 1000 kb/s, Mobile: 2000 kb/s, 30 Hz). (a) "Foreman." (b) "Mobile."

(ρ) is directly related to the bit rate (R) as $R = \theta(1-\rho)$. The conventional R-D model defined the relationship between bit rate and QP are named Q-domain model, such as the linear and quadratic R-D models in [7] and [12]. The ρ-domain model can be established immediately after a preanalysis with only 16×16 motion estimation and DCT transform, so the complexity increases slightly when compared to the traditional one-pass encoding. In addition, it is a very simple linear function which is of high accuracy in bit control error [18], while the Q-domain model has a complex nonlinear behavior, which makes it hard to develop an accurate and robust model in Q-domain. Therefore, the ρ-domain model is chosen for the implementation of proposed window model.

Employing ρ-domain model, the preanalysis on a single mode, usually 16×16 inter mode [18] is required. But it is very simple and of low complexity. So the preanalysis is usually applied in the single-pass encoding [18]. Our algorithm is also a single-pass processing, it is with much less complexity comparing with the two-pass processing firstly. Second, the window size in our method usually ranges from 32 to 64, whose delay is about 1 s to 2.5 s. Such a delay is admissible in real-time coding. Third, the window size is under the controlling of buffer constraint and dynamically regulated. According to CAT-LK [40], the initial encoder delay can be utilized into our proposed window-level processing without additional time delay introduced. While the traditional two-pass method usually performs encoding twice on the whole sequence with the full-complexity operations of encoder.

TABLE I
PROPOSED WINDOW-LEVEL ρ-DOMAIN RATE CONTROL ALGORITHM

Step 1: *If* the window is the first one, initializing window size L=L_0, L_0 is the initial window size; *else*, updating L from the latest α and β according to (8);

Step 2: Getting the total bits of the current window $B=(r/f)\times L$;

Step 3: Pre-analysis of only 16x16 inter prediction for P and B frames and 16x16 intra prediction for I frames in a given window;

Step 4: Building ρ-QP tables for each frame in this window and ρ-QP tables for this window according to (10)-(12);

Step 5: Computing QP of the *j*-th frame by $B_j=\theta_j(1-\rho_j)$ and (12), where B_j is the remaining bits, θ_j is updated for each frame. ρ_j is firstly calculated by $\rho_j=1-(B_j/\theta_j)$, and then, QP is obtained by looking up table (12);

Step 6: Performing encoding for *j*-th frame with the obtained QP of Step 5;

Step 7: Updating θ as $\theta_{j+1}=b_j/(1-\rho_j)$, where b_j is the coding bits of frame *j*. $B_{j+1}=B_j-b_j$. The window-level ρ-QP table is updated by subtracting the ρ-QP table of frame j. If the last frame of current window is reached, go to Step 8; else go to Step 5;

Step 8: Updating window parameters α and β from (5) and (7) as:

$$\alpha = {L\times\delta Q^2}/{\sigma_Q^2} \quad \text{and} \quad \beta = -{\delta R\times\bar{Q}}/{\delta Q},$$

where $\sigma_Q^2=\sum_{j=0}^{L-1}(Q_j-\frac{1}{L}\sum_{j=0}^{L-1}Q_j)^2$, $\bar{Q}=\frac{1}{L}\sum_{j=0}^{L-1}Q_j$. δQ and δR represent the ranges of QP and bits variations respectively;

Step 9: *If* the sequence ends, terminates procedure; *else* go to Step 2.

A. M-QP Table

Employing ρ-domain model, the relation between ρ and QP (named ρ-QP table) should be provided beforehand, which is usually obtained by a preanalysis process. For building ρ-QP table, the quantization is performed many times for each DCT coefficient to find the minimum QP which can quantize the coefficient into zero. Such processing is usually of high computational complexity. Actually, we can establish a magnitude-QP (M-QP) table beforehand, where any possible magnitude of DCT coefficients is mapped into a minimum QP which can quantize the DCT coefficient to zero. Thus, in the stage of building ρ-QP table, the QP for each coefficient can be obtained immediately by looking up the M-QP table.

The popular transform-based coding standards employ integer DCT, from which the transformed coefficients should be normalized/scaled before quantization due to the nonnormalized scaling from float DCT to integer DCT. For a 4×4 DCT coefficient of H.264/AVC, there are three possible quantized values according to its position in the 4×4 matrix [5]. Thus, we should provide three M-QP tables for each 4×4 DCT coefficients. Given a coefficient y, the minimum QP which quantize it into zero can be obtained by

$$\text{Table}_{i,j}(y) = QP_t \; if(Q\text{step}_{t-1} \le |y| < Q\text{step}_t) \qquad (10)$$

where $t=0,1,\ldots,51$ and $Qstep_{-1}=0$, $Qstep$ is quantization step which is indexed by QP in conventional coding system, $(i,\ j)$ represents the coordinate in MB. $Table_{i,j}(y)$ is a monotonically increasing function ranging from the minimum QP to the maximum QP.

TABLE II
QP AND BUFFER VARIATIONS WITH THE ADAPTIVE WINDOW SIZE($\Delta Q = 2$; 1000 kb/s FOR "FOREMAN" AND "LINKER", 2000 kb/s FOR "MOBILE")

Sequence	Buffer delay (Second)	Percentage of each window size (%)					STD of QP	Buffer fullness		Performance	
		=16	[16,32]	[32,48]	[48,64]	[64,80]		MAX/MIN(%)		PSNR/bit rate	
Mobile	0.18	100.00	0.00	0.00	0.00	0.00	1.45	122	67	34.33	2000.41
	0.20	80.00	20.00	0.00	0.00	0.00	1.42	98	53	34.30	2000.77
	0.30	40.00	40.00	20.00	0.00	0.00	1.30	81	44	34.28	2000.48
	0.40	20.00	20.00	10.00	10.00	40.00	0.90	72	33	34.25	1999.87
Foreman	0.18	100.00	0.00	0.00	0.00	0.00	1.36	110	49	39.19	997.63
	0.20	40.00	40.00	20.00	0.00	0.00	1.19	101	44	39.09	996.47
	0.25	16.70	16.70	50.00	16.70	0.00	1.11	80	35	39.01	996.71
	0.30	16.70	16.70	16.70	50.00	0.00	0.87	73	29	38.99	996.16
Linker	0.40	30.77	17.23	14.77	24.62	12.31	1.40	104	29	36.09	1000.68
	0.50	2.23	3.46	4.69	10.85	78.86	1.15	96	17	35.71	994.48
	0.60	1.23	4.92	0.00	14.77	80.00	1.09	97	15	35.69	990.99
	0.80	0.45	2.07	0.00	7.02	91.31	1.06	78	11	35.73	1000.44

B. Window-Level ρ-QP Table

In [18], the preanalysis with only 16×16 inter mode is performed on P frames for establishing the ρ-QP table by looking up the corresponding M-QP table according to the coordinate of the coefficient in MB as

$$\rho_m(QP) = \sum_{y \in \text{MB}} \{if(\text{Table}_{i,j}(|y|) < QP)\}. \quad (11)$$

From (11), the ρ-QP table for each MB is progressive, i.e., the coefficients quantized into zero by a smaller QP must be zero by a bigger QP. After obtained ρ-QP tables of all MB, a frame-level ρ-QP table can be obtained by the sum of all MB's ρ-QP tables, i.e., $\rho_f(QP) = \sum_{m=0}^{m=N-1} \rho_m^{MB}(QP)$. Given the ρ-QP tables of all frames $\{\rho_f(QP), f = 0, 1, \ldots, L-1\}$ in a window, the window-level ρ-QP table is obtained by the sum of all frame's ρ-QP tables, i.e.,

$$\rho_w(QP) = \sum_{f=0}^{f=L-1} \rho_f(QP). \quad (12)$$

C. Window-Level ρ-Domain Rate Control

Assuming target bit rate is r (bits/s), frame rate is f, our proposed algorithm is described in detail as follows.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

We implemented the proposed window-level ρ-domain rate control algorithm on JM11.0 of H.264/AVC under the conditions: *Profile/Level: 100/40, Reference frames: 2, Full search*, Search range: *16, RDO on* and *CABAC, IPPP* encoding structure. The following experiments will be arranged into three parts: the first one is for the validation of the proposed window model, and the second one is for exhibiting the coding efficiency of the proposed algorithm and the third one is for encoding time-delay reduction.

A. Validation of Window Model

From discussion in Section II about window model, picture quality smoothness and buffer smoothness are contrary to each other at a given bit rate. In addition, the tradeoff between them

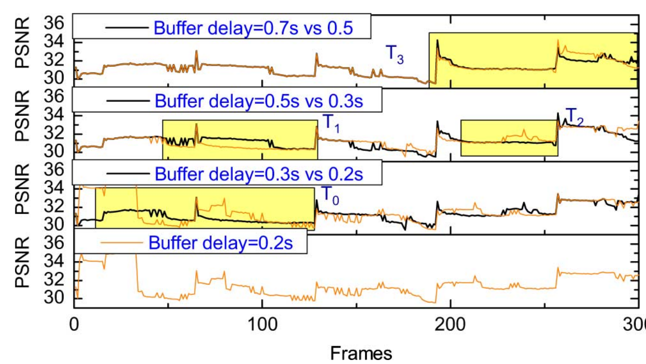


Fig. 3. Frame PSNR curve and buffer delay with the adaptive window size for "Mobile" (1000 kb/s, 30 Hz).

can be achieved by the proposed window model (8). We firstly perform experiments on CIF sequences "Foreman" and "Mobile" to verify the contrary relationship between visual quality smoothness and buffer smoothness. By changing the window size, the algorithm proposed in Section III can achieve different smoothness of visual quality as shown in Fig. 1, where the larger window is, the smoother visual quality is. However, the corresponding buffer shown in Fig. 2 behaves inversely to visual quality smoothness. The experiments also validate (9) which is an extension of window model (8), i.e., the larger window is, the smoother QP variation, and vice versa.

Second, we compare the proposed algorithm with the state-of-the-art algorithms. JVT-H017r3 [19], [20] on frame-level and ρ-domain algorithm [18] are implemented to be benchmarks. From Fig. 1, the proposed algorithm produces much smoother QP/PSNR than benchmarks, especially for window size of 64 and 128, but large buffer fluctuation exists concurrently. In addition, the dramatic buffer fluctuation can be observed at the start of buffer curves in Fig. 2 for benchmarks, due to the bad initial QP for both of them. While the proposed algorithm can produce a good initial QP by importing I frame into R-D model. Most importantly, the buffer smoothness of the proposed algorithm is almost the same as those of benchmarks as window size is not bigger than 64. That means the proposed algorithm can achieves much better visual quality than the benchmarks under the same buffer resource. In Fig. 2,

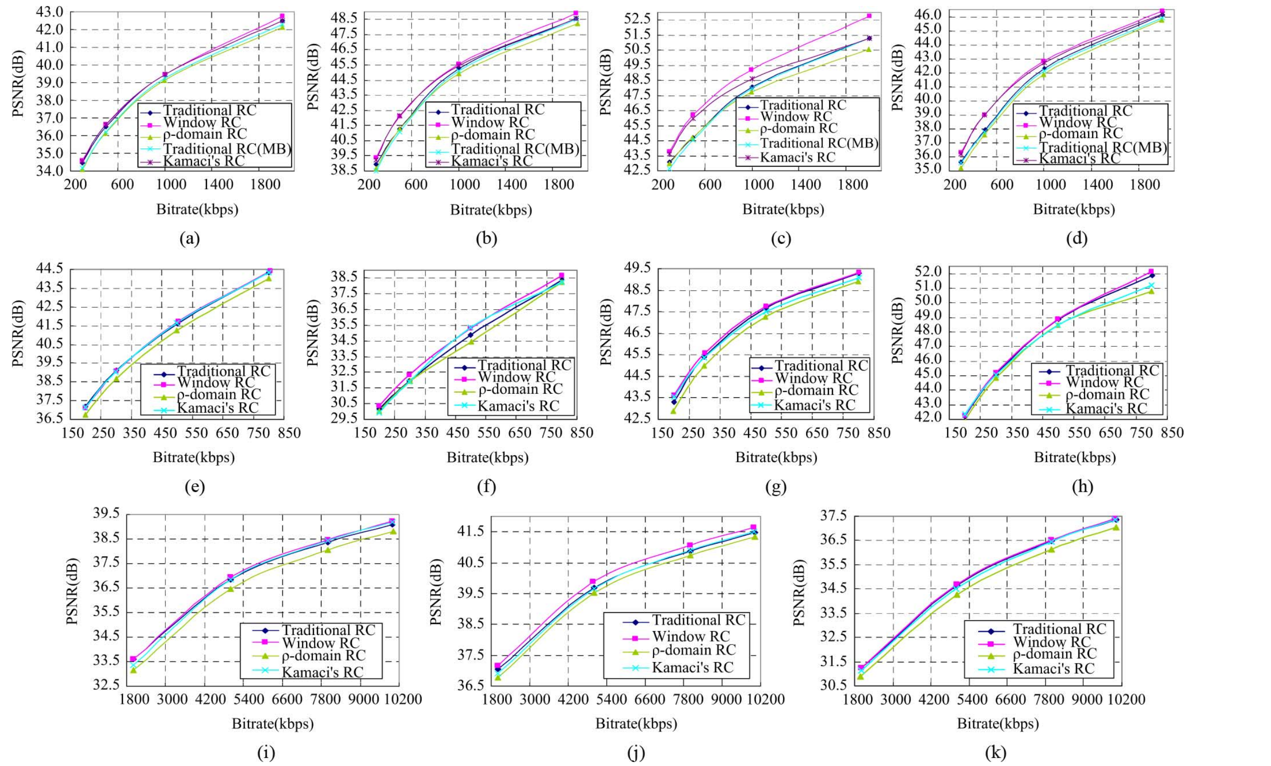


Fig. 4. Comparison of R-D performance between the proposed algorithm and benchmarks. (a)-(d) CIF sequences. (e)-(h) QCIF sequences. (i)-(k) 720P (1280×720) HD sequences). (a) "Foreman." (b) "News." (c) "Akiyo." (d) "Silent." (e) "Foreman." (f) "Football." (g) "Grandma." (h) "News." (i) "Night." (j) "Crew." (k) "Harbour."

the buffer curve of traditional algorithm is drawn additionally along all other curves for comparison.

The experiments with adaptive window size are further performed on CIF sequences "Foreman," "Mobile," and "Linker" ("Linker" consists of five standard CIF sequences "Foreman," "Football," "Tennis," "News," and "Silent" in order, 1300 frames in total) to show the good tradeoff between visual quality smoothness and buffer smoothness. Such tradeoff is achieved by choosing a proper window size derived from the proposed window model (8). In Section III, ΔR is used in window model instead of buffer capacity because it can be easily imported by considering R-Q model. However the buffer usage can be easily detected during the actual encoding process, so the buffer constraint is taken as a requirement instead of bits fluctuation cooperating with ΔQ in the following experiments. The experiments conditions are as follows: L ranges from 16 to 80 and is initialized to 16, QP variation bound $\Delta Q = 2$, and buffer constraint is in the column labeled "Buffer delay" in Table II.

The experimental results are listed in Table II, where the larger the buffer is, the larger the percentage of large windows is, accordingly the less QP variation represented by its standard deviation (STD) is. Taking "Linker" as an example, the largest number of windows chooses the largest window size of 80 given the largest buffer delay of 0.8 s, and QP varies least at the same time among all cases. We also notice that buffer overflows if too small buffer delay is considered, such as 0.18 s delay for "Foreman" and "Mobile," due to the restriction of minimum window size of 16 in our simulating system. Such observation accords with window model. In addition, the adaptive window size is achieved by choosing the window sizes as Step.8 of Table I, where α and β are updated with the bit used and QP used after coding each window. Then a new window size can be deduced from (8). The bit used is also used to update current buffer fullness for further clipping window size. If the current buffer is at a high-risk of overflow, the less window size than the previous one is chosen. Conversely, if buffer drops to a low level, we will retrieve the previous window size to make full use of buffer resource for smooth visual quality. Furthermore the R-D performance tends to decrease when more windows chose large size. The reason is that the uniform quantization on multiple scenes with much different complexity will degrade the overall R-D performance, which will be discussed in detail in the next subsection. Therefore, the typical window size ranging from 32 to 64 is recommended in our algorithm. At such time interval of a window, the single scene is assumed in our consideration.

In addition, the tradeoff between visual quality and buffer constraint is explained visually by the use of frame PSNR curve of "Mobile" in Fig. 3, where the segments shadowed (marked by "$\mathrm{T}i (i = 0, 1, 2, 3)$") show the obvious visual quality improvements when the buffer delay increases. The window sizes are initialized to 16 for all cases, and then they are updated dynamically after encoding each window.

B. Performance of the Proposed Rate Control Algorithm

The comparisons are performed between the proposed window-level algorithm and three benchmark algorithms, including the traditional one (JVT-H017r3) [20], ρ-domain [18] and Kamaci's algorithm [43]–[45]. Table III summarizes the

TABLE III
PERFORMANCE COMPARISON FOR ALL TESTING ALGORITHMS IN TERMS OF BIT CONTRL ERROR AND PSNR GAIN

Resolution	Sequence	Target bit rate (kbps)	Traditional rate control			ρ-domain rate control			Kamaci's rate control			Proposed rate control		
			Bit rate (kbps)	PSNR	Error (%)	Bit rate (kbps)	PSNR	Error (%)	Bit rate (kbps)	PSNR	Error (%)	Bit rate (kbps)	PSNR	Error (%)
CIF	Foreman	2000	2001.77	42.49	0.09	2000.83	42.14	0.04	1998.99	42.48	-0.05	1999.94	42.77	0.00
		1000	1003.43	39.45	0.44	998.83	39.12	-0.02	1001.38	39.46	0.24	999.84	39.45	0.08
		500	501.77	36.48	0.35	497.82	36.13	-0.44	501.00	36.62	0.20	500.26	36.62	0.05
		300	303.08	34.46	1.03	299.53	34.08	-0.16	300.90	34.55	0.30	301.44	34.60	0.48
	News	2000	2001.70	48.57	0.09	2010.90	48.20	0.55	1999.43	48.55	-0.03	1999.81	48.91	-0.01
		1000	1001.16	45.24	0.22	1000.83	44.93	0.18	999.67	45.45	0.07	1000.10	45.57	0.11
		500	501.66	41.29	0.33	502.76	41.29	0.55	500.12	42.16	0.02	500.29	42.11	0.06
		300	301.11	38.92	0.37	301.90	38.67	0.63	300.06	39.21	0.02	300.30	39.38	0.10
	Akiyo	2000	1999.22	51.27	-0.04	1995.74	50.58	-0.21	2000.13	51.29	0.01	1999.32	52.77	-0.03
		1000	1000.22	48.05	0.12	994.81	47.74	-0.42	999.97	48.59	0.10	998.03	49.20	-0.10
		500	500.50	44.69	0.10	497.03	44.72	-0.59	500.84	45.96	0.17	499.21	46.21	-0.16
		300	301.02	43.10	0.34	297.66	42.97	-0.78	300.49	43.60	0.16	300.47	43.76	0.16
	Silent	2000	2001.88	46.15	0.09	1994.40	45.76	-0.28	2000.13	46.19	0.01	2000.38	46.42	0.02
		1000	1002.51	42.31	0.35	997.83	41.86	-0.12	1000.50	42.73	0.15	1000.29	42.82	0.13
		500	502.52	37.92	0.50	498.77	37.55	-0.25	501.13	38.99	0.23	499.92	38.98	-0.02
		300	301.58	35.65	0.53	300.07	35.22	0.02	300.93	36.19	0.31	299.96	36.29	-0.01
Average				**42.25**	**0.31**		**41.94**	**0.33**		**42.63**	**0.13**		**42.87**	**0.09**
QCIF	Foreman	800	799.86	44.35	-0.02	801.71	44.00	0.21	799.50	44.33	-0.06	804.64	44.42	0.58
		500	500.09	41.60	0.02	499.92	41.25	-0.02	499.57	41.65	-0.09	503.76	41.71	0.75
		300	300.02	39.10	0.01	299.90	38.66	-0.03	299.42	39.05	-0.19	301.73	39.06	0.57
		200	200.08	37.17	0.04	200.00	36.75	0.00	199.37	37.15	-0.31	200.34	37.10	0.17
	Football	800	799.97	38.38	0.00	801.43	38.24	0.18	799.63	38.21	-0.05	799.83	38.67	-0.02
		500	499.89	34.86	-0.02	501.27	34.45	0.25	499.49	35.35	-0.10	499.85	35.30	-0.03
		300	300.04	31.97	0.01	300.86	31.97	0.29	299.42	31.88	-0.19	299.88	32.32	-0.04
		200	200.18	30.15	0.09	200.61	30.02	0.31	199.55	29.94	-0.22	199.92	30.36	-0.04
	Grandma	800	800.54	49.28	0.07	798.74	48.91	-0.16	800.49	49.07	0.06	800.14	49.32	0.02
		500	500.32	47.68	0.06	497.77	47.26	-0.45	500.64	47.49	0.13	500.94	47.75	0.19
		300	300.45	45.40	0.15	298.50	44.97	-0.50	299.93	45.38	-0.02	300.65	45.59	0.22
		200	200.46	43.30	0.23	199.19	42.88	-0.41	200.30	43.54	0.15	200.60	43.61	0.30
	News	800	800.48	51.87	0.06	800.29	50.81	0.04	799.69	51.20	-0.04	799.23	52.15	-0.10
		500	500.54	48.90	0.11	498.69	48.47	-0.26	499.80	48.47	-0.04	499.96	48.89	-0.01
		300	300.28	45.08	0.09	300.37	44.87	0.12	299.92	45.06	-0.03	300.41	45.21	0.14
		200	200.68	42.22	0.34	200.42	41.93	0.21	200.03	42.38	0.02	200.07	42.25	0.03
Average				**41.96**	**0.08**		**41.59**	**0.21**		**41.88**	**0.11**		**42.11**	**0.17**
720P	Night	10000	10001.67	39.10	0.02	10009.68	38.80	0.10	9999.40	39.18	-0.01	9999.87	39.22	0.00
		8000	8000.36	38.36	0.00	8000.17	38.05	0.00	7995.29	38.42	-0.06	7999.89	38.47	0.00
		5000	5000.95	36.83	0.02	4992.78	36.47	-0.14	4998.84	36.83	-0.02	4998.10	36.94	-0.04
		2000	2003.66	33.60	0.18	1994.03	33.15	-0.30	1999.36	33.37	-0.03	2000.95	33.61	0.05
	Crew	10000	10007.22	41.47	0.07	10013.80	41.34	0.14	9994.74	41.51	-0.05	9995.13	41.65	-0.05
		8000	8006.89	40.88	0.09	8011.27	40.74	0.14	8013.48	40.91	0.17	7998.49	41.06	-0.02
		5000	5006.23	39.70	0.12	5007.53	39.52	0.15	5007.03	39.67	0.14	4994.18	39.87	-0.12
		2000	2002.82	37.04	0.14	2003.08	36.78	0.15	2002.71	36.92	0.14	1996.85	37.17	-0.16
	Harbour	10000	10003.49	37.37	0.03	10007.44	37.05	0.07	9991.02	37.37	-0.09	9997.35	37.41	-0.03
		8000	8001.01	36.49	0.01	7997.51	36.13	-0.03	7995.46	36.45	-0.06	7997.60	36.52	-0.03
		5000	5000.80	34.65	0.02	5000.40	34.28	0.01	4998.73	34.50	-0.03	5001.53	34.68	0.03
		2000	2000.74	31.24	0.04	1996.48	30.90	-0.18	1999.84	31.15	-0.01	2000.59	31.25	0.03
Average				**37.23**	**0.06**		**36.93**	**0.12**		**37.19**	**0.07**		**37.32**	**0.05**

experimental results, where the proposed algorithm achieves an average of 0.62 dB, 0.15 dB, and 0.1 dB PSNR gain on QCIF, CIF and HD sequences, respectively, over the traditional one. The largest PSNR gain can be up to 1.52 dB on "Akiyo" (CIF). The R-D curves are drawn in Fig. 4, where the traditional algorithm on MB level is better than the one on frame-level in terms of PSNR gain. Thus, we just consider the traditional algorithm on frame-level in the following comparisons. From Fig. 4, the coding efficiency of ρ-domain algorithm is inferior to other three algorithms. Kamaci's algorithm can achieve the R-D performance comparable to that of the proposed algorithm on "Foreman (QCIF)," "Night," and "Harbour," but there is obvious lost on the other sequences. In Table III, the row labeled "Average" represents absolute error of bit control and PSNR improvement averagely. All testing algorithms are with bit control error less than 0.33%. The proposed algorithm is

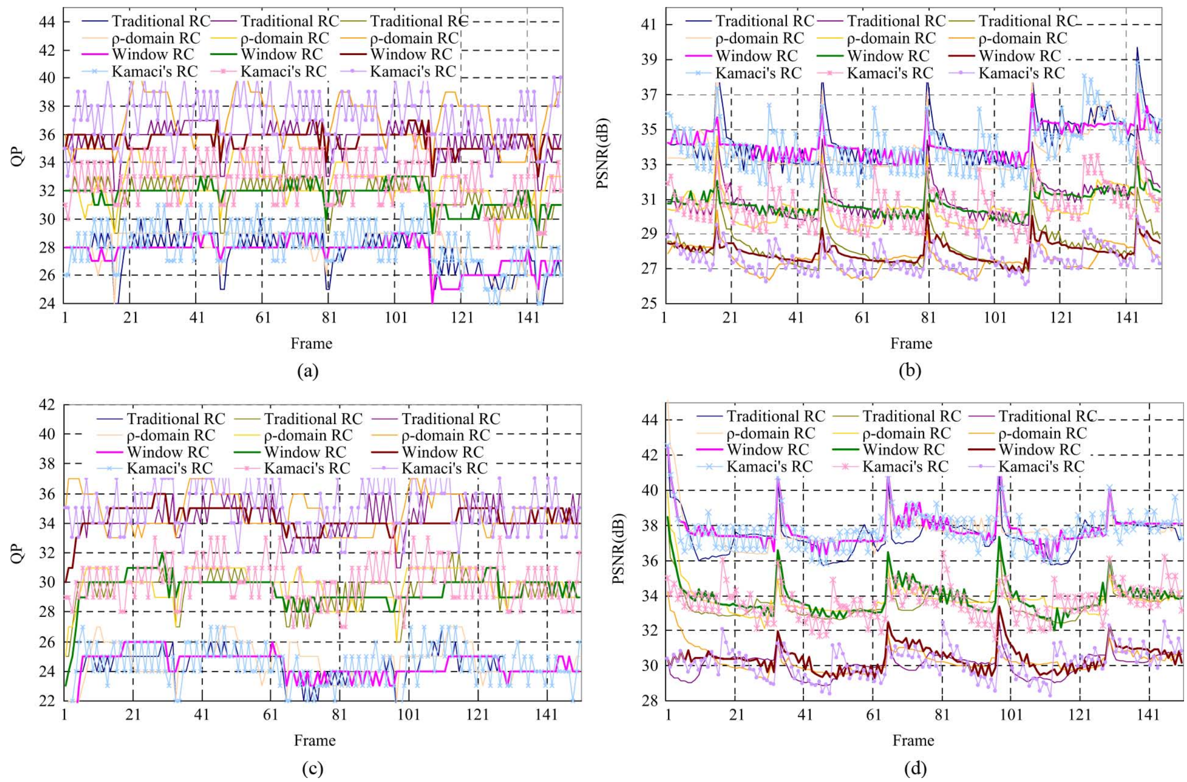


Fig. 5. Comparison of frame QP/PSNR. (a) and (b) "Mobile." (c) and (d) "Bus;" bit rate: 2000 kb/s, 1000 kb/s, and 500 kb/s).

better than others on HD and CIF sequences with only 0.05% bit control error, but little worse than Kamaci's algorithm on QCIF sequence.

Fig. 5 shows frame QP/PSNR curves on two CIF sequences for all testing algorithms, where it can be observed that the proposed algorithm achieves much smoother QP/PSNR performances than benchmarks. The ρ-domain algorithm is with the significant QP variation. Both traditional and Kamaci's algorithms are with the persistent QP fluctuation along the QP curves. The QP/PSNR fluctuations represented by their STDs $(\sigma_Q/\sigma_{\mathrm{psnr}})$ are listed in Table IV. In the medium and high bit rate, the significantly smooth PSNR curves can be observed for the proposed algorithm from Fig. 5(b) and (d). And also, the least PSNR STD for the proposed algorithm can be observed from Table IV.

TABLE IV
VARIATION OF FRAME QP/PSNR FOR ALL TEST ALGORITHMS

Sequence	Target bit rate (kbps)	Traditional RC		ρ-domain RC		Kamaci's RC		Window RC	
		σ_Q	σ_{psnr}	σ_Q	σ_{psnr}	σ_Q	σ_{psnr}	σ_Q	σ_{psnr}
Mobile	2000	1.73	1.61	1.72	1.57	1.72	1.54	1.18	1.03
	1000	1.43	1.32	1.66	1.24	1.72	1.21	0.93	0.83
	500	1.43	0.95	1.87	0.74	1.84	0.8	0.99	0.71
	300	2.32	0.99	2.31	0.78	2.08	0.69	1.98	0.7
Bus	2000	1.13	0.97	1.29	1.06	1.44	1.1	1.25	1.05
	1000	1.21	0.98	1.13	0.89	1.5	0.95	1.12	0.91
	500	1.14	0.74	1.27	0.64	1.63	0.87	0.86	0.73
	300	1.34	0.9	1.66	0.77	1.7	0.85	0.59	0.53
Average		**1.47**	**1.06**	**1.61**	**0.96**	**1.7**	**1**	**1.11**	**0.81**

The proposed algorithm performs much better than the traditional one in the circumstance of frequent scene change. The comparative experiments of them are performed on a concatenated sequence combined by the first 32 frames of seven CIF sequences ("Foreman," "Football," "Mobile," "Silent," "Stefan," "Bus," and "News"). The experimental results are shown in Fig. 6, where the frame QP/PSNR variation of the proposed algorithm with window size of 32 is much less than that of traditional one. As window size increases further, the smoother QP/PSNR can be observed. For traditional algorithm, the QP variation may overreact if scene changes significantly. But the proposed algorithm shows a robust performance in controlling picture quality smoothness for multiple scenes. Noted that the smooth QP/PSNR is for each window, but there are obvious QP variations between far from windows. The reason is that the bits quota is allocated averagely to each window at a window-level CBR mode, but the video content of windows is much different from each other. For VBR encoding or enough buffer capacity/delay, the consistent visual quality on sequence-level can be achieved by choosing window size to the length of sequence. From Fig. 6(c), the R-D performance of the proposed algorithm is up to 1.0 dB higher than that of traditional one when the window size of 32 is selected, but it is less than that of traditional one if the window size is beyond 64. The reason is that the PSNR improvement of simple scene is more than that of complex scene with the same bits increased. Therefore, the more PSNR gain can be obtained if the more bits are assigned to simple scenes. At least the same bits should be allocated to each scene for R-D performance. But in the proposed algorithm, all frames in a window share the bits budget without limitation on an individual frame; meanwhile, as smooth as possible QPs are for all frames in a window. So,

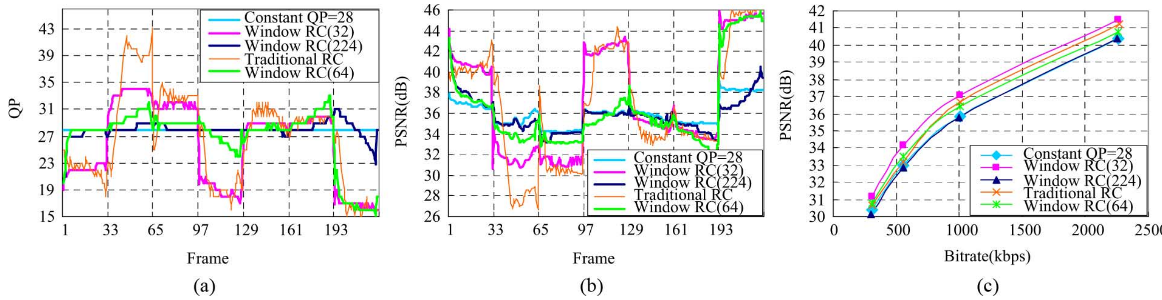


Fig. 6. Comparison of the proposed and traditional algorithms for multiple scenes. (a) Frame QP variation. (b) Frame PSNR variation. (c) R-D performance.

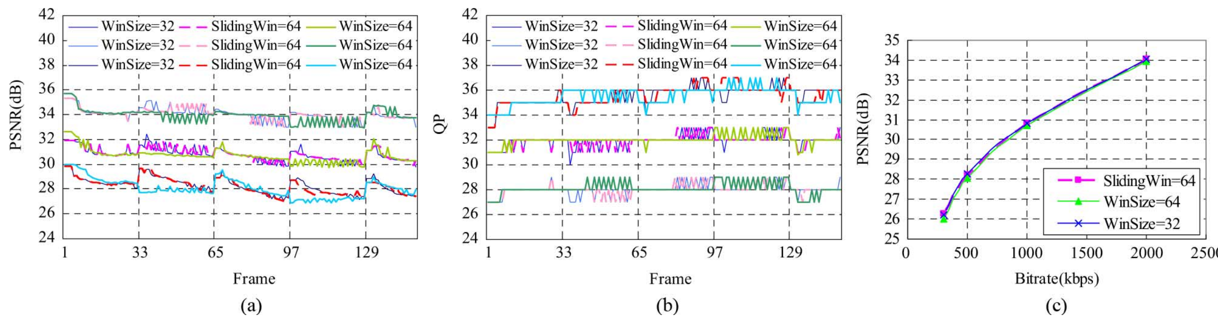


Fig. 7. Comparison of the proposed window and sliding-window on "Mobile" (CIF). (a) and (b) Coded at 2000, 1000, and 500 kb/s. (a) Frame PSNR variation. (b) Frame QP variation. (c) R-D performance.

actually, more bits are consumed by complex frames. Thus, the R-D performance is compromised for large window. But the much smooth visual quality is obtained for them.

The complexity should be considered for applying a rate control algorithm to the real-time video coding systems. In the proposed algorithm, the preanalysis for establishing ρ-QP tables is required. But the additional complexity is slightly increased from both complexity analysis and experiments as follows. Only 16×16 block ME and 4×4 integer DCT transform are performed for P/B frame in preanalysis. The fast ME at integer-pixel with only one reference (B uses one forward reference) is employed in this stage, and SAD is used to select the best MV. For I frame, only 4×4 Intra predictions with few directions (such as horizontal, vertical and diagonal) are compared to select a best one with smallest SAD. The complexity of fast ME is about 10% of the total complexity for P frame encoding with one reference frame [46], and DCT complexity is negligible $(< 1\%)$ [46]. Thus, the complexity of preanalysis is very little in the proposed algorithm. From the time statistic of software implementation under the conditions aforementioned, the time of preanalysis is about 4%–7% of the total encoding time as fast ME is enabled, and less than 4% as full ME is enabled. In addition, the time complexity of the proposed algorithm is only 1.8%–2.1% higher than that of traditional one. Kamaci's algorithm is comparable to the traditional one in the aspect of complexity. The complexity of ρ-domain is over 20% higher than other three methods due to complex control of MB-level and preanalysis. However, the startup delay is required to preanalyzing a window for the proposed algorithm. We estimate such delay by the proportion of time consuming for preanalysis to total encoding time. Due to about 4%–7% additional time consuming for preanalysis, the startup delay of preanalyzing a window of 32 frames is just corresponding to the encoding time of 1.28–2.24, i.e., $32 \times (4\% - 7\%)$ frames. It is worthwhile to achieve smooth visual quality with the cost of such a delay. In addition, such a delay can be further reduced or eliminated by introduce multiple threads processing because the process of preanalysis is independent to final encoding. Even MV information of preanalysis can be reused in the final encoding process.

C. Reducing Startup Delay of the Proposed Algorithm

Given L and ΔR, the quality smoothness can be expected by the extension of window model (9). From (9), the larger L is, the smoother the picture quality is. However, the larger L means the larger startup delay and buffer delay. In addition, the video content may vary notably between windows, so the significant QP/PSNR fluctuation between windows inherently exists for CBR encoding.

For a smooth transition of picture quality between windows and reducing startup delay and buffer delay, we further introduce sliding-window which is composed of parts of frames from the last coded window and parts of frames for forthcoming coding. Firstly, the startup delay of sliding window is less than that of window which only consists of the forthcoming frames, because only the forthcoming frames need to be preanalyzed in sliding window. Second, the size of sliding window is larger than that of window with only forthcoming frames, so smoother QP can be achieved according to window model (9). Third, there may be significant change of video content between two neighboring windows, which will results in large QP difference across windows. It can be eliminated by sliding-window as the frames involved are processed in a same window.

The comparative experiments are performed with the proposed window and sliding-window algorithms. The size of sliding-window equals 64, including 32 frames previously coded and 32 forthcoming frames The experimental results shown in Fig. 7 exhibit sliding-window method is better than window method with window size of 32, but little worse than window method with window size of 64 from the aspect of quality smoothness. Meanwhile, they have the comparable R-D performance from Fig. 7(c).

V. Conclusion

In this paper, we firstly propose a theoretical model to show the relations among window size, QP variation and bits variation. Then, a window-level ρ-domain rate control algorithm incorporating the window model into the traditional ρ-domain rate control is introduced. In our proposed rate control algorithm, the picture quality and buffer, which are the two key criteria in rate control, are controllable in theory and practical applications by selecting the proper window size. Furthermore, if the window size and buffer constraint are given, as smooth as possible picture quality can be obtained. As the proposed rate control algorithm operates on window level, in which the bits quota of a window is allocated according to the average bit rate, the traditional bit allocation on frame level is not necessary anymore. Besides, the window-level rate control using sliding window is also introduced to reduce the time-delay of encoder, and smooth the picture quality variation between windows. The future work will be done by introducing a new criterion for the smoothness of picture quality and buffer occupancy. On the other hand, the picture quality under the assumption of multiple scenes needs to be studied further.

References

[1] *ISO/IEC Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbits/s Part2: Video,(MPEG1)*, JTC1/SC29/WG11, Mar. 18, 2004.

[2] *ISO/IEC Video Recommendation ITU-T H.262, Generic Coding of Moving Pictures and Associated Audio Information*, MPEG2/H.262, 1995.

[3] *ITU-T /Study Group XV(1988-1993), Video Codec for Audiovisual Services at px64 kbits*, H.261, Mar. 1–12, 1993.

[4] *ITU-T/Study Group 15, Video coding for Low Bit Rate Communication*, H.263, May 2, 1996.

[5] *ISO/IEC JVT, ITU-T VCEG, ISO/IEC 14496-10 or ITU-T Rec. H.264, Doc. JVT-B118r8*, MPEG4/H.264, Feb. 1, 2002.

[6] *Information Technology Advanced Coding of Audio and Video Part2: Video. Standard Draft, (AVS)*, AVS-P2, Mar. 2005.

[7] *ISO/IEC, Coding of Moving Pictures and Associated Audio, Test model 5,(MPEG2) MPEG*, JTC1/SC29/WG11, 1994.

[8] *ITU-T/ Study Group 15, Video Codec Test Model TMN8*, H.263, Jun. 1997.

[9] in *Proc. 10th Meeting H.263 Version 3, Test Model 12*, Portland, OR, Aug. 22–24, 2000.

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

[11] *Text of ISO/IEC 14496-2 MPEG4 video VM-version 8.0 ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Associated Audio MPEG 97/W1796*, , Jul. 1997.

[12] 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. 287–311, Feb. 1997.

[13] H. Gish and J. N. Pierce, "Asymptotically efficiency quantizing," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 5, pp. 676–683, Sep. 1968.

[14] E. Lam and J. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[15] L. Xu, X. Y. Ji, W. Gao, and D. B. Zhao, "Laplacian distortion model (LDM) for rate control in video coding," in *Proc. PCM*, Dec. 2007, pp. 638–646.

[16] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[17] H. Wang, S. Kwong, and C. W. Kok, "Efficient prediction algorithm of integer DCT coefficients for H.264/AVC optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 2, pp. 547–552, Apr. 2006.

[18] Z. H. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via ρ-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, pp. 928–940, Aug. 2001.

[19] Z. G. Li, W. Gao, F. Pan, S. W. Ma, K. P. Lim, G. N. Feng, X. Lin, S. Rahardja, H. Q. Lu, and Y. Lu, "Adaptive rate control for H.264," *J. Vis. Commun. Image Represent.*, vol. 17, no. 2, pp. 376–406, Apr. 2006.

[20] S. W. Ma, Z. G. Li, and F. Wu, "Proposed draft adaptive rate control," in *Proc. 8th Meeting Joint Video Team ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-H017r3*, Geneva, May 20–26, 2003.

[21] Y. Liu and Z. G. Li, "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.

[22] S. W. Ma, W. Gao, and Y. Lu, "Proposed draft description of rate control on JVT standard," in *Proc. 6th Meeting JVT, Doc. JVT-F086-L, 6th Meeting*, Awaji, JP, Island, Dec. 2002.

[23] S. W. 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.

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

[25] T. Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," in *Proc. Int. Conf. Image Process.*, Thessaloniki, Greece, Oct. 2001, pp. 542–545.

[26] J. P. Dong and N. Ling, "A model parameter and MAD prediction scheme for H.264 macroblock layer rate control," in *Proc. ISCAS*, May 18–21, 2008, pp. 628–631.

[27] S. J. Wu, Y. Q. Huang, and T. Ikenaga, "A macroblock-level rate control algorithm for H.264/AVC video coding with context-adaptive MAD prediction model," in *Proc. ICCMS*, Feb. 20–22, 2009, pp. 124–128.

[28] Y. Liu, Z. G. Li, and Y. C. Soh, "Rate control of H.264/AVC scalable extension," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 116–121, Jan. 2008.

[29] Z. H. He, W. Zeng, and C. W. Chen, "Low-pass filtering of rate-distortion functions for quality smoothing in real-time video communication," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 973–981, Aug. 2005.

[30] 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.

[31] Z. Z. Chen and K. N. Ngan, "Towards rate-distortion tradeoff in real-time color video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 1, pp. 158–167, Jan. 2007.

[32] L. Xu, Z. H. Wang, L. Deng, X. Y. Ji, D. B. Zhao, and W. Gao, "Window-level rate control for smooth picture quality and smooth buffer occupancy," in *Proc. IEEE PCS*, Chicago, IL, May 2009, pp. 125–128.

[33] P. H. Westerink, R. Rajagopalan, and C. A. Gonzales, "Two-pass MPEG-2 variable-bit-rate encoding," *IBM J. Res. Develop.*, vol. 43, no. 4, pp. 471–488, Jul. 1999.

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

[35] 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. 1–11, Mar. 2008.

[36] Y. Yu, J. Zhou, Y. Wang, and C. W. 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.

[37] D. Bagni, B. Biffi, and R. Ramalho, "A constant-quality, single-pass VBR control for DVD recoders," *IEEE Trans. Consum. Electron.*, vol. 49, no. 3, pp. 653–662, Aug. 2003.

[38] J. F. Cai, Z. H. He, and C. W. Chen, "A novel frame-level bit allocation based on two-pass video encoding for low bit rate video streaming applications," *J. Vis. Commun. Image Represent.*, vol. 17, no. 4, pp. 783–798, 2006.

[39] J. R. Corbera, P. A. Chou, and S. L. Regunathan, "A generalized hypothetical reference decoder for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 674–687, Jul. 2003.

[40] "Time-shift causality constraint on the CAT-LB HRD," in *Proc. 5th Meeting Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-E133.doc, 5th Meeting*, Geneva, Switzerland, Oct. 9–17, 2002.

[41] F. Cucker and D. X. Zhou*, Learning Theory: An Approximation Theory Viewpoint*. Cambridge, U.K.: Cambridge Univ. Press.

[42] Z. S. Wei*, Probability and Statistics*. Beijing, China: Higher Education Press.

[43] 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.

[44] N. Kamaci and Y. Altunbasak, "Frame bit allocation for H.264 using cauchy-distribution based source modeling," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Philadelphia, PA, Mar. 2005, pp. 57–60.

[45] Y. Altunbasak and N. Kamaci, "An analysis of the DCT coefficient distribution with the H.264 video coder," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Montreal, QC, Canada, May 2004, pp. 177–180.

[46] L. Su, Y. Lu, F. Wu, S. P. Li, and W. Gao, "Complexity-constrained H.264 video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 1–15, Apr. 2009.

Long Xu received the M.S. degree in applied mathematics from Xidian University, Shaanxi, China, in 2002, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China.

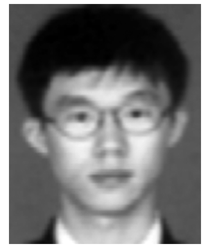
He is currently a Postdoctoral Researcher for the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. His research interests include image/video coding, wavelet-based image/video coding, and computer vision.

Debin Zhao received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1985, 1988, and 1998, respectively.

He was a Research Fellow in the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, from 1989 to 1993. He is currently a Professor in the Department of Computer Science, Harbin Institute of Technology. His research interests include data compression, image processing, and human-machine interface. He has published two books and over 60 scientific papers.

Xiangyang Ji received the M.S. degree from the Harbin Institute of Technology, Harbin, China, 2001, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Science, Beijing, China.

Currently, he is an Assistant Professor at Tsinghua University, Beijing, China. His research interests include video coding and video transmission.



Lei Deng received the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 2002 and 2007, respectively.

He was a Postdoctoral Researcher at the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, from 2007 to 2009. Currently, he is with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. His research interests include computer architecture, digital signal processing, and video compression.

Sam Kwong (M'93–SM'04) received the B.S. and M.S. degrees in electrical engineering from the State University of New York at Buffalo in 1983, the University of Waterloo, Waterloo, ON, Canada, in 1985, and the Ph.D. degree from the University of Hagen, Hagen, Germany, in 1996.

From 1985 to 1987, he was a Diagnostic Engineer with Control Data Canada. He joined Bell Northern Research Canada as a Member of Scientific Staff. In 1990, he became a Lecturer in the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong, where he is currently a Professor in the Department of Computer Science. His research interests are video and image coding and evolutionary algorithms.

Wen Gao (M'92–SM'05–F'08) received the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He is a Professor of computer science with Peking University, Beijing, China. Before joining Peking University, he was a Professor of computer science with Harbin Institute of Technology from 1991 to 1995, and a Professor with the Institute of Computing Technology of the Chinese Academy of Sciences. He has published extensively, including four books and over 600 technical articles in refereed journals and proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics. He serves or has served on the editorial board of several journals, including *EURASIP Journal of Image Communications* and the *Journal of Visual Communication and Image Representation.*

Dr. Gao serves or has served on the editorial board of several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, THE IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as IEEE ICME and ACM Multimedia, and has also served on the advisory and technical committees of numerous professional organizations.