Region-Based Rate Control for H.264/AVC for Low Bit-Rate Applications

Hai-Miao Hu, Bo Li, Weiyao Lin, Wei Li, and Ming-Ting Sun, *Fellow, IEEE*

Abstract—Rate control plays an important role in video coding. However, in the conventional rate control algorithms, the number and position of macroblocks (MBs) inside one basic unit for rate control is inflexible and predetermined. The different characteristics of the MBs are not fully considered. Also, there is no overall optimization of the coding of basic units. This paper proposes a new region-based rate control scheme for H.264/advanced video coding to improve the coding efficiency. The inter-frame information is explored to objectively divide one frame into multiple regions based on their rate-distortion (R-D) behaviors. The MBs with similar characteristics are classified into the same region, and the entire region, instead of a single MB or a group of contiguous MBs, is treated as a basic unit for rate control. A linear rate-quantization stepsize model and a linear distortion-quantization stepsize model are proposed to accurately describe the R-D characteristics for the region-based basic units. Moreover, based on the above linear models, an overall optimization model is proposed to obtain suitable quantization parameters for the region-based basic units. Experimental results demonstrate that the proposed region-based rate control approach can achieve both better subjective and objective quality by performing the rate control adaptively with the content, compared to the conventional rate control approaches.

Index Terms—Basic unit rate control, bit-allocation, H.264/advanced video coding (AVC), low bit-rate applications, optimal quantization parameter (QP) achievement, rate-distortion (R-D) optimization, region division.

I. INTRODUCTION

WITH the increasing demand of various video transmission applications, video coding with a high coding efficiency is important. Comparing with exiting video coding standards, H.264 advanced video coding (AVC) [1] can achieve a

H.-M. Hu and W. Li are with the Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China (e-mail: huhaimiao@cse.buaa.edu.cn; lwei@buaa.edu.cn).

B. Li is with the Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China, and also with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China (e-mail: boli@buaa.edu.cn).

W. Lin is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200030, China (e-mail: wylin@sjtu.edu.cn).

M.-T. Sun is with the Department of Electrical Engineering, University of Washington, Seattle, WA 98105 USA (e-mail: mts@u.washington.edu).

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/TCSVT.2012.2199398

better coding efficiency under the transmission capacity constraint, due to the adoption of several novel coding techniques. In video coding, rate control plays an important role. The major purpose of rate control is to determine the quantization stepsizes (or quantization parameters) to regulate the bit-rate produced by the video encoder according to the available channel bandwidth to maximize the video quality. There are many previous works on the rate control of H.264/AVC [2]–[23]. Most of them were based on JVT-G012, which was adopted as the rate control scheme in the H.264/AVC recommended software Joint Model (JM) [9], [24].

JVT-G012 introduced a concept of basic unit rate control. The basic unit is defined to be a group of contiguous macroblocks (MBs). The number of MBs in a basic unit is fixed, and it can be an MB, an MB row, a slice, or a frame [9]. The basic unit rate control is a top-down ordinal process and mainly consists of the following steps. First, the remaining bits for the current frame are allocated to not-yet coded basic units equally. The mean absolute difference (MAD) of the current basic unit is then predicted through the MAD prediction model using the actual MAD of the co-located basic unit in the previous frame. Finally, a quadratic rate-distortion (R-D) model is used to compute the quantization parameter (QP) for the basic unit according to the predicted MAD.

The basic unit rate control makes the H.264/AVC encoder more flexible and can achieve good performance in video transmission applications, in which the subjective quality of encoded frames is often seriously degraded due to the restricted channel bandwidth. In applications such as videophone and remote video surveillance, the perceptual quality of the foreground moving object is more important than the background. However, given the same QP, the perceptual quality of these regions may be quite different due to the different characteristics of the regions. In order to reduce the transmission bits while keeping good perceptual quality, the features of the human visual system (HVS) are often exploited to allocate more bits to the basic units located in the heuristically determined regions of interest (ROI), compared to the other basic units located in the non-ROI [2]–[5], [25]–[29].

Although the basic unit rate control in JVT-G012 gives good performance, its coding efficiency decreases with the decrease of basic unit size. Generally, the MB-layer rate control results in the worst coding efficiency [6], [7], for several reasons. First, since the number of MBs in a basic unit inside one frame is inflexible and their locations are predetermined before encoding [7], [9], the correlation

between the current basic unit and its co-located basic unit of the previous frame may not be so strong due to local object motions or the global camera motion. Second, the basic unit rate control scheme is a top-down ordinal process without the overall optimization among the basic units in one frame. This may lead to the overdraft of bit-budget for the beginning basic units in a frame, and result in bit starvation for the later basic units. Third, since the MBs inside one basic unit are predetermined regardless of their different characteristics, adopting one uniform model for an entire basic unit cannot produce the optimal result. Although the frame-layer rate control can achieve better coding efficiency compared to the small size basic unit rate control, its coding efficiency can be further improved by considering different characteristics of the MBs. The above conclusions will be further confirmed by the experiments in the following sections.

In order to improve the coding efficiency, this paper proposes a new region-based rate control scheme for H.264/AVC, in which the MBs with similar R-D characteristics are classified into the same basic unit, and an overall optimization is applied to the basic units by using the proposed linear rate-quantization stepsize (R-QS) model and the linear distortion-quantization stepsize (D-QS) model. Experimental results demonstrate that the proposed region-based rate control scheme can achieve both better subjective and objective quality by performing the rate control adaptively with the content, compared to the conventional basic unit layer rate control.

Note that in our proposed approach, the numbers of MBs in different basic units are different and the location of MBs belonging to a basic unit may be discontinuous, which is different from conventional basic unit rate control approaches. Although basic unit rate control is an active research topic [11], [22], [23], to the best of our knowledge, no region-based basic unit rate control, such as our proposed approach, has ever been proposed.

The remainder of this paper is organized as follows. Some observations and justifications to support our proposed region-based rate control are provided in Section II. Section III discusses the proposed region-based basic unit rate control scheme. The proposed region division algorithm is presented in Section III-A. The proposed linear R-QS and D-QS models are described in Section III-B and the QP determination method is described in Section III-C. The proposed approach is summarized in Section III-D. Section IV presents the experimental results. Some discussions are drawn in Section V. Finally, this paper is concluded in Section VI.

II. OBSERVATIONS AND JUSTIFICATIONS

The proposed region-based rate control is related to three important issues, namely, region division, bit-allocation among coding units, and the QP determination. Some observations related to these issues and justifications to our proposed approach are discussed in detail as follows.

A. Region Division

In a video frame, a moving region (MR) usually attracts more attention. In many previous works [2]–[5], [25]–[28],
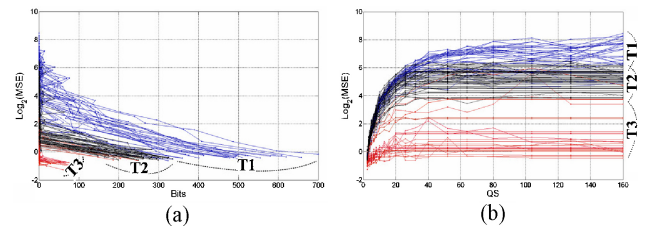

Fig. 1. R-D curves and D-Q curves of all the MBs in one frame (*Grandma*, 90th, QCIF, QP: 12:2:48). (a) R-D curves. (b) D-Q curves.
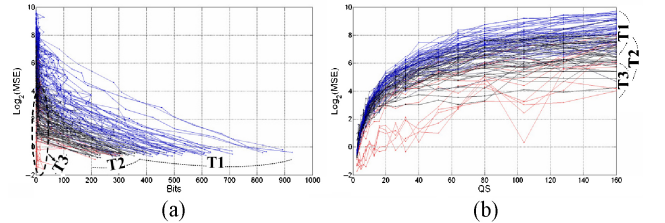

Fig. 2. R-D curves and D-Q curves of all the MBs in one frame (*Foreman*, 114th, QCIF, QP: 12:2:48). (a) R-D curves. (b) D-Q curves.

an MR is allocated more bits to keep a good perceptual quality. However, it is difficult to extract reliable MRs. Several techniques [4], [26], [27] have been adopted to extract the MRs. These methods are performed after motion estimation (ME), and the QPs are adjusted based on the MR extraction result. Since the H.264/AVC reference software uses R-D optimization in the ME, which requires the QP information, these MR extraction methods cannot be directly used due to the "region extraction-ME" dilemma. In order to solve this dilemma, difference-based methods [2], [3], [5], [25] are proposed, which utilize the difference between the current frame and the previous frame. However, when there is a significant temporal variation due to camera motions, these methods cannot produce appropriate results [5], [28].

We observe that besides attracting more attention, an MR is more difficult to code and its R-D characteristic is significantly different from non-moving (non-MR) regions. For an MR, it is not likely that the motion compensation is perfect, and so, there are usually more prediction residuals needing more bits to code given the same QP. The R-D curves of different MBs within a video frame are different, and can be classified into different types with different R-D characteristics. Two examples are shown in Figs. 1 and 2. It can be confirmed that the MBs with the R-D curves, labeled as T1 in Figs. 1 and 2, correspond to MRs. These curves have long fat tails, which implies that MRs will consume more bits than the other regions. Also, it can be observed from Figs. 1(b) and 2(b) that the coding quality of the MRs is worse than the other regions with the same QP, which may result in a nonuniform quality for the entire frame. The effect is especially significant in the low bit-rate applications, where relative large QPs are utilized.

Moreover, we observe that the characteristics of MBs inside the non-MR may not be uniform, as shown in Figs. 1 and 2. The MBs within the non-MR can be classified into two types, that are labeled as T2 and T3 in Figs. 1 and 2, with different R-D characteristics. The R-D curves in T3 have shorter tails and descend more rapidly. This implies that the T3 region is easy to code and needs only a small number of bits. The

TABLE I

COMPARISON OF CODING EFFICIENCY BETWEEN THE ORIGINAL
SEQUENCES AND THE INVERTED VERSION (I.E., THE ORIGINAL
FRAMES ROTATED 180°)

Sequence (Target Bit-Rate)		PSNR (dB)		Bit-Rate (kb/s)	
		MB [9]	Frame [9]	MB [9]	Frame [9]
Claire (16 kb/s)	Original	38.17	39.35	16.05	16.04
	Rotated	38.84	39.24	16.06	16.01
Grandma (16 kb/s)	Original	34.97	35.87	16.05	16.00
	Rotated	35.16	35.91	16.05	15.97
City (24 kb/s)	Original	29.04	29.56	24.08	24.03
	Rotated	29.18	29.55	24.08	24.04
Football (24 kb/s)	Original	25.62	25.76	24.18	24.03
	Rotated	25.33	25.81	24.11	24.06
Avg. absolute difference		*0.32*	*0.05*		

T2 regions are transition regions and have a medium tail. In order to improve the coding efficiency, it is advantageous for the non-MR to be further divided into two sub-regions and to apply different bit-allocation strategies for the different sub-regions according to their characteristics [28], [30].

B. Bit-Allocation Among Coding Units

Accurate bit-allocation is an important but difficult issue in rate control in terms of the overall coding performance. In the conventional basic unit rate control, the remaining bits are equally allocated to the not-yet coded basic units and there is no overall optimization among the basic coding units [9]. However, this top-down ordinal process may lead to overdraft of the bit-budget for the beginning basic units in a frame and bit starvation for the later basic units as mentioned previously, thus resulting in inappropriate coding results. This is more obvious in low bit-rate applications, where the numbers of bits allocated to each frame are relatively limited. This conclusion is confirmed by the experimental results shown in Table I, where four video sequences and their corresponding upside-down inverted versions are coded twice with an MB-layer rate control scheme and a frame-layer rate control scheme, respectively. The experimental setting is the same as that described in Section IV. We can see that compared to the frame-layer rate control, the peak signal-to-noise ratio (PSNR) differences between an original sequence and the corresponding inverted one are much more obvious using the MB-layer rate control scheme than using the frame-layer rate control scheme. This implies that the MB-layer rate control scheme may lead to less stable results even for the same video content. It also shows that the bit-allocation among basic coding units is indeed an issue.

In some rate control schemes [2]–[5], [25]–[28], the MBs located in an MR are allocated more bits by adopting larger weighting factors in the bit-allocation strategy than the MBs located in the non-MR. However, the weighting factors are usually constants and determined in heuristic ways that may not achieve good results.

C. QP Determination

An R-D model is usually employed to achieve the QP assignment based on the given bit-allocation. Based on
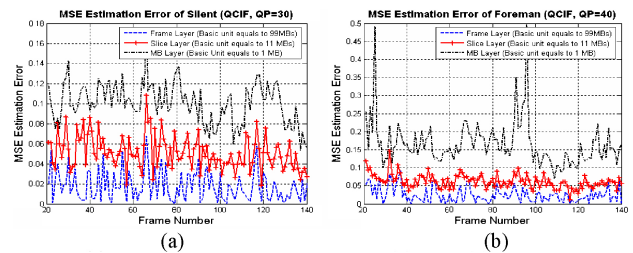

Fig. 3. Inter-frame correlation comparison among the different layer rate control schemes. (a) *Silent* (QP = 30). (b) *Foreman* (QP = 40).

different assumptions of the discrete cosine transform (DCT) coefficient distribution, many R-D models have been proposed [10], [11], [31]–[33] for rate control. A logarithmic rate-quantization (R-Q) model was proposed in [31] by assuming the DCT coefficients have a Gaussian distribution. In [32], an R-D model based on the fraction of zeros among the quantized DCT coefficients was proposed by assuming that the DCT coefficients have a Laplacian distribution. Also, based on the assumption of Laplacian distribution, a quadratic R-D model was proposed in [33] and was adopted as the rate control scheme in JVT-G012 for H.264/AVC. Recently, the Cauchy distribution was reported to be a better fit than the Laplacian distribution for H.264/AVC. Several R-D models have been developed for H.264 by using the Cauchy distribution of DCT coefficients [10]–[12]. However, it is relatively difficult to apply since the mean and variance of Cauchy distribution are not well defined [34]. Moreover, the entropy of the quantized transformed residuals is regarded as the actual rate in the above R-D models. However, the entropy is for the case of independent coding, while the quantized transformed residuals are always dependently entropy coded at the block level, such as the case of run-length coding [34].

The inter-frame correlation is another important issue that affects the accuracy of the R-D model. Since the R-D characteristic of the not-yet coded basic units is not known, the R-D model applied to the current basic unit is established based on the coding results of the co-located basic unit in the previous coded frames. As mentioned, rate control with a smaller size basic unit cannot achieve as good performance as with a large size basic unit (e.g., the frame-layer rate control) [6], [7]. This is because the correlation between the current basic unit and its co-located basic unit of the previous frame may not be so strong due to local object motions or global camera motions. This statement can be confirmed by the experimental results shown in Fig. 3. In Fig. 3, the mean square error (MSE) mismatch between different sizes of basic units in the current frame and the co-located basic unit in the previous frame is used as an example to implicitly compare the inter-frame correlation among different layer rate control schemes. It can be observed that the MB-layer rate control has the largest estimation error.

III. PROPOSED REGION-BASED BASIC UNIT RATE CONTROL FOR H.264/AVC

A. Proposed Region Division Algorithm

The difference-based MR extraction methods are widely utilized and can achieve good MR extraction when the

background holds still [3], [5]. However, when there is a significant temporal variation due to the global motions, these methods cannot produce satisfactory results [5]. Therefore, it is reasonable to perform the global motion compensation before the MR extraction.

The gray projection method (GPM) is a simple and effective global motion vector (GMV) estimation method and has been widely used in the digital image stabilization [35]. Its basic idea is to project the two-dimensional image into two independent one-dimensional curves, namely, the row projection curve and the column projection curve, and to estimate the GMV by searching the maximum cross-correlation of the projection curves between the current frame and its previous frame. The detailed description of GPM can be found in [35]. Since the GPM utilizes the statistical information, it is robust to local object motions and can achieve good results. Moreover, the GPM has a low computational complexity, which can be easily applied to real-time applications. Therefore, GPM is used in the MR extraction. Based on the estimated GMV, the difference between the current frame and its previous frame for an MB is calculated by

$$Diff_k(p) = \frac{1}{256} \times \sum_{(i,j) \in p} \left| G_k(i,j) - G_{k-1}(i + GV_k^x, j + GV_k^y) \right|$$

(1)

where $Diff_k(p)$ denotes the difference of the pth MB in the kth frame. (GV_k^x, GV_k^y) represent the horizontal and vertical components of the GMV, respectively. $G_k(i,j)$ denotes the luminance value of the pixel (i,j) in the kth frame. Note that $Diff_k(p)$ will be further employed as the complexity measure for the proposed linear R-QS model, which will be described in the following section.

The MR is extracted with a predefined threshold based on the calculated difference. Furthermore, according to the HVS feature, the central region of the picture attracts more attention than other parts [2]. Thus, the MBs that have larger temporal differences and are at the central location of the frame are treated as the MR. The proposed MR extraction method is described by (2), where $MR_k(p)$ equals 1 if the pth MB belongs to the MR, and equals 0 otherwise

$$MR_k(p) = \begin{cases} 1, & \text{if } \omega \times Diff_k(p)/Diff_k^{\mathrm{avg}} > Th1 \\ 0, & \text{else} \end{cases}$$

(2)

where $Diff_k^{\mathrm{avg}}$ represents the average difference of MBs, $Th1$ is a predefined threshold ($Th1$ is set to 0.75 in this paper), and ω is a constant decided by the location of the current MB. In our experiments, ω is divided into three levels based on experiments: 1.0 for the central region, 0.1 for the border region, and 0.55 for the transition region. For the quarter common intermediate format (QCIF) resolution, one MB-wide strip along the edge of the frame is defined as the border region. Two MB-wide strip next to the border region is defined as the transition region. The remaining part is defined as the central region. For the common intermediate format (CIF) resolution, the strip widths of the border region and transition region are doubled.

The non-MR is subdivided into the complex region and the flat region based on the observation from Figs. 1 and 2. Due
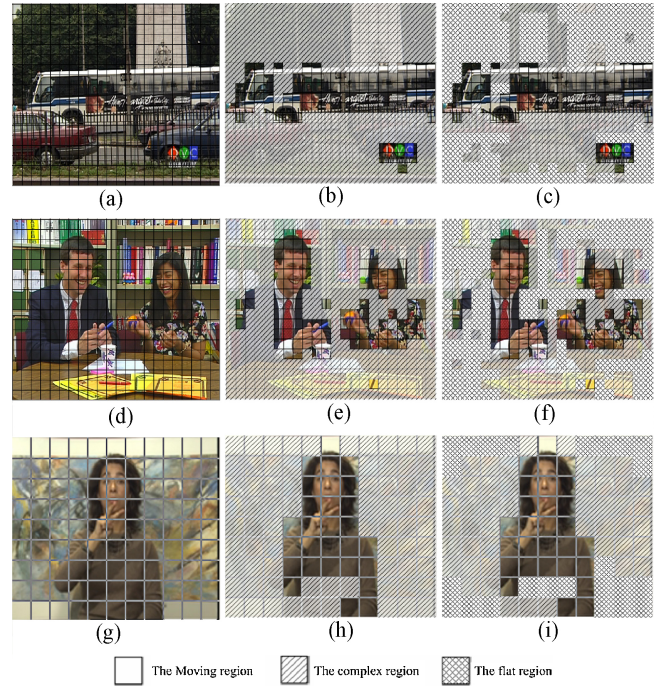

Fig. 4. Results of the proposed region division method. (a) Original picture, *Bus* (70th frame, CIF). (d) Original picture, *Paris* (93rd frame, CIF). (g) Original picture, *Silent* (93rd frame, QCIF). (b), (e), and (h) Our MR division result. (c), (f), and (i) Our non-MR sub-division result.

to the "chicken and egg dilemma" described in [9], the MSE information of the current frame is unavailable before the QP value is determined. Since there is no large motion existing in the non-MR, the variation of the content in the non-MR is relative stable. Thus, the MBs in the non-MR have a strong inter-frame correlation, and the MSE distribution among the MBs in the non-MR is similar to that in the previous frame. Therefore, the MSE information of the co-located MBs in the previous frames is used as the complex measurement for the non-MR sub-division. A predefined threshold $Th2$ is used to extract the flat region from the non-MR by

$$NMR_k(p) = \begin{cases} 1, & \text{if } MSE_{k-1}(p)/MSE_{k-1}^{\mathrm{avg}} > Th2 \\ 0, & \text{else} \end{cases}$$

(3)

where $MSE_k(p)$ is the actual MSE of the pth MB in the kth frame, MSE_k^{avg} represents the average MSE, and $NMR_k(p)$ is equal to 1 if the pth MB belongs to the complex region and is equal to 0 otherwise. $Th2$ is set to 0.5 in this paper.

In Fig. 4, the *Bus* sequence contains global motions. As shown in the figure, the experimental results demonstrate that based on the GPM, the proposed method can extract the important MR well not only when the background holds still, but also when the background is dynamic. Also, the proposed method can efficiently extract the flat region. Although the extracted MR may not always correspond to the objects completely, it will not affect the coding efficiency. Moreover, with the proposed non-MR sub-division, the remaining part of the object (i.e., object parts that are not included in the MR) usually will be classified into the complex region, but not mixed with the flat region.
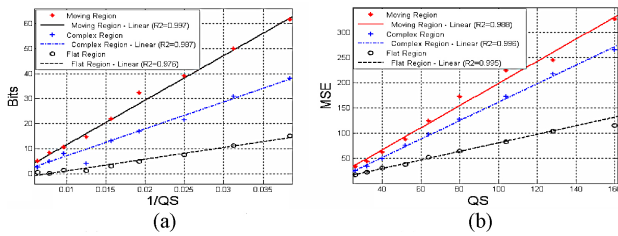
Fig. 5. R-D curves of different regions in the 145th frame of *Foreman* (QP: 32:2:48). (a) R-QS curves. (b) D-QS curves.

Fig. 6. R-D curves of different regions in the 37th frame of *Silent* (QP: 33:3:51). (a) R-QS curves. (b) D-QS curves.

In the next section, we will show that the proposed region division method can result in good partitions of MBs with accurate linear R-QS and D-QS models. Note that the predefined thresholds, such as *Th*1 and *Th*2, are chosen according to the statistics from the experimental results. Although they can achieve acceptable results for most test sequences, a refining threshold determination method can give better results, as will be discussed in Section V.

B. Proposed Linear R-QS Model and D-QS Model

According to the region-division result, we propose to treat the entire region as the basic unit for the rate control. The proposed scheme includes two important parts, which are described as follows. In our proposed approach, the MBs in each region have a similar R-D behavior, and the MBs in different regions have different R-D behaviors. The R-D models suitable for the region-based basic units should be established correspondingly.

Conventionally, the R-D models are derived based on the statistical properties of video signals and R-D theories [6]. Recently, based on the observation that the distribution of the DCT transform coefficients follows a Cauchy distribution, the R-QS and D-QS relationships have been modeled in the form of an exponential function for H.264/AVC [10], [11]. Although Cauchy distribution shows higher accuracy in the case of heavy tails of transformed residuals, it is hard to be applied since the mean and variance of Cauchy distribution are not well defined [34]. Therefore, it is desirable to propose an appropriate R-QS and D-QS model suitable for the region-based basic units.

Several experiments have been carried out to exploit the R-D relationship. The experimental results are shown in Figs. 5 and 6, where "Bits" and "MSE" represent the average bits and the average MSE of MBs in each region, respectively. The R-QS and D-QS models can be represented accurately by linear functions. The accuracy of the curve fitting is specified in Figs. 5 and 6 by the R-squared value (i.e., R2 in the figures), which is an indicator from 0 to 1 that reveals how close the approximated linear function is to the actual data [21]. The R-square values of both the approximated linear R-QS and D-QS models are close to 1, which demonstrates their validity and reliability.

In order to make the R-QS model more adaptive to video contents, a parameter related to video complexity, such as the MAD, can be associated with the R-QS model [33]. We investigate the relationship between the rate and the
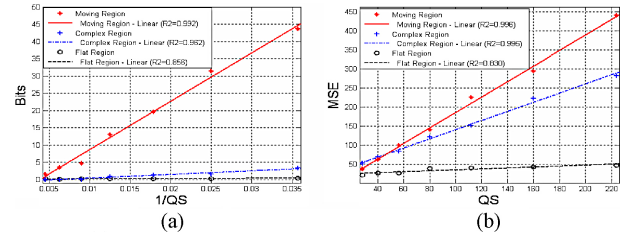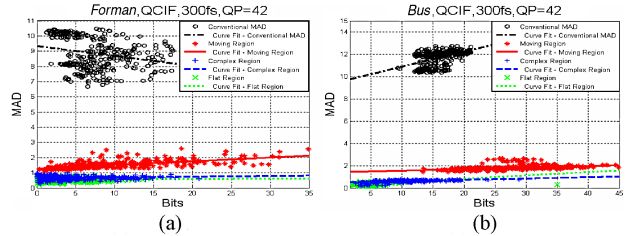

Fig. 7. Relationship between bit-rate and MAD (QCIF, 300 frames, QP = 42). "Bits" represents the average bits of MBs, and "conventional" represents the predicted MAD of one frame with the method in [9]. (a) *Foreman*. (b) *Bus*.

MAD $MAD_{k,r}$ of the rth basic unit in the kth frame, which equals the average $Diff_k(p)$ of MBs located in the rth basic unit [calculated by (1)]. It should be noted that $MAD_{k,r}$ is different from the conventional MAD used in the quadratic R-Q model in JVT-G012, where MAD is predicted based on the actual MAD of the previous frame [9]. Comparing with the conventional MAD, the proposed $MAD_{k,r}$ is also close to having a linear relationship with the bit-rates, which can be seen from Fig. 7.

Therefore, the R-QS and D-QS relationships for each basic unit are modeled in a linear form as follows:

$$\begin{aligned} \text{R-QS}: R_{k,r} &= a_{k,r} \times MAD_{k,r} \times QS_{k,r}^{-1} + b_{k,r} \\ \text{D-QS}: D_{k,r} &= c_{k,r} \times QS_{k,r} + d_{k,r} \end{aligned} \quad (4)$$

where $R_{k,r}$ denotes the average rate (including both the texture and nontexture information) of MBs located in the rth basic unit of the kth frame, $D_{k,r}$ is the average distortion of MBs located in the rth basic unit of the kth frame, and $QS_{k,r}$ is the QS (that will be converted to QP) for the rth basic unit of the kth frame. $a_{k,r}$, $b_{k,r}$, $c_{k,r}$, and $d_{k,r}$ are the parameters of the R-QS and D-QS models, and they can be updated through the linear regression method after coding one frame. For the MR, the complex region, and the flat region, r equals 1, 2, and 3, respectively.

Compared to similar works in [2], [8], [21], and [36], our proposed linear R-QS model and D-QS model have the following major differences. First, the proposed models are linear models performed in the quantization stepsize domain. Second, the difference between the current frame and the global motion compensated previous frame is directly employed as the complexity measurement of the current frame. This outperforms the conventional methods that use the co-located MB in the previous frame, and can produce a better result when there are camera motions. Third, the impact of the nontexture bits (e.g., headers and overheads such as motion

vectors) is significant at low bit-rates. In the proposed R-Q model, the nontexture information is taken into consideration together with the texture information, which can efficiently remove the negative impact from the inaccurate nontexture estimation [21] and can outperform the simple estimation method in [36]. Fourth, three linear models are established individually for the three basic units, which can efficiently exploit the different R-D characteristics of the MBs. Also, since the MBs with similar characteristics are classified into one basic unit, this method can make full use of the inter-frame correlation to accurately establish the R-D models.

C. QP Determination

In order to implement an overall bit-allocation among the basic units and to minimize the distortion under the given bit-allocation, an overall optimization model [as shown in (5)] is established based on the proposed linear R-QS and D-QS models to obtain the QPs for the three basic units

$$\arg\min_{QS_{k,r}\in Q} \sum_{r=1}^{3} \left[N_{k,r} \times \left(c_{k,r} \times QS_{k,r} + d_{k,r} \right) \right]$$
$$\text{s.t.} \quad \sum_{r=1}^{3} \left[N_{k,r} \times \left(a_{k,r} \times MAD_{k,r} \times QS_{k,r}^{-1} + b_{k,r} \right) \right] \leq R_k \tag{5}$$

where $N_{k,r}$ denotes the number of MBs in the rth basic unit of the kth frame, Q is the set of all possible quantization steps, R_k is the total bit-allocation for the entire kth frame, which can be obtained through the frame-layer bit-allocation method in [9]. The calculated QS will be further converted into QP. The dynamic programming [6], [39] is employed to solve the above problem and to obtain the QPs for the three basic units. The above scheme is labeled as region-based basic unit rate control, type 1 (RBUR_T1).

Although RBUR_T1 can achieve an appropriate objective quality, the subjective quality, especially for the MR, sometimes does not look good. Since the MR usually attracts more attention and needs more bits, two amendments (namely QS constraint and Lagrange multiplier adjustment) are explicitly added to RBUR_T1 in order to further improve the subjective quality. The modified scheme is labeled as RBUR_T2.

First, for the QS constraint, the optimization model (5) is modified to (6) to assign a relative small QS (more bits) to the MR, and to assign a relative large QS (fewer bits) to the complex region and the flat region

$$\arg\min_{QS_{k,r}\in Q} \sum_{r=1}^{3} \left[N_{k,r} \times \left(c_{k,r} \times QS_{k,r} + d_{k,r} \right) \right]$$
$$\text{s.t.} \quad \sum_{r=1}^{3} \left[N_{k,r} \times \left(a_{k,r} \times MAD_{k,r} \times QS_{k,r}^{-1} + b_{k,r} \right) \right] \leq R_k$$
$$\text{and } QS_{k,r} \leq QS_{k,r+1}. \tag{6}$$

Second, the Lagrange multiplier (adopted for mode decision and ME) in H.264/AVC [1], [9] plays an important role for the R-D optimization (RDO). With the cost function $J = D + \lambda \times R$ where D represents the distortion, R represents the rate, and λ represents the Lagrange multiplier. However, this Lagrange multiplier is only related to QP and no property of the input signal is considered. The Lagrange multiplier should be adjusted to dynamically adapt to different videos, as well as different regions [13], [34], [38]. Since the characteristics of

different regions are separately represented with the individual linear R-QS and D-QS models, it is reasonable to adjust the Lagrange multiplier corresponding to different regions by taking these models into consideration. On the other hand, the RDO can be treated as a refining process to further keep the perceptual quality of the MR. Since a smaller multiplier corresponds to a lower distortion and a higher rate and vice versa [13], a relatively smaller multiplier should be assigned to MR. Therefore, with the consideration of above issues the Lagrange multiplier for mode decision is proposed to be adjusted through the following equations:

$$\lambda_{k,r} = \beta_{k,r} \times \lambda_{k,r}^{\text{Org}} = \beta_{k,r} \times \alpha \times 2^{(QP_{k,r}-12)/3} \tag{7}$$

where $\lambda_{k,r}$ is the adjusted Lagrange multiplier for the rth basic unit in the kth frame, $\lambda_{k,r}^{\text{Org}}$ is the original Lagrange multiplier for mode decision in H.264/AVC, α is a constant that is recommended to be 0.85 for H.264/AVC [9], and $QP_{k,r}$ is the QP obtained through (6). $\beta_{k,r}$ is the calculated by referring to both the adjustment method in [13] and the conventional λ determination method (by setting the derivative of the cost function J to zero [34])

$$\delta_{k,r} = |c_{k,r}/a_{k,r}|$$
$$\beta_{k,r} = N_{k,r} \times \delta_{k,r} / \sum_{i=1}^{3} \left(N_{k,i} \times \delta_{k,i} \right) \tag{8}$$

where $a_{k,r}$ and $c_{k,r}$ are the first-order parameters of the proposed R-QS and D-QS models for the rth basic unit in (4), respectively. $N_{k,r}$ denotes the amount of MBs. The Lagrange multiplier for ME equals the square root of $\lambda_{k,r}$.

Although the coding efficiency of RBUR_T2 is slightly degraded compared to RBUR_T1, the perceptual quality of RBUR_T2 is improved, which will be confirmed by the experimental results in Section IV.

D. Summary of the Proposed Region-Based Rate Control

The proposed region-based rate control scheme mainly includes five steps, which are summarized as follows. Note that the other steps of our scheme (e.g., the rate control for intra-frames and GOP-layer rate control) are the same as JVT-G012 described in [9].

1) *Region Division:* GMP is performed to estimate the GMV. Based on the GMV, the difference between the current frame and the corresponding global motion compensated previous frame for an MB is evaluated by (1). According to the calculated difference, MBs are classified into three regions, namely, the MR, the flat region, and the complex region through (2) and (3).

2) *Frame-Layer Bit-Allocation:* The frame-layer bit-allocation method of JVT-G012 [9] is employed to determine the target bit for a frame, which is described as follows:

$$R_k = \mu \times RM_k / NF_k + (1 - \mu) \times B_k \tag{9}$$

where R_k denotes the target number of bits of the kth frame, RM_k and NF_k are the amount of remaining bits and the number of remaining frames in the current GOP before encoding the kth frame, μ is a constant and is set as 0.5 according to JVT-G012, and B_k is the available bit-rate considering the current buffer occupancy and the target buffer level.

TABLE II
COMPARISON OF CODING PERFORMANCE (QCIF)

Target Bit-Rate		Bit-Rate (kb/s)				PSNR (dB)			
(Sequence)		MBL	FL	RBUR T1	RBUR T2	MBL	FL	RBUR T1	RBUR T2
20 kb/s	*News*	20.14	20.02	20.07	19.99	31.63	32.09	32.73	32.47 (+0.86)
	Grandma	20.07	20.01	19.96	19.94	35.79	36.58	36.92	36.78 (+0.99)
28 kb/s	*Container*	28.03	28.09	28.22	28.19	36.37	36.48	37.03	36.91 (+0.54)
	Salesman	28.06	27.98	27.98	28.21	33.91	34.17	35.11	34.55 (+0.64)
36 kb/s	*Akiyo*	36.09	36.10	35.99	35.93	40.81	41.65	42.22	42.09 (+1.28)
	Hall	36.13	36.04	36.06	36.08	37.80	37.91	38.48	38.31 (+0.51)
40 kb/s	*Mobile*	40.18	40.16	40.17	40.16	23.64	23.71	23.97	23.92 (+0.28)
	City	40.11	40.06	40.12	40.10	31.03	31.21	31.37	31.31 (+0.28)
Avg. gain							*+0.35*	*+0.86*	*+0.67*

3) *Optimal QP Determination:* In order to implement an overall optimal bit-allocation among the basic units and to minimize the distortion under the given bit-allocation R_k, the optimization model (6) is established based on the proposed linear R-QS and D-QS models described in (4). In (6), a QS constraint is explicitly added to differentiate the importance of different regions. The dynamic programming is used to solve (6) and to obtain the optimal QPs for the three regions. To reduce the blocking artifacts and limit the calculated $QP_{k,r}$ within the range of [1, 51], the final QP of the rth basic unit in the kth frame $QP_{k,r}$ is bounded by

$$QP_{k,r} = \min\{\max\{QP_{k,r}, QP_{k-1,r} - a, 1\}, QP_{k-1,r} + b, 51\}$$
(10)

where a equals 3, 3, 2 and b equals 2, 3, 3 for the MR, the complex region, and the flat region, respectively.

4) *Performing RDO:* According to the obtained optimal QP, RDO is performed for each MB in the current frame with the adjusted Lagrange multiplier.

5) *Parameters Update:* After the encoding of each frame, the parameters of the proposed linear R-QS and D-QS models are updated using the linear regression method.

Note that there are several region-based rate control schemes proposed recently [2]–[5]. Most of them exploit the features of HVS to allocate more bits to the basic units located in the heuristically determined ROI compared to the other basic units located in the non-ROI. However, the proposed region-based basic unit rate control is different from the conventional region-based rate control in the following three aspects.

First, the conventional region-based rate control schemes are MB-layer rate control schemes. Since the correlation between the current MB and its co-located MB of the previous frame may not be so strong due to local object motions or the global camera motion, MB-layer rate control schemes usually result in the worst coding efficiency when compared with the larger size basic unit rate control schemes. In the proposed method, MBs with similar characteristics are classified into the same region, and the entire region is treated as a basic unit for rate control. Thus, the proposed method is a kind of "region-based basic unit" rate control, rather than a "region of interest" rate control.

Second, in the conventional region-based methods, the MBs located in an MR are allocated with more bits by adopting larger weighting factors in the bit-allocation strategy than the MBs located in the non-MR. However, the weighting factors are usually constants and determined in heuristic ways that may not achieve good results. In order to implement an overall bit-allocation among the region-based basic units, an overall optimization model is established based on the proposed linear R-QS and D-QS models to obtain the QPs for the three kinds of basic units.

Third, the conventional region-based methods focus on the subjective quality of the ROI, while the overall objective quality may be degraded. However, it is hard to exactly define the ROI. ROI may vary with different applications and different observers. On the other hand, it is hard to find a widely acceptable ROI extraction method. The skin-tone detection method was employed to extract the ROI in [2], which is limited to the conversational video communication applications. The region-based rate control scheme proposed in [3] (originally proposed for MPEG 4) extracted the ROI only according to the inter-frame difference, which may not work well when there is a camera motion. Differently, the proposed region-based basic unit rate control scheme is aiming at improving the overall coding efficiency. In the proposed scheme, the region-division criterion is the different characteristics of different MBs, and the MBs with similar R-D characteristics are classified into the same region. The experimental results demonstrate that the proposed region-based rate control scheme can achieve both better subjective and the overall objective quality (especially for the MR and the complex region) by performing the rate control adaptively with the content.

In our previous work [28], we proposed a region-based rate control scheme using inter-layer information for scalable video coding (SVC). However, it should be noted that this paper is quite different from [28]. First, [28] was mainly aiming at SVC. The main contributions were to fully exploit the inter-layer information to help the MR division and the non-MR sub-division. In the work, the rate control was the MB-layer rate control scheme. However, in this paper, it is not related to exploiting the inter-layer information at all, and a region-based basic unit rate control is proposed to improve the coding efficiency. It is mainly based on the observation that different regions have different RD behaviors and should use different RD models. Second, in [28], a group of predefined weighting factors were used to allocate different bits to the MBs located in different regions. The R-D model adopted to calculate QP was the conventional quadratic model. However,

TABLE III
COMPARISON OF CODING PERFORMANCE (CIF)

Target Bit-Rate (Sequence)		Bit-Rate (kb/s)				PSNR (dB)			
		MBL	FL	RBUR T1	RBUR T2	MBL	FL	RBUR T1	RBUR T2
32 kb/s	*Container*	32.05	32.00	32.23	32.20	32.11	32.31	32.53	32.53 (+0.42)
	Foreman	32.20	32.62	32.89	32.68	28.73	28.87	28.98	28.89 (+0.16)
64 kb/s	*Hall*	64.21	64.17	64.10	64.08	35.55	35.84	36.24	36.17 (+0.62)
	News	64.32	63.95	64.31	64.03	34.50	34.94	35.42	35.19 (+0.69)
96 kb/s	*City*	96.27	95.87	96.43	96.38	29.79	30.40	30.52	30.51 (+0.72)
	Akiyo	96.21	96.04	96.27	96.35	41.36	42.02	42.30	42.30 (+0.94)
128 kb/s	*Paris*	128.51	128.07	128.03	128.21	30.39	30.46	31.51	31.30 (+0.91)
	Waterfall	128.27	128.28	128.17	128.45	33.80	33.85	34.34	34.25 (+0.45)
Avg. gain							*+0.31*	*+0.70*	*+0.61*

Fig. 8. Comparison of the subjective visual quality of the 38th frame of *Paris* at CIF 128 kb/s. (a) Original. (b) Frame layer (FL). (c) RBUR T1. (d) RBUR T2.

in this paper, in order to implement an overall bit-allocation among the region-based basic units, an overall optimization model is established based on the proposed linear R-QS and D-QS models to obtain the QPs for the three basic units. So, the contributions of this paper are quite different from those of the previous publication.

IV. EXPERIMENTS

In this section, we evaluate the performance of the proposed region-based rate control scheme for H.264/AVC. The experiments are carried out based on the H.264/AVC reference software JM16.0 [24]. The frame-layer rate control (that has the largest basic unit, labeled as "FL") and the MB-layer rate control (that has the smallest basic unit, labeled as "MBL") of JVT-G012 [9] are adopted to compare with the proposed scheme. The test sequences are intra-coded for the first frame (I-frame) and followed with subsequent inter-coded frames (P-frames). The buffer size is set as 0.5*bit-rate. The frame rate is set to 15 f/s. RDO and CABAC are enabled. The experimental results are shown in Tables II and III, where "RBUR T2" denotes the proposed region-based rate control scheme and "RBUR T1" denotes the proposed scheme excluding the QS constraint and the Lagrange multiplier

adjustment. The average gain in the table is calculated by comparing with the MB-layer rate control scheme.

The experimental results demonstrate that our scheme outperforms both the MB-layer and the frame-layer rate control schemes of JVT-G012. For the QCIF resolution, the proposed scheme (RBUR T2) can gain 0.67 dB and 0.32 dB in the average PSNR when comparing with the MB-layer scheme and the frame-layer scheme, respectively. For the CIF resolution, the proposed scheme can gain 0.61 dB and 0.30 dB in the average PSNR, when comparing with the MB-layer scheme and the frame-layer scheme, respectively.

Because of adopting the QS constraint and the Lagrange multiplier adjustment, the objective quality (in terms of the PSNR) of "RBUR T2" is not as good as that of the "RBUR T1." It can be observed from Tables II and III that compared with the "RBUR T1," the PSNR gains of "RBUR T2" are degraded by 0.19 dB and 0.09 dB for QCIF resolution and CIF resolution, respectively. However, the subjective quality of "RBUR T2" is better than that of "RBUR T1," which can be observed through some reconstructed frames in Fig. 8. When comparing with the frame-layer scheme, both the proposed schemes "RBUR T1" and "RBUR T2" can achieve a better perceptual quality of the man and woman in Fig. 8. Moreover, when comparing with "RBUR T1," the perceptual quality of

TABLE IV

PSNR COMPARISON OF DIFFERENT REGIONS (150 FRAMES)

Sequence	Algorithms	Bit-Rate (kb/s)	PSNR (dB)			
			Whole	Flat Region	Complex Region	Moving Region
Hall (QCIF, 36 kb/s)	MBL	36.13	37.80	40.49	37.87	36.89
	FL	36.04	37.91	40.93	37.79	36.87
	RBUR T2	36.08	38.31	40.88	38.46	37.31
Mobile (QCIF, 36 kb/s)	MBL	36.11	24.73	26.51	25.06	24.55
	FL	36.05	24.85	26.37	25.13	24.54
	RBUR T2	36.23	25.04	26.50	25.37	24.74
Container (CIF, 96 kb/s)	MBL	96.05	34.76	39.47	33.04	35.17
	FL	96.06	35.73	40.17	35.16	35.35
	RBUR T2	97.44	35.90	40.23	35.40	35.57
Waterfall (CIF, 96 kb/s)	MBL	96.33	32.49	33.76	33.16	32.23
	FL	95.89	32.68	33.09	32.91	32.49
	RBUR T2	96.08	33.05	33.43	33.31	32.91
City (4CIF, 512 kb/s)	MBL	513.11	32.82	35.31	32.95	32.42
	FL	514.14	33.33	35.41	33.44	32.92
	RBUR T2	514.58	33.50	35.52	33.64	33.12
Mobile (4CIF, 512 kb/s)	MBL	514.33	25.57	30.35	25.64	24.85
	FL	511.93	25.74	30.51	25.82	24.96
	RBUR T2	515.61	25.93	30.71	26.04	25.12
Mobcal (1280*720, 1024 kb/s)	MBL	1025.55	32.47	36.07	32.56	31.73
	FL	1020.61	33.01	36.61	33.24	32.11
	RBUR T2	1026.11	33.30	36.74	33.66	32.38
Shield (1280*720, 1024 kb/s)	MBL	1025.83	33.68	36.05	34.24	32.18
	FL	1022.20	34.14	36.17	34.52	32.74
	RBUR T2	1018.44	34.28	36.23	34.70	32.86
Avg. gain	*FL*		*+0.38*	*+0.16*	*+0.44*	*+0.24*
	RBUR T2		*+0.62*	*+0.28*	*+0.76*	*+0.50*

TABLE V

COMPARISON OF CODING PERFORMANCE WITH DIFFERENT FRAME SKIP NUMBERS (CIF, 100 FRAMES)

Frame Skip Number	Sequence	Bit-Rate (kb/s)		PSNR (dB)							
		FL	RBUR T2	Whole		Flat Region		Complex Region		Moving Region	
				FL	RBUR T2	FL	RBUR T2	FL	RBUR T2	FL	RBUR T2
2	*Hall*	64.15	63.99	37.84	38.00	40.20	40.38	38.20	38.56	36.37	36.35
	Mobile	128.02	128.02	27.70	27.80	30.53	30.55	27.87	28.01	27.41	27.46
Avg. gain					*+0.13*		*+0.10*		*+0.25*		*+0.02*
1	*Hall*	63.89	64.14	36.94	37.18	39.63	39.87	37.16	37.55	35.67	35.87
	Mobile	128.18	128.01	27.38	27.50	30.48	30.56	27.62	27.79	26.81	26.94
Avg. gain					*+0.18*		*+0.16*		*+0.28*		*+0.16*
0	*Hall*	64.04	64.17	35.82	36.08	39.06	39.24	35.74	36.09	34.81	35.08
	Mobile	128.39	127.77	25.80	25.96	29.54	29.64	25.94	26.13	25.29	25.48
Avg. gain					*+0.21*		*+0.14*		*+0.27*		*+0.23*

the tie of the man and the hands and the arms of the woman in Fig. 8(d) is further improved.

The PSNRs of different regions are compared in Table IV, and the frame-by-frame PSNR curves are shown in Fig. 9. The experimental results show that the proposed method can improve the PSNR of both the MR and the complex region, while the PSNR gain of the flat region is not as obvious as the other two regions. Compared to the frame-layer rate control scheme, the proposed scheme can gain 0.26 dB and 0.32 dB in the average PSNR for the MR and the complex region, respectively, while the proposed scheme can gain 0.12 dB in the average PSNR for the flat region.

Note that the intention of the proposed region-based basic unit rate control is to fully explore the inter-frame correlation. With the increase of the interval between two frames, the inter-frame correlation is decreased and the improvement of the proposed method is correspondingly decreased. Some experiments are carried out with different frame skip numbers (the number of frames to be skipped in input sequence, e.g., 2 will code every third frame). The result is shown in Table V. The average gain is calculated by comparing it with the frame-layer rate control scheme. It is obvious that the objective quality of the MR decreases more rapidly than the other two regions. This is because the inter-frame correlation of the MR between the neighboring frames is decreased and the motion compensation for the MR is worse with the increase of frame skip numbers. However, in terms of the overall quality, the proposed method can outperform the frame-layer rate control regardless of different frame skip numbers.
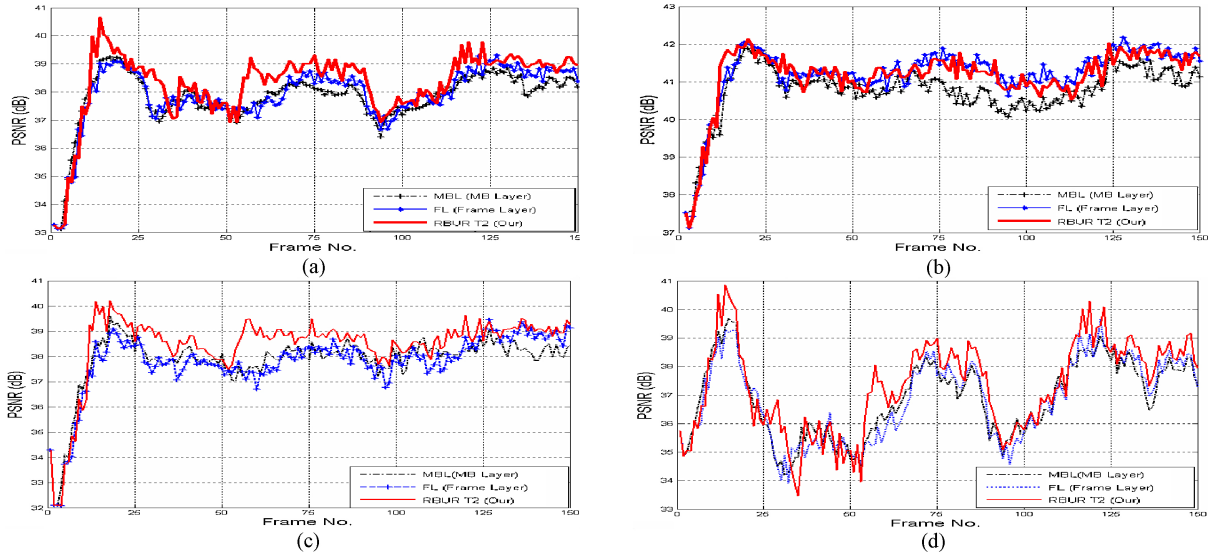
Fig. 9. Frame-by-frame PSNR curves of different regions (*Hall* QCIF 36 kb/s). (a) PSNR curve of the whole picture. (b) PSNR curve of the flat region. (c) PSNR curve of the complex region. (d) PSNR curve of the MR.

When comparing with the conventional basic unit rate control, the computational complexity of the proposed scheme is increased since the proposed region division process employs GPM to obtain the GMV. More experiments are performed to evaluate the computational complexity, and the experimental results are shown in Table VI. Compared to the MB-layer and the frame-layer schemes, the encoding time of the proposed scheme is increased by 1.15% and 0.87%, respectively. Therefore, the computational complexity increase is negligible. Furthermore, it can be reduced by applying the SKIP mode early determination for the flat region, which will be discussed in detail in Section V.

In conclusion, the above experimental results demonstrate that the proposed approaches can achieve both higher subjective and objective quality and can outperform the conventional MB-layer rate control and the frame-layer rate control schemes.

V. DISCUSSION

Although the proposed approach is originally aiming at low bit-rate applications, it can be easily extended to applications with a wider range of bit-rates. Several experiments have been performed to evaluate the efficiency of the proposed approach in a wider range of bit-rates. The buffer size is set to 0.5*bit-rate. In order to evaluate the buffer performance, when the buffer fullness is larger than a safety margin (conventionally, 0.8*Buffer-Size is used) of the buffer size, the encoder skips frames and the previous frames are used in the PSNR computation. The frame-layer rate control scheme (labeled as "FL") and the MB-layer rate control scheme (labeled as "MBL") of JVT-G012 [9] are adopted to compare with the proposed scheme (labeled as "RBUR T2"). The experimental results are shown in Table VII. The frame-by-frame PSNR curves and buffer fullness are compared in Fig. 10. The average gain is calculated by comparing with the MB-layer rate control.

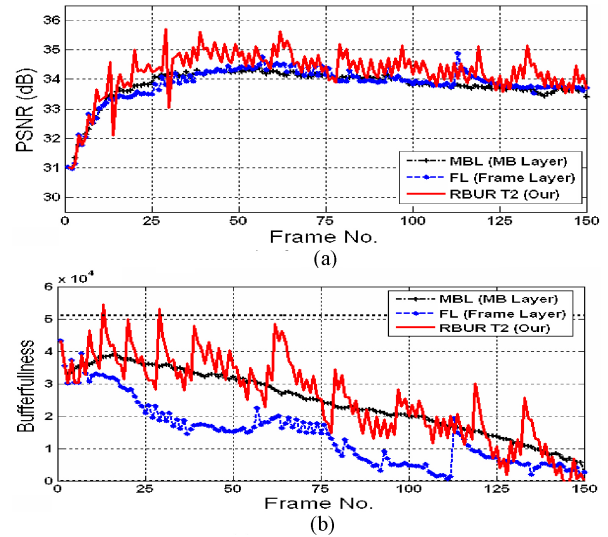The experimental results demonstrate that the proposed scheme can obtain a better coding result in a wider range of

Fig. 10. Frame-by-frame PSNR curves and buffer fullness (*Waterfall*, CIF, 128 kb/s, buffer size = 0.5*128000). (a) Frame-by-frame PSNR curves. (b) Buffer fullness.

bit-rates. Note that the buffer fullness fluctuation of the proposed scheme is slightly larger than the other two schemes and the frame skip number of the proposed scheme is also slightly increased. However, the increase is limited and, compared with the MB-layer and the frame-layer rate control schemes, the proposed approach can gain 0.34 dB and 0.19 dB in average PSNR, respectively. Moreover, the proposed scheme is aimed at the low bit-rate applications instead of the delay-sensitive applications and the proposed scheme can be easily extended to maintain a smooth buffer fullness, which will be discussed in detail as follows.

It can be observed from Fig. 11 that the bit-rate of the MR has a larger fluctuation than the other two kinds of regions. This implies that the bit-rate fluctuation of the whole picture mainly comes from the MR. In compliance with the proposed region-based basic unit, a low-delay rate control

TABLE VI

COMPARISON OF COMPUTATIONAL COMPLEXITY (150 FRAMES)

Resolution (Target Bit-Rate)	Sequence	Bit-Rate (kb/s)			PSNR (dB)			Encoding Time (ms/f)*		
		MBL	FL	RBUR T2	MBL	FL	RBUR T2	MBL	FL	RBUR T2
QCIF 88 kb/s	*Mobile*	88.09	88.07	88.40	28.14	28.28	28.49	391.60	391.93	397.40
	News	88.22	87.71	87.90	40.98	41.00	41.61	329.00	333.40	337.33
CIF 224 kb/s	*Container*	224.10	224.84	224.05	38.42	38.60	38.91	1345.93	1349.47	1344.87
	Waterfall	224.39	224.24	223.81	36.09	36.10	36.49	1398.53	1401.73	1404.27
4CIF 1536 kb/s	*Piano*	1537.19	1536.49	1532.05	32.08	32.31	32.37	5840.80	5830.53	5911.73
	Mobile	1536.06	1538.74	1535.33	29.32	29.51	29.59	6042.93	6040.20	6123.53
Avg.						+0.13	+0.41		+0.28%	+1.15%

*Average encoding time by encoding each sequence for five times. Hardware platform: CPU: Intel(R) Core (TM) 2 at 2.66 GHz, memory: 3.37 GB.

TABLE VII

COMPARISON OF CODING PERFORMANCE FOR A WIDE RANGE OF BIT-RATES

Sequence (Target Bit-Rate, kb/s)		Bit-Rate (kb/s)			Frame Skip Number			PSNR (dB)		
		MBL	FL	RBUR T2	MBL	FL	RBUR T2	MBL	FL	RBUR T2
Container (QCIF)	160	160.11	160.11	160.45	0	0	0	44.42	44.80	44.99
	96	96.05	95.97	96.60	0	0	0	41.87	42.04	42.29
	36	36.04	36.22	35.98	0	0	0	37.40	37.50	37.87
News (QCIF)	160	160.68	159.73	160.17	0	0	0	45.12	45.19	45.58
	96	96.20	95.77	95.90	0	0	0	41.52	41.70	42.14
	36	36.22	35.99	36.09	0	0	0	35.07	35.29	35.46
Bus (CIF)	512	512.07	512.66	512.29	0	0	0	33.44	33.49	33.57
	256	256.77	256.02	255.47	0	0	0	29.95	29.93	30.09
	128	128.10	128.16	131.44	2	5	5	26.78	26.63	26.76
Waterfall (CIF)	512	511.99	511.04	511.44	0	0	0	39.41	39.60	39.76
	256	256.23	255.70	256.57	0	0	0	36.60	36.70	36.98
	128	128.27	128.28	127.60	0	0	2	33.80	33.85	34.24
City (4CIF)	2048	2047.60	2050.14	2049.55	0	0	0	37.34	37.53	37.55
	1280	1280.01	1283.66	1282.54	0	0	0	35.86	36.13	36.18
	512	513.11	514.14	514.58	0	0	0	32.82	33.33	33.50
Mobile (4CIF)	2048	2047.94	2053.38	2047.13	0	0	0	30.51	30.71	30.72
	1280	1280.12	1282.24	1278.52	0	0	0	28.61	28.79	28.89
	512	513.55	512.40	510.93	8	10	10	25.22	25.29	25.34
Avg.						+0.28	+0.39		+0.15	+0.34

can be implemented by separately carrying out the different processes for different regions, such as the MAD prediction process, which is important for the low-delay rate control. As for the MR, the switched MAD prediction method proposed in [21] can be employed to enhance the traditional linear MAD prediction model [9], which is not suitable for predicting abrupt MAD fluctuations. As for the other two regions, the linear MAD prediction may achieve an accurate prediction result since there is no high motion existing in these two regions. Therefore, the proposed region-based basic unit rate control can be slightly modified to justify the low-delay constraints.

Although the computational complexity of the proposed rate control scheme is increased compared to the conventional basic unit rate control, the computation complexity increase is negligible and it can be reduced with the consideration of the following two issues. First, in terms of the mode decision, the SKIP mode occupies a large proportion, especially, in low bit-rate applications. Statistical information is listed in Tables VIII and IX. It can be observed that around 90% of the MBs located in the flat region adopt the SKIP mode. Over a half SKIP mode distributes in the flat region that is less important than the other two regions. Therefore, a SKIP mode early determination
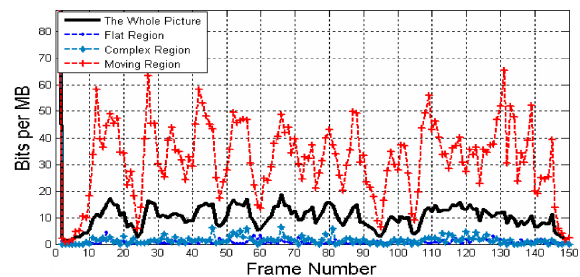

Fig. 11. Average bits of different regions (*Silent* QCIF QP = 36).

process can be implemented for the flat region, which can reduce the computational complexity. Second, the obtained GMV can be referred to by the ME process [37]. This can further reduce the computational complexity.

Although the thresholds used in the region division can achieve acceptable results for most test sequences, a refining threshold determination method is desired. These thresholds could be updated according to different contents, and a more complicated threshold determination method can be implemented (e.g., taking more HVS features into consideration or employing video content analysis methods) to achieve a

TABLE VIII
SKIP MODE DISTRIBUTION RATIO FOR THREE REGIONS
(QCIF, CONSTANT QP = 40)

Sequence	SKIP Mode Distribution Ratio (%)			Prediction Accuracy of SKIP Mode for Flat Region (%)	Coding Results	
	MR	Complex Region	Flat Region		Bit-Rate (kb/s)	PSNR (dB)
Silent	0.17	0.42	0.41	0.99	9.53	29.27
News	0.16	0.30	0.54	0.99	10.64	29.57
Miss America	0.32	0.08	0.59	0.99	4.4	33.47
Mother-daughter	0.29	0.23	0.48	0.94	4.95	30.82
Carphone	0.29	0.22	0.48	0.94	9.72	29.44
Table	0.11	0.26	0.63	0.86	20.62	28.54
Football	0.09	0.24	0.67	0.72	59.04	28.43
Avg.	*0.21*	*0.25*	*0.54*	*0.93*		

TABLE IX
SKIP MODE DISTRIBUTION RATIO FOR THREE REGIONS
(QCIF, CONSTANT BIT-RATE = 32 KB/S)

Sequence	SKIP Mode Distribution Ratio (%)			Prediction Accuracy of SKIP Mode for Flat Region (%)	Coding Results	
	MR	Complex Region	Flat Region		Bit-Rate (kb/s)	PSNR (dB)
Silent	0.08	0.55	0.37	0.95	32.09	33.94
News	0.11	0.46	0.43	0.98	32.07	34.53
Miss America	0.10	0.14	0.64	0.87	31.94	42.20
Mother-daughter	0.15	0.28	0.57	0.95	32.08	37.76
Carphone	0.19	0.24	0.58	0.81	31.94	34.63
Table	0.09	0.34	0.57	0.78	32.01	30.42
Football	0.17	0.24	0.59	0.77	32.06	26.58
Avg.	*0.13*	*0.32*	*0.53*	*0.87*		

more accurate region division. Moreover, some mathematical methods, such as cluster analysis, can also be included to achieve more accurate MB classification. The above methods may result in an increase of computational complexity. However, they may further improve the coding efficiency and could be one of our future works.

VI. CONCLUSION

In this paper, we proposed a region-based rate control scheme for H.264/AVC. The video frames were divided into three regions according to their different characteristics. Each entire region was treated as a basic unit, and a linear R-QS model and a linear D-QS model were proposed for different basic units. Based on this, an optimization model was established to calculate the optimal QPs for the three basic units. Two amendments were added to achieve not only a good objective quality, but also a high subjective quality. The experimental results demonstrated that the proposed region-based rate control outperformed the conventional basic unit rate control schemes. Compared with the MB-layer scheme and the frame-layer scheme, the proposed scheme can gain around 0.6 dB and 0.25 dB in the average PSNR, respectively.

Although the proposed scheme was originally aiming at the low bit-rate applications, it can be easily extended to appli-

cations with a wider range of bit-rates. Moreover, the region-based scheme can also be easily extended to smooth the bit-rate fluctuation, which is important for low-delay applications. More complicated threshold determination methods could also be employed to achieve a more accurate region division, which may improve the coding efficiency further.

REFERENCES

[1] JVT, *Advanced Video Coding (AVC)*, ITU-T Rec.H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10), 3rd ed., 2004.

[2] Y. Liu, Z. G. Li, and Y. C. Soh, "Region-of-interest based resource allocation for conversational video communication of H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 134–139, Jan. 2008.

[3] Y. Sun, I. Ahmad, D. Li, and Y. Q. Zhang, "Region-based rate control and bit allocation for wireless video transmission," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 1–10, Feb. 2006.

[4] M. C. Chi, M. J. Chen, C. H. Yeh, and J. A. Jhu, "Region-of-interest video coding based on rate and distortion variations for H.263+," *Signal Process.: Image Commun.*, vol. 23, no. 2, pp. 127–142, Feb. 2008.

[5] H. Song and C. C. J. Kuo, "A region-based H.263+ codec and its rate control for low VBR video," *IEEE Trans. Multimedia*, vol. 6, no. 3, pp. 489–500, Feb. 2004.

[6] Z. Chen and K. N. Ngan, "Recent advances in rate control for video coding," *Signal Process. Image Commun.*, vol. 22, no. 1, pp. 19–38, Jan. 2007.

[7] 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 Representation*, vol. 17, no. 2, pp. 376–406, Apr. 2006.

[8] D. K. Kwon, M. Y. Shen, and C. C. J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 5, pp. 517–529, May 2007.

[9] Z. G. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," presented at the 7th JVT Meeting, JVT-G012-rl, Pattaya, Thailand, Mar. 2003.

[10] 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.

[11] S. R. Sergio, A. E. Óscar, F. L. Manuel, and D. M. Fernando, "Cauchy-density-based basic unit layer rate controller for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 8, pp. 1139–1143, Aug. 2010.

[12] J. Liu, Y. Cho, Z. Guo, and C. C. J. Kuo, "Bit allocation for spatial scalability coding of H.264/SVC with dependent rate-distortion analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 7, pp. 967–981, Jul. 2010.

[13] M. Jiang and N. Ling, "On Lagrange multiplier and quantizer adjustment for H.264 frame-layer video rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 5, pp. 663–669, May 2006.

[14] H. Wang and S. Kwon, "Rate-distortion optimization of rate control for H.264 with adaptive initial quantization parameter determination," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 140–144, May 2008.

[15] J. Dong and N. Ling, "A context-adaptive prediction scheme for parameter estimation in H.264/AVC macroblock layer rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1108–1117, May 2009.

[16] K. Y. Liao, J. F. Yang, and M. T. Sun, "Rate-distortion cost estimation for H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 1, pp. 38–49, May 2010.

[17] Y. Zhou, Y. Sun, Z. Feng, and S. Sun, "New rate-distortion modeling and efficient rate control for H.264/AVC video coding," *Signal Process. Image Commun.*, vol. 24, no. 5, pp. 345–356, May 2009.

[18] X. Li, A. Hutter, and A. Kaup, "Efficient one-pass frame level rate control for H.264/AVC," *J. Vis. Commun. Image Representation*, vol. 20, no. 8, pp. 585–594, Sep. 2009.

[19] S. C. Hsia and S. H. Wang, "Adaptive video coding control for real-time H.264/AVC encoder," *J. Visual Commun. Image Representation*, vol. 20, no. 7, pp. 463–477, 2009.

[20] J. Dong and N. Ling, "A context-adaptive prediction scheme for parameter estimation in H.264/AVC macroblock layer rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 8, pp. 1108–1117, Aug. 2009.

[21] 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.

[22] B. T. Wu and J. I. Guo, "Low compute complexity BU-based rate control algorithm for H.264/AVC encoder," in *Proc. IEEE APCCAS*, Dec. 2010, pp. 564–567.

[23] Y. Sun, Y. Zhou, Z. Feng, and Z. He, "An incremental basic unit level QP determination algorithm for H.264/AVC rate control," *Proc. SPIE*, vol. 7257, p. 72571N, Jan. 2009.

[24] JVT. (2009, Sep.). *H.264/AVC Reference Software (JM16.0)* [Online]. Available: http://iphome.hhi.de/suehring/tml

[25] Y. Liu, Z. G. Li, Y. C. Soh, and M. H. Loke, "Conversational video communication of H.264/AVC with region of interest concern," in *Proc. IEEE ICIP*, Oct. 2006, pp. 3129–3132.

[26] N. Doulamis, A. Doulamis, D. Kalogeras, and S. Kollias, "Low bit-rate coding of image sequences using adaptive regions of interest," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 8, pp. 928–934, Jan. 1998.

[27] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao, "Modeling visual attention's modulatory after effects on visual sensitivity and quality evaluation," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1928–1942, Nov. 2005.

[28] H. M. Hu, W. Lin, B. Li, and M. T. Sun, "A region-based rate control scheme using inter-layer information for H.264/SVC," *J. Vis. Commun. Image Representation*, vol. 22, no. 7, pp. 615–626, Oct. 2011.

[29] H. Li, Z. B. Wang, H. J. Cui, and K. Tang, "An improved ROI-based rate control algorithm for H.264/AVC," in *Proc. IEEE ICSP*, vol. 2. Nov. 2006, pp. 16–20.

[30] D. Barreto, L. D. Alvarez, and R. Molina, "Region-based super-resolution for compression," *Multidimensional Syst. Signal Process*, vol. 18, nos. 2–3, pp. 59–81, Mar. 2007.

[31] B. Tao, B. W. Dickinson, and H. A. Peterson, "Adaptive model-driven bit allocation for MPEG video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 1, pp. 147–157, Feb. 2000.

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

[33] H. J. Lee, T. Chiang, and Y.-Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 878–894, Sep. 2000.

[34] X. Li, N. Oertel, A. Hutter, and A. Kaup, "Laplace distribution based Lagrangian rate distortion optimization for hybrid video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 2, pp. 193–205, Feb. 2009.

[35] G. Ren, P. Li, and G. Wang, "A novel hybrid coarse-to-fine digital image stabilization algorithm," *Inform. Technol. J.*, vol. 9, no. 7, pp. 1390–1396, 2010.

[36] 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.

[37] W. Lin, K. Panusopone, D. Baylon, and M. T. Sun, "A computation control motion estimation method for complexity scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1533–1543, Nov. 2010.

[38] M. Wang and B. Yan, "Lagrangian multiplier based join three-layer rate control for H.264," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 679–682, Aug. 2009.

[39] S. P. Bradly, A. C. Hax, and T. L. Magnanti, *Applied Mathematical Programming*. Reading, MA: Addison-Wesley, 1977, pp. 320–362.

Hai-Miao Hu received the B.S. degree from Central South University, Changsha, China, in 2005, and the Ph.D. degree from Beihang University, Beijing, China, in 2012, all in computer science.

Currently, he is with the Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University. He was a Visiting Student with the University of Washington, Seattle, from 2008 to 2009. His current research interests include video coding and networking, image and video processing, video surveillance, video analysis, and high-efficiency video coding.

Bo Li received the B.S. degree from Chongqing University, Chongqing, China, in 1986, the M.S. degree from Xi'an Jiaotong University, Xi'an, China, in 1989, and the Ph.D. degree from Beihang University, Beijing, China, in 1993, all in computer science.

He is currently a Professor of computer science and engineering with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, and the Director of the Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering. He has published over 100 conference and journal papers in diversified research fields, including digital video and image compression, video analysis and understanding, remote sensing image fusion, and embedded digital image processors.

Weiyao Lin received the B.E. and M.E degrees from Shanghai Jiao Tong University, Shanghai, China, in 2003 and 2005, respectively, and the Ph.D. degree from the University of Washington, Seattle, in 2010, all in electrical engineering.

Since 2010, he has been an Assistant Professor with the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University. His current research interests include video processing, machine learning, computer vision, video coding, and compression.

Wei Li received the B.S. and Ph.D. degrees in computer science and engineering from Beihang University, Beijing, China, in 1995 and 2003, respectively.

He is an Associate Professor with the Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University. His current research interests include video coding, multimedia systems, very large scale integration design for video coding, and video analysis and understanding.
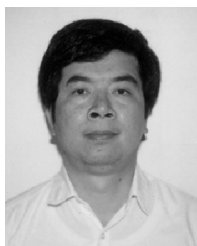
Ming-Ting Sun (S'79–M'81–SM'89–F'96) received the B.S. degree from National Taiwan University, Taipei, Taiwan, in 1976, the M.S. degree from the University of Texas, Arlington, in 1981, and the Ph.D. degree from the University of California, Los Angeles, in 1985, all in electrical engineering.

He joined the University of Washington, Seattle, in August 1996, where he is currently a Professor with the Department of Electrical Engineering. Previously, he was the Director of the Video Signal Processing Research Group, Bellcore (now Telcordia). He has been a Chair Professor with Tsinghwa University, Beijing, China, and a Visiting Professor with Tokyo University, Tokyo, Japan, and National Taiwan University. He holds 11 patents and has published over 200 technical papers, including 13 book chapters in the areas of video and multimedia technologies. He co-edited the book, *Compressed Video Over Networks*.

Dr. Sun was the Editor-in-Chief of the IEEE TRANSACTIONS ON MULTIMEDIA and a Distinguished Lecturer of the Circuits and Systems Society from 2000 to 2001. He received the IEEE CASS Golden Jubilee Medal in 2000, and was the General Co-Chair of the Visual Communications and Image Processing 2000 Conference. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 1995 to 1997. He received the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award in 1993. From 1988 to 1991, he was the Chairman of the IEEE CAS Standards Committee, and established the IEEE Inverse Discrete Cosine Transform Standard. He received the Award of Excellence from Bellcore for his work on the digital subscriber line in 1987.