λ-Domain Rate Control Algorithm for HEVC Scalable Extension

Li Li, Bin Li, *Member, IEEE*, Dong Liu, *Member, IEEE*, and Houqiang Li, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a λ-domain rate control algorithm for high efficiency video coding (HEVC) scalable extension. All the commonly used scalabilities including temporal, spatial, and quality scalability are taken into consideration. The proposed algorithm mainly has three key contributions. First, we propose an optimal initial target bits and initial encoding parameters determination algorithm for the first frame of each layer to achieve the best rate-distortion (R-D) performance. Second, an optimal bit allocation algorithm taking both the intra and inter layer dependence into consideration is proposed for the inter frames under spatial and quality scalability cases. Third, since the coding scheme of HEVC scalable extension with multiple layers is even more flexible than HEVC, an adaptive updating algorithm for R-λ model is proposed to control the bits per frame even more precisely. The experimental results demonstrate that the proposed λ-domain rate control algorithm can bring both more precise bitrate accuracy and better R-D performance compared with the previous rate control algorithms for HEVC scalable extension.**

***Index Terms*—High efficiency video coding (HEVC), rate control, R-λ model, R-Q model, scalable HEVC (SHVC), scalable video coding (SVC).**

I. Introduction

MULTIMEDIA streaming becomes quite widely used in these days. A number of multimedia services such as video conferencing, Internet protocol television, and other real time video transmission applications rely on the multimedia streaming techniques. However, along with the increasing number of users, the multimedia streaming techniques are now facing more and more challenges. One of the greatest challenges is how to deal with various characteristics of different end devices and various network conditions of different users.

In fact, scalable video coding (SVC) [1] can provide the property that parts of the bitstream can form another valid bitstream for some target decoders. Therefore, SVC is a highly attractive solution for applications where the end devices of different users are with various characteristics. The usual types of scalability include temporal, spatial, and quality scalability. Temporal scalability shows the situation in which part of the bitstream represents the video with reduced frame rate. Spatial scalability describes the case in which subset of the bitstream represents the video with a reduced picture size. Quality scalability provides the bitstream with the same spatio-temporal resolution but lower fidelity.

Although SVC can deal with the various characteristics of different end devices, the rate adaptation capability of SVC to the given target bitrate is far from enough, since the users can only derive different bitrates by selecting different layers. Therefore, a finer bitrate adjustment scheme is necessary for SVC to adapt to the various target bitrate brought by different users. It should be noted that the coarse rate adaptation capability is just a challenge for the scalable extensions of standard video codecs, which is not inherently related to scalability. The scalable video codecs such as 3D-SPIHT [2] and MC-EZBC [3], which are based on the wavelet transform, have extremely flexible rate adaptation capabilities.

In fact, rate control is usually employed in video coding framework to adapt the bitstream to the varying target bitrate. A rate control algorithm can dynamically adjust encoder parameters to achieve a target bitrate. However, rate control is unable to deal with the various characteristics of different end devices. For example, a mobilephone is only able to display video with frame rate of $15fps$, while a laptop can display video with frame rate of $30fps$. An encoder without scalability cannot generate a bitstream to adapt to the mobilephone and the laptop at the same time.

In essence, the main targets of SVC and rate control are different. To be adaptable to the end device with different frame rate, diverse spatial resolution, and various decoding capability is the main goal of SVC, while rate control focuses on adjusting the bitrate of video bitstream. Although they have some overlapped functionality, it is far from enough by only using rate control or SVC to provide the adaptability to both varying target bitrate and diverse capability of end devices. Therefore, we propose to combine these two schemes together. However, it is not easy to design a rate control scheme for SVC, especially for high efficiency video coding (HEVC) [4] scalable extension [5] which supports a number of coding modes by enabling quadtree based coding unit splitting and transform unit splitting.

Firstly, it is difficult to choose a suitable rate-distortion (R-D) function. One class of rate control algorithms builds a relationship between bitrate R and quantization Q [6]–[8], which is named as R-Q model. It is assumed that the target bitrate R can be achieved by selecting a proper Q. That is to say, Q is the key

Manuscript received July 30, 2015; revised March 26, 2016 and May 23, 2016; accepted July 9, 2016. Date of publication July 27, 2016; date of current version September 15, 2016. This work was supported in part by 973 Program under Contract 2013CB329004, in part by the National Key Research and Development Plan under Grant 2016YFC0801001, and in part by the Natural Science Foundation of China under Contract 61325009, Contract 61390510, and Contract 61272316. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shahram Shirani.

L. Li, D. Liu, and H. Li are with the CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China, Hefei 230027, China (e-mail: lilimao@mail.ustc.edu.cn; dongeliu@ustc.edu.cn; lihq@ustc.edu.cn).

B. Li is with Microsoft Research Asia, Beijing 100080, China (e-mail: libin@microsoft.com).

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/TMM.2016.2595264

factor to determine R. Another class of rate control algorithms presents a relationship between R and the percentage of zeros among quantized transform coefficients $\boldsymbol{\rho}$ [9]–[11], which is called as R-$\boldsymbol{\rho}$ model. As it is assumed that there is a one-to-one correspondence between $\boldsymbol{\rho}$ and Q, the R-$\boldsymbol{\rho}$ model also implies that Q is the key factor to determine R. However, the assumption that Q is the key factor to determine R cannot hold for very flexible coding scheme HEVC and its scalable extension. In fact, as analyzed in [12], [13], the Lagrange Multiplier λ instead of Q is the key factor to determine R and a hyperbolic R-λ relationship is proposed for rate control of HEVC. Therefore, in this paper, we will first validate the effectiveness of the hyperbolic R-λ model in HEVC scalable extension context and then propose the λ-domain rate control algorithm for HEVC scalable extension.

Secondly, the initial target bits and initial encoding parameters for each layer for HEVC scalable extension are difficult to determine. The simplest method to determine the initial encoding parameters is to build a relationship between the initial quantization parameter (QP) and target bits per pixel (bpp) [14]. Such a method is too simple to achieve the optimal R-D performance. Therefore, several works propose to take the complexity of the first frame into consideration [15]. However, the initial QP is not only influenced by the complexity of the first frame, but also affected by the influence of the current frame on the subsequent frames. Recently, some initial encoding parameters determination methods begin to take the temporal correlation into consideration [16]. However, the temporal correlation is utilized in a coarse way and the achieved R-D performance is still far from the best.

Thirdly, how to estimate the parameters of a rate control model properly is also a serious problem. The Q-domain rate control algorithm always uses the regressive method to update the model parameters [17], [18]. In our previous work, an adaptive least mean square method [13] is proposed to update the model parameters of the R-λ model in HEVC. However, while rate control for HEVC just requires the bitrate of several continuous frames to be steady, rate control for SVC requires that, in any extracted subset bitstreams, the bitrate of several continuous frames should be kept steady. Besides, the inaccuracy of the bits of one layer may be propagated to the other layers due to the reference relationship between different layers. Therefore, the bits of each frame should be kept more steady, which requires the model parameters to be more accurate.

A. Related Work

There have already been many researches on rate control for temporal scalability. One group of them pays attention to improving the overall R-D performance without considering the bitrate accuracy for each layer. For instance, Hu *et al.* [19] propose to use adaptive weighting factors for different frames according to their importance in the bit allocation scheme. Seo *et al.* [20] propose to apply the adaptive weighting factors not only to different frames, but also to different CUs. However, such rate control schemes for SVC cannot satisfy the needs of controlling the bitrate of each layer precisely. Another group of them tries to improve the R-D performance of each layer on the premise of accurate bitrate control. For example, H.264/SVC reference software JSVM [21] adopts a quadratic R-Q model based rate control scheme supporting temporal scalability [22]. However, both the bitrate accuracy and R-D performance are far from the best. Liu *et al.* [23] propose an improved rate control scheme which shows both better bitrate accuracy and R-D performance. However, a quite complex pre-analysis process for the whole sequence, which is unwelcome in a rate control scheme, is used to estimate the model parameters. Binh and Yang [24] propose an intra frame complexity based method to better determine the initial encoding parameters for H.264/SVC.

Besides the rate control algorithms for temporal scalability, there are also many researches focusing on spatial and quality scalability. For example, Xu *et al.* [25] propose a rate control scheme for spatial scalability by employing an improved TMN8 model based on the mode analysis of P/B frames. Liu *et al.* [26] present a rate control algorithm for spatial and quality scalability based on the classic linear R-Q model. Since the above methods always extend the R-D model in the base layer to the enhancement layer and have not taken the inter layer dependency into consideration, the R-D performance achieved is far from the best. To solve this problem, Mansour *et al.* [27] propose to take the inter layer dependency into consideration and model the R-D behavior for the base layer and enhancement layer separately based on the Laplace distribution. Liu *et al.* [28] propose a bit allocation scheme for spatial scalability with inter layer dependent R-D analysis. However, such a bit allocation scheme is unable to achieve the target bitrate for each layer precisely.

All the above listed works consider only temporal or spatial/quality scalability. There are also many researches supporting all kinds of common used scalabilities. Some of them build a relationship between R and Q. For example, Liu *et al.* [29] introduce an accurate linear R-Q model to build the relationship between R and Q. In addition, Yang *et al.* [30] adopt an incremental approach to determine the QP for inter frames. Others build a relationship between R and $\boldsymbol{\rho}$. For instance, Pitrey *et al.* [31] propose a two-pass rate control scheme via R-$\boldsymbol{\rho}$ model, and each frame in a specified layer is allocated the same number of target bits. To improve the R-D performance, the algorithm is further improved that the target bits of different frames in the same layer are determined according to their frame complexity [32]. Besides, Liu *et al.* [33] adopt a linear model between $\boldsymbol{\rho}$ and Q to better characterize their relationship. In summary, although the accuracy of the R-Q or R-$\boldsymbol{\rho}$ model improves along with the development, the essence is that Q is the key factor to determine R. However, as we pointed out, λ other than Q is the key factor to determine R for HEVC.

B. Contribution

In this paper, to overcome the difficulty of rate control for HEVC scalable extension and keep the bits per frame steady, we propose a λ-domain rate control algorithm for HEVC scalable extension. All the commonly used scalabilities including temporal, spatial, and quality scalability are taken into consideration. The proposed algorithm mainly has three key contributions.

Firstly, we propose an optimal initial target bits and initial encoding parameters determination algorithm for the first frame of each layer to achieve the best R-D performance. To determine the best initial target bits, we build a model to characterize the relationship between the optimal initial target bits ratio (the optimal initial target bits ratio is defined as the ratio of the optimal initial target bits to the average bits) and three related factors: the complexity of the current frame, the influence of the current frame on the subsequent frames, and the target bpp of a specified layer. Besides, to achieve the assigned target bits precisely for the first frame, we design a method to estimate the hyperbolic R-λ model parameters of the first frame accurately.

Secondly, an optimal bit allocation algorithm is developed for spatial and quality scalability using the fundamental rate distortion optimization (RDO) theory. When designing the bit allocation algorithm, both the intra and inter layer dependency are taken into consideration. For the base layer, the optimal bit allocation can be achieved by setting the ratio of λ between different frames inversely proportional to the sum of their influence to the base layer and enhancement layer. For the enhancement layer, the optimal bit allocation can be achieved by setting the ratio of λ between different frames inversely proportional to their influence to the enhancement layer. The early work has been published in [34] and this paper provides a solution in a more general case with the importance of the base layer and enhancement layer taken into consideration. Besides, both the ratios of λ between different frames in IBBB/IPPP and hierarchical-B cases are shown in this paper.

Thirdly, to achieve better bitrate accuracy and overcome the difficulty of estimating the proper model parameters for inter frames, an adaptive updating algorithm for R-λ model is proposed. The adaptive updating algorithm can accommodate to various video content and different bitrate, and thus can make each layer reach the target bitrate precisely.

C. Paper Organization

This paper is organized as follows. Section II introduces the R-Q and R-λ model briefly and verifies the hyperbolic R-λ model for HEVC scalable extension. The proposed λ-domain rate control scheme for HEVC scalable extension is described in Section III. The initial encoding parameters determination algorithm, the optimal bit allocation algorithm for spatial/quality scalability, and the adaptive updating algorithm for R-λ model are all explained in details in this section. The experimental results are shown in Section IV. Both the bitrate accuracy and R-D performance improvement of the proposed rate control algorithm for HEVC scalable extension are introduced. Section V concludes the whole paper.

II. Choosing the Suitable R-D Function

In this section, we will firstly introduce the brief idea of R-Q and R-λ model in HEVC and discuss their advantages and disadvantages in Section II-A. Then, in Section II-B, the hyperbolic R-λ model which can lead to accurate bits per frame for HEVC will be verified for its scalable extension.

A. A Brief Introduction to R-Q and R-λ Model in HEVC

The R-Q model is proposed in [35] for MPEG-4 in 1997. At that time, the relationship between R and Q is modeled as

$$R = a \times Q^{-1} + b \times Q^{-2} \tag{1}$$

where a and b are the model parameters related to the video content. Choi *et al.* [36] develop the above algorithm for HEVC in both picture level and basic unit level. Using picture level as an example, the relationship between R and Q in HEVC is modeled as

$$bpp_i(j) = a \times \frac{MAD_{pred,i}(j)}{QP_i(j)} + b \times \frac{MAD_{pred,i}(j)}{QP_i^2(j)} \tag{2}$$

where $bpp_i(j)$ is the target bpp of the jth frame in the ith Group Of Pictures (GOP), $MAD_{pred,i}(j)$ and $QP_i(j)$ are the estimated mean absolute difference (MAD) and QP for that frame respectively, a and b are the model parameters. The $MAD_{pred,i}(j)$ is calculated by

$$MAD_{pred,i}(j) = \frac{1}{N_{pixels}} \sum_{AllPixelsInPic} |P_{org} - P_{pred}| \tag{3}$$

where N_{pixels} is the number of pixels of the jth frame in the ith GOP, P_{org} is the pixel value of the original signal, and P_{pred} is the pixel value of the predicted signal. The relationships built in (1) and (2) are different in outward form, but the same in essence. They all build a relationship between R and Q (or QP). That is to say, both (1) and (2) hold the opinion that Q is the key factor to determine the bitrate. However, it has been analyzed in [12] that the bitrate can be influenced by many factors; not only QP, but also mode, motion and other coding parameters may have significant impact on the bitrate.

To overcome the disadvantages of R-Q model, another rate control scheme, which is based on the R-λ model in HEVC, is proposed in [12], and the relationship between R and λ is modeled as

$$R = \alpha \times \lambda^{\beta} \tag{4}$$

where α and β are two model parameters related to the video content. As β is less than 0, the relationship between R and λ is hyperbolic. Without regard to the model parameters, the relationship in (4) presents a one-to-one correspondence between R and λ, and according to the analysis in [12], the one-to-one correspondence does exist as the R-D curve [37] is convex and λ is the slope of the R-D curve. Certainly, many other parameters such as QP, mode, and motion do have influence on the bitrate R. However, all those parameters are needed to be determined through RDO while λ determines the optimization target of RDO.

After the above introduction and analysis of the models in (2) and (4), we know that λ instead of Q is the key factor on determining the bitrate R in HEVC. The R-λ model presents much more benefits compared with the R-Q model.

B. Validation of R-λ Model for HEVC Scalable Extension

As it has been explained that λ is the key factor to determine R, the next step will be determining the R-λ relationship.

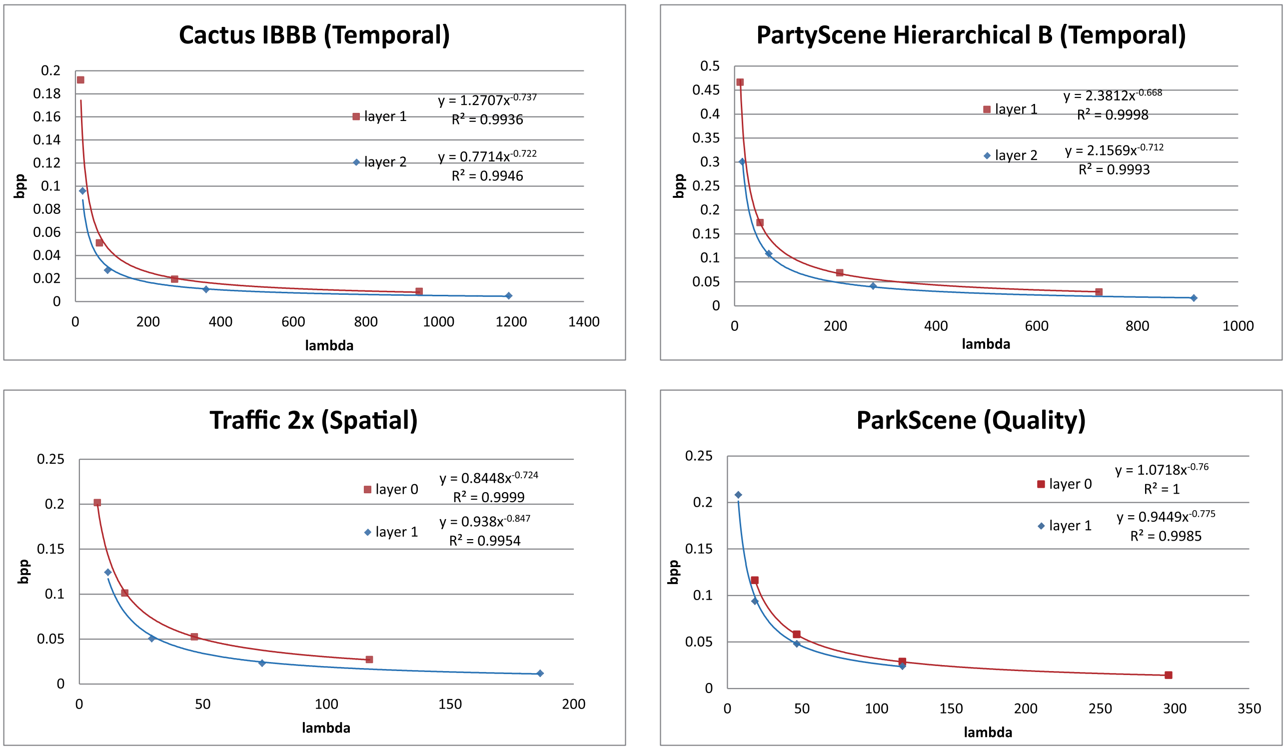


Fig. 1. Verification of hyperbolic R-λ model for HEVC scalable extension.

As we know, the hyperbolic relationship between R and λ has been demonstrated for HEVC in [12]. However, the behavior of λ in HEVC scalable extension context has never been studied. Although we still hold the opinion that λ is the key factor to determine R even for HEVC scalable extension, the relationship between R and λ becomes unknown due to some new features such as inter layer prediction brought by temporal, spatial, and quality scalability. To be able to use an R-λ model based rate control algorithm in HEVC scalable extension, we should carefully validate that the relationship between R and λ is still hyperbolic.

We test our assumptions on both temporal, spatial, and quality scalability. For temporal scalability, we first generate the HM [38] anchor with temporal scalability for both IBBB and hierarchical-B cases and then fit the R-λ model for each layer according to (4). Low delay main profile configuration [39] supporting temporal scalability is used for IBBB case and random access main profile configuration [39] supporting temporal scalability with only one intra frame at the beginning of each sequence is used for hierarchical-B case. For both spatial and quality scalability, to make it more convenient to fit the relationship between bpp and λ, we generate the SHM [40] anchor using the simple IBBB coding structure (only the first frame is intra frame, followed by a number of B frames) with only one reference frame and flat QP (the QPs of all the B frames are set as the same value, i.e., QP of the I frame plus 3) for both the base and enhancement layers. We also fit the R-λ models for both layers according to (4). The results are shown in Fig. 1. For temporal scalability, we show the results of layer 1 and layer 2, while for spatial and quality scalability, we show the results of layer 0 and layer 1. It should be noticed that R is expressed in terms of bpp in Fig. 1. We use the coefficient of determination R^2 to measure how well (4) is. The result shows that the average R^2 for each layer is greater than 0.99. Therefore, we hold that the hyperbolic R-λ model is also suitable for HEVC scalable extension.

III. Proposed Rate Control Scheme for HEVC Scalable Extension Using R-λ Model

The λ-domain rate control algorithm for HEVC has been described clearly in [13]. In this paper, a λ-domain rate control algorithm for HEVC scalable extension is proposed to keep the bitrate for each layer steady as well as achieve quite good R-D performance. The "layer" here means a temporal, spatial or quality layer. The overall rate control algorithm for temporal scalability and spatial/quality scalability will be described in Section III-A and Section III-B, respectively. Moreover, an adaptive updating algorithm for R-λ model is proposed in Section III-C to control the bits per frame even more precisely. The notations used in the following parts are explained in Table I.

A. The Rate Control Algorithm for HEVC Temporal Scalability

Generally speaking, the rate control algorithm can be divided into two parts: bit allocation and how to achieve the target bitrate. They will be explained separately in the following.

1) Bit Allocation: Before discussing the bit allocation scheme, the encoding structure of temporal scalability for both IBBB and hierarchical-B cases should be explained first. Fig. 2 shows the typical encoding structures for temporal scalability.

TABLE I
NOTATION LIST AND ITS MEANINGS

Notation	Explanation
N_S	the number of frames in a sequence
N_G	the number of frames in a GOP
$RG_{l,i}$	remained bits of a GOP before coding frame i in layer l
$R_{l,G}$	the GOP level target bits for layer l
TAP_l	average target bits for a frame in layer l
$TP_{l,i}$	target bits for frame i in layer l
$D_{l,i}$	distortion for frame i in layer l before coding the GOP
$R_{l,i}$	target bits for frame i in layer l before coding the GOP
$\lambda_{l,i}$	λ for frame i in layer l before coding the GOP
$F_{k,i}$	buffer fullness of the kth CPB before encoding frame i
R_k	sum of the target bitrate lower than layer k
B_k	buffer size of layer k
$TBpp_l$	target average bpp for a frame or BU in layer l
$RBpp_l$	actual average bpp for a frame or BU in layer l
λ_l^R	real used λ for a frame or a BU in layer l
λ_l^E	estimated λ for a frame or a BU in layer l

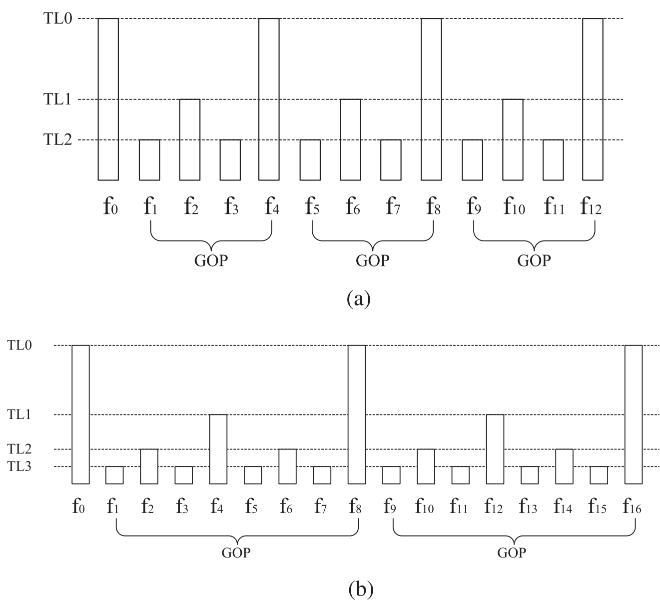


Fig. 2. GOP structure of temporal scalability. (a) GOP structure for IBBB case. (b) GOP structure for hierarchical-B case.

Fig. 2(a) is for IBBB case, and GOP size is 4, while Fig. 2(b) is for hierarchical-B case, and GOP size is 8. In Fig. 2, different heights of the rectangles mean the frames belong to different temporal layers. Higher rectangle means that the corresponding frame has smaller temporal ID (TID) and vice versa. TID is an identification of the temporal layer to which the current frame belongs.

As we know, most of the bit allocation schemes for a whole sequence include GOP level, picture level, and Basic Unit (BU, for example, one or more continuous CUs) level. Since the target bitrate of each layer is specified independently according to application, various layers of temporal scalability can be thought as independent sequences no matter in terms of bit allocation or how to achieve the target bitrate. Therefore, the GOP level, picture level and BU level bit allocation schemes for temporal scalability are quite similar to that of HEVC in [13]. The main difference is that we try to optimize the initial target bits of the first frame of temporal layer 0 since the first frame is an intra frame which may have significant influence on the overall performance of a rate control scheme. Since the optimal initial target bits may vary in a very large range, we use the initial target bits ratio (the ratio of the initial target bits relative to the average bits of temporal layer 0) to represent the initial target bits.

Generally speaking, there are mainly three parameters influencing the optimal initial target bits ratio. The first parameter is the complexity of the current frame. The complexity reflects the number of bits the current frame will consume under a certain quality. In general, the more complex a frame is, the more bits it will cost. Considering both accuracy and simplification, we propose to use the Sum of the Absolute Transformed Difference of the intra frame $SATD_I$ to measure the complexity. The $SATD_I$ is the sum of the absolute values of the coefficients $h_{i,j}$ after applying Hadamard transform to each 8×8 block minus the DC value

$$SATD_I = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |h_{i,j}| - |h_{0,0}|, \quad N = 8. \tag{5}$$

The second parameter is the influence of the current frame on the subsequent frames. The larger the influence, the more bits will be assigned to the current frame. The accurate determination of the influence is a quite complex process. Considering simplification, we propose to use the sum of the absolute transformed difference between the first frame and several subsequent frames ($SATD_B$) to measure the influence. The larger the influence, the smaller the value of $SATD_B$. To calculate $SATD_B$, a number of 8×8 residue blocks between the first frame and several subsequent frames are firstly generated using a simple motion estimation process. $SATD_B$ is then calculated using the sum of the absolute values of the coefficients $r_{i,j}$ after applying Hadamard transform to the 8×8 residue blocks

$$SATD_B = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |r_{i,j}|, \quad N = 8. \tag{6}$$

Under different coding structures, the frames used to calculate $SATD_B$ are different. For the IBBB case with GOP size 4, the average $SATD_B$ between the first frame and the second, third, and fifth frame is used to measure the influence. For the hierarchical-B case with GOP size 8, the average $SATD_B$ between the first frame and the second, third, fifth, and ninth frame is used to measure the influence. The third parameter is the $TBpp_l$ of temporal layer 0. Since the residue after quantization is quite small in low bitrate case, the prediction becomes more important for low bitrate case compared with the high bitrate case. Consequently the influence of the first frame on the subsequent frames also becomes more significant. Therefore, the smaller the target bitrate, the larger the optimal initial target bits ratio will be.

In summary, the optimal initial target bits ratio can be calculated through

$$bitsRatio = a \cdot SATD_I^b \cdot SATD_B^c \cdot TBpp_l^d \tag{7}$$

where a, b, c, and d are the model parameters. According to our analysis above, a and b are larger than 0, c and d are less than 0. After fitting the formula with 8 sequences from Class B, C, D, E (two for each class), we can obtain two groups of suitable values of a, b, c, and d for IBBB and hierarchical-B cases, respectively. The values of a, b, c, and d are set as 0.34, 1.40, -1.11, -0.54 in IBBB case. In hierarchical-B case, the values of a, b, c, and d are set as 1.32, 1.07, -1.52, -0.41 accordingly.

Besides the bit allocation for the first frame, the scene change problem is also quite important for temporal scalability especially in IBBB case due to the fact that the frames with smaller TID cannot reference the frames with larger TID. As shown in Fig. 2(a), all the frames f_{4n+1}, f_{4n+2}, and f_{4n+4} can only reference the frame f_{4n}. Therefore, if the scene change happens to appear in frame f_{4n+1}, besides f_{4n+1}, f_{4n+2} and f_{4n+4} should be also handled as scene change frames since the frames f_{4n+2} and f_{4n+4} are unable to reference any frames after scene change. Without special treatment, the continuous three scene change frames may lead to some serious problems such as R-D performance degradation and buffer overflow since the bits of these frames are out of control.

To deal with the scene change of temporal scalability, we should first detect the scene change. Therefore, a simple scene change detection method is firstly proposed. Our scene change detection method mainly has two constraints. Firstly, for the kth frame, the MAD_k between the current frame and its nearest reference frame must be larger than a threshold T_{k-2} multiplied by 2. The initial value of the threshold T_1 is set as the MAD_1 between the second frame and the first frame. Then it is updated along with the encoding process using the following formula:

$$T_k = \frac{T_k \times (k-1) + MAD_k}{k}. \tag{8}$$

Secondly, the MAD_k between the current frame and its nearest reference frame must be larger than a predefined fixed value T to avoid the case of sudden change from static to large motion.

After the scene change is detected, the scene change frame will be handled as an intra frame. Since the scene change frames in different temporal layers have totally different influences on the subsequent frames, they should be assigned different amount of bits. If the scene change frame is in temporal layer 0, it will have significant influence on the subsequent frames thus it will be handled similarly as the first frame. However, if the scene change frame is in other temporal layers, the influence of such kind of scene change frames to the subsequent frames will be much less. For example, as shown in Fig. 2(a), the frames f_{4n+1} and f_{4n+2} can only influence f_{4n+3} and the influence cannot propagate to the subsequent frames due to the existence of f_{4n+4}. Therefore, much less bits will be assigned to the scene change frames in other temporal layers. In our implementation, the less important scene change frames are assigned $\frac{1}{6}$ of the target bits of the first frame.

After the bit allocation process, the assigned bits should follow some constraints brought by the hypothesis reference decoder (HRD). We extend the simple buffer control algorithm in HM16.7 to support temporal scalability to meet the HRD constraints [41]. As SVC can provide the property that a subset bitstream can form another valid bitstream, all the subset bitstreams should follow the HRD constraints. When determining the target bits of a specified frame, the following two constraints, which try to make the occupancy of a coded picture buffer (CPB) stay between 10% and 90% of the CPB size, will be followed to avoid buffer overflow and underflow for each subset bitstream:

$$F_{k,i} - TP_{l,i} + \frac{R_k}{f} < B_k \times 0.9, \quad k \geq l \tag{9}$$

$$F_{k,i} - TP_{l,i} > B_k \times 0.1, \quad k \geq l \tag{10}$$

where f is the frame rate. The formulas imply that the current frame should follow the buffer constraint of not only the current layer but also the layers higher than the current layer. The target bits assigned by the bit allocation algorithm can meet the requirements of formula (9) and (10) in most cases. However, if the target bits make formula (9) invalid, the target bits will be regulated as

$$TP_{l,i} = F_{k,i} + \frac{R_k}{f} - B_k \times 0.9, \quad k \geq l. \tag{11}$$

In addition, if the target bits make formula (10) invalid, the target bits will be regulated as

$$TP_{l,i} = max(200, F_{k,i} - B_k \times 0.1), \quad k \geq l. \tag{12}$$

2) How to Achieve the Target Bitrate: To achieve the target bitrate, λ determination is the most important part. After the assigned number of bits to a picture or a BU is determined, λ can be computed through the following equation:

$$\lambda_l^R = \alpha_l \times TBpp_l^{\beta_l}. \tag{13}$$

For the first frame, the accurate determination of the model parameters α_l and β_l is of great importance since the first frame has significant influence on the overall R-D performance. To find the accurate model parameters for the first frame, we encode the first frame of several test sequences using QP equal to 22, 27, 32, and 37. The corresponding λ and bpp are collected to fit the R-λ model to find the model parameters. The model parameters of some typical sequences with different resolutions and characteristics are shown in Table II. From Table II, we can obviously see that the values of α_l change significantly from one sequence to another while the values of β_l are all very close to -2.0. Therefore, we fix the approximated value of β_l as -2.0 in our scheme. After determining an approximated value of β_l,

TABLE II
TYPICAL MODEL PARAMETERS OF INTRA FRAME

Sequences	α_l	β_l	$\alpha_{l,adjust}$	$SATD_I$
Traffic	3.91	–2.04	4.06	10.53
PeopleOnStreet	4.03	–2.10	4.44	13.29
Cactus	5.85	–1.76	4.70	11.48
BQTerrace	11.52	–1.89	10.87	17.26
BasketballDrill	5.85	–1.87	5.21	12.37
PartyScene	50.16	–2.23	48.53	26.12
BasketballPass	6.99	–2.01	7.05	12.28
BlowingBubbles	21.43	–1.97	21.34	18.14
Johnny	0.62	–2.07	0.72	7.37

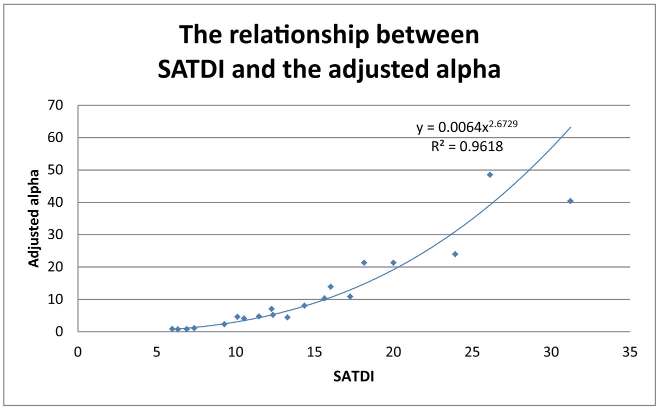


Fig. 3. Relationship between $\alpha_{l,adjust}$ and $SATD_I$.

α_l should be adjusted using the following formula to make the new model parameters approximate the original model,

$$\alpha_{l,adjust} = \frac{\lambda}{TBpp_l^{-2.0}}. \tag{14}$$

Different combinations of λ and $TBpp_l$ for different QP values (22, 27, 32, and 37) may result in different $\alpha_{l,adjust}$. We choose the average $\alpha_{l,adjust}$ of the four points as the final $\alpha_{l,adjust}$. The detailed values of $\alpha_{l,adjust}$ of the typical sequences are shown in the fourth column of Table II. After careful observation and analysis, we find that the $\alpha_{l,adjust}$ is actually related to the $SATD_I$ of the intra frame. The values of the $SATD_I$ of the typical sequences are shown in the last column of Table II. Then we try to find the relationship between α_l and $SATD_I$ through fitting a curve for all the sequences under the HEVC common test condition. The fitted curve is shown in Fig. 3. From the curve, we can see that the coefficient of determination R^2 is as high as 0.96, which means that the curve fits all the test sequences quite well. Therefore, the relationship between the $\alpha_{l,adjust}$ and $SATD_I$ is determined as

$$\alpha_{l,adjust} = 0.0064 \cdot SATD_I^{2.6729}. \tag{15}$$

In summary, the λ_l^R of the first frame can be calculated as

$$\lambda_l^R = (0.0064 \cdot SATD_I^{2.6729}) \cdot TBpp_l^{-2.0}. \tag{16}$$

For the other frames except for the first frame, the accurate determination of initial model parameters α_l and β_l is not so important. Therefore, α_l and β_l are initialized as 3.2003 and –1.367 for various sequences, and will adapt to the specified video content through the following updating scheme [12]:

$$\lambda_l^E = \alpha_l^{old} \times RBpp_l^{\beta_l^{old}} \tag{17}$$

$$\alpha_l^{new} = \alpha_l^{old} + \delta_{\alpha_l} \times (\ln(\lambda_l^E) - \ln(\lambda_l^R)) \times \alpha_l^{old} \tag{18}$$

$$\beta_l^{new} = \beta_l^{old} + \delta_{\beta_l} \times (\ln(\lambda_l^E) - \ln(\lambda_l^R)) \times \ln(RBpp_l) \tag{19}$$

where δ_{α_l} and δ_{β_l} are set as 0.1 and 0.05 respectively in the rate control algorithm for HEVC. However, as mentioned in Section I, in the rate control scheme for a specified temporal layer, if δ_{α_l} and δ_{β_l} are still fixed at 0.1 and 0.05, it may cause some serious problems. The detailed settings of δ_{α_l} and δ_{β_l} will be discussed in Section III-C in details. After λ is determined, QP will be determined by the following equation [42]:

$$QP = c_1 \times \ln(\lambda) + c_2 \tag{20}$$

where c_1 and c_2 are set as 4.2005 and 13.7122, respectively.

B. The Rate Control Algorithm for HEVC Spatial and Quality Scalability

In the case of spatial and quality scalability, HEVC scalable extension follows the conventional approach of multi-layer coding. Each layer consists of a GOP structure similar to that shown in Fig. 2. Since the target bitrate of each layer is specified independently according to application, various layers of spatial and quality scalability can be thought as independent sequences no matter in terms of bit allocation or how to achieve the target bitrate. Therefore, the basic concepts of the GOP level, picture level, and BU level rate control schemes are quite similar to that of HEVC. Similar to temporal scalability, the introduction of the rate control algorithm for HEVC spatial and quality scalability will be also divided into two parts: bit allocation and how to achieve the target bitrate.

1) Bit Allocation: For the first frame of the base layer, we estimate the optimal initial target bits ratio in a similar way as the first frame of temporal layer 0 according to (7). It should be noted that the $bitsRatio$ means the ratio of the optimal initial target bits to the average bits of the base layer. Besides, since the encoding structure is different from the case in temporal scalability, the model parameters will become different. To determine the optimal initial target bits ratio of the base layer, we fix the initial QP of the enhancement layer and traverse all the QPs of the base layer to find the optimal initial target bits ratio. After fitting (7) using several Class B sequences in IBBB 2x and hierarchical-B 2x cases, we can obtain several suitable model parameters for IBBB/IPPP and hierarchical-B cases. For both IBBB and IPPP cases, the values of the model parameters a, b, c, and d are set as 0.13, 2.06, -0.92, and -0.33, respectively. In hierarchical-B cases, the values of the model parameters are set as 0.12, 3.86, -3.62, and -0.39 accordingly.

The optimal initial target bits ratio determination of the first frame of the enhancement layer is also according to (7). The main difference is that the $SATD_I$ is replaced by $SATD_P$ in (7), since the first frame of the enhancement layer is a P frame which can take the corresponding base layer frame as reference. To calculate $SATD_P$, we will first generate several 8×8 residue blocks through subtracting the reconstructed base layer block from the enhancement layer block (for spatial scalability, up-sampling of the base layer frame is needed, while for quality scalability, the up-sampling operation is not needed). Then $SATD_P$ is calculated through the sum of the absolute values of the coefficients after applying Hadamard transform to each 8×8 residue block. Similar to the base layer, to determine the optimal initial target bits ratio of the enhancement layer, we fix the initial QP of the base layer and traverse all the QPs of the enhancement layer to find the optimal initial target bits ratio. After fitting (7) using several Class B sequences in IBBB 2x and hierarchical-B 2x cases, we can obtain several suitable model parameters for IBBB/IPPP and hierarchical-B cases. For both IBBB and IPPP cases, the values of the model parameters a, b, c, and d are set as 0.02, 2.94, -0.47, and -0.33, respectively. In hierarchical-B cases, the values of the model parameters are set as 0.32, 2.96, -1.82, and -0.30 accordingly.

Besides the determination of optimal initial target bits ratio, due to the existence of the intra and inter layer dependency, the optimization of bit allocation of the inter frames is also very important. In this paper, we use an adaptive ratio bit allocation algorithm in picture level to improve the overall R-D performance taking both the intra and inter layer dependency into consideration. The target bits of each frame for a specified layer can be computed using the following formula:

$$TP_{l,i} = \Omega_{l,i} \times RG_{l,i}. \tag{21}$$

The adaptive weighting factors of picture level bit allocation $\Omega_{l,i}$s are optimized through fundamental RDO theory.

The picture level bit allocation aims to minimize the weighted sum of the distortion of the base layer and enhancement layer under the constraint of the GOP level target bits for both the base layer and enhancement layer. Therefore, the picture level bit allocation problem can be formulated as

$$\begin{aligned} &\min_{\lambda_{l,i}} \sum_{i=1}^{N_S} (\omega_0 D_{0,i}(\lambda_{l,i}) + \omega_1 D_{1,i}(\lambda_{l,i})) \\ &s.t. \sum_{i=1}^{N_G} R_{0,i}(\lambda_{l,i}) \le R_{0,G}, \quad \sum_{i=1}^{N_G} R_{1,i}(\lambda_{l,i}) \le R_{1,G}. \end{aligned} \tag{22}$$

It should be noted that both $D_{l,i}$ and $R_{l,i}$ are related to $\lambda_{l,i}$. However, in the following, to make the formulas concise, we ignore the $\lambda_{l,i}$ in the cost function. The Lagrange multiplier method can convert the constrained optimization problem in (22) to an equivalent unconstraint problem by introducing the Lagrangian cost function as

$$\min_{\lambda_{l,i}} \sum_{i=1}^{N_S} (\omega_0 D_{0,i} + \omega_1 D_{1,i}) + \lambda \sum_{i=1}^{N_G} R_{0,i} + \mu \sum_{i=1}^{N_G} R_{1,i} \tag{23}$$

where λ and μ are the Lagrangian multipliers. According to the Lagrangian method, we can obtain a formula as follows:

$$\frac{\partial \sum_{i=1}^{N_S} (\omega_0 D_{0,i} + \omega_1 D_{1,i})}{\partial \lambda_{l,j}} + \lambda \frac{\partial \sum_{i=1}^{N_G} R_{0,i}}{\partial \lambda_{l,j}} + \mu \frac{\partial \sum_{i=1}^{N_G} R_{1,i}}{\partial \lambda_{l,j}} = 0. \tag{24}$$

If we consider the base layer and enhancement layer separately, (24) can be converted into the following two formulas:

$$\frac{\partial \sum_{i=1}^{N_S} (\omega_0 D_{0,i} + \omega_1 D_{1,i})}{\partial \lambda_{0,j}} + \lambda \frac{\partial \sum_{i=1}^{N_G} R_{0,i}}{\partial \lambda_{0,j}} = 0 \tag{25}$$

$$\frac{\partial \sum_{i=1}^{N_S} (\omega_0 D_{0,i} + \omega_1 D_{1,i})}{\partial \lambda_{1,j}} + \mu \frac{\partial \sum_{i=1}^{N_G} R_{1,i}}{\partial \lambda_{1,j}} = 0. \tag{26}$$

For the base layer frame, since the frames decoded after the current frame in both the base layer and enhancement layer may take the current frame as reference, the quality of the current frame may have significant influence on the subsequent frames in both the base layer and enhancement layer

$$\frac{\partial D_{0,i}}{\partial D_{0,j}} \ge 0, \quad \frac{\partial D_{1,i}}{\partial D_{0,j}} \ge 0, \quad \text{if } i \ge j. \tag{27}$$

Therefore, (25) can be rewritten as

$$\omega_0 \cdot \frac{\partial \sum_{i=j}^{N_S} D_{0,i}}{\partial D_{0,j}} \cdot \frac{\partial D_{0,j}}{\partial \lambda_{0,j}} + \omega_1 \cdot \frac{\partial \sum_{i=j}^{N_S} D_{1,i}}{\partial D_{0,j}} \cdot \frac{\partial D_{0,j}}{\partial \lambda_{0,j}} + \lambda \frac{\partial R_{0,j}}{\partial \lambda_{0,j}} = 0. \tag{28}$$

Then the $\lambda_{0,j}$ can be solved as

$$\lambda_{0,j} = \frac{\lambda}{\omega_0 \cdot \frac{\partial \sum_{i=j}^{N_S} D_{0,i}}{\partial D_{0,j}} + \omega_1 \cdot \frac{\partial \sum_{i=j}^{N_S} D_{1,i}}{\partial D_{0,j}}} = \lambda \cdot \omega_{0,j}. \tag{29}$$

It should be emphasized that the $\frac{\partial \sum_{i=j}^{N_S} D_{0,i}}{\partial D_{0,j}}$ and $\frac{\partial \sum_{i=j}^{N_S} D_{1,i}}{\partial D_{0,j}}$ mean the influence of the current frame on the subsequent frames in the base layer and the enhancement layer, respectively. Therefore, the $\omega_{0,j}$ is inversely proportional to the weighted influence of the current frame on its subsequent frames for both the base layer and enhancement layer. The larger the influence of the current frame on the subsequent frames, the smaller the value of $\omega_{0,j}$ will be. The smaller value of $\omega_{0,j}$ will lead to smaller $\lambda_{0,j}$ and higher quality for frame j in base layer. The more important frames allocating more bits is good for the quality of the whole sequence. Except for the form in (29), the relationship of the λ between different frames for the base layer can also be expressed as

$$\frac{\lambda_{0,i}}{\lambda_{0,j}} = \frac{\omega_0 \cdot \frac{\partial \sum_{k=j}^{N_S} D_{0,k}}{\partial D_{0,j}} + \omega_1 \cdot \frac{\partial \sum_{k=j}^{N_S} D_{1,k}}{\partial D_{0,j}}}{\omega_0 \cdot \frac{\partial \sum_{k=i}^{N_S} D_{0,k}}{\partial D_{0,i}} + \omega_1 \cdot \frac{\partial \sum_{k=i}^{N_S} D_{1,k}}{\partial D_{0,i}}} = \frac{\omega_{0,i}}{\omega_{0,j}}. \tag{30}$$

Equation (30) obviously shows that the ratio of λ between different frames should be inversely proportional to the weighted sum of their influence to the base layer and enhancement layer.

For the enhancement layer, only the frames decoded after the current frame in the enhancement layer will take the current frame as reference

$$\frac{\partial D_{0,i}}{\partial D_{1,j}} = 0, \quad \frac{\partial D_{1,i}}{\partial D_{1,j}} \ge 0, \quad \text{if } i \ge j. \tag{31}$$

So (26) can be rewritten as

$$\omega_1 \cdot \frac{\partial \sum_{i=j}^{N_S} D_{1,i}}{\partial D_{1,j}} \cdot \frac{\partial D_{1,j}}{\partial \lambda_{1,j}} + \mu \frac{\partial R_{1,j}}{\partial \lambda_{1,j}} = 0. \tag{32}$$

Then the $\lambda_{1,j}$ can be solved as

$$\lambda_{1,j} = \frac{\mu}{\omega_1 \cdot \frac{\partial \sum_{i=j}^{N_S} D_{1,i}}{\partial D_{1,j}}} = \mu \cdot \omega_{1,j}. \tag{33}$$

The conclusion also reflects that more important frames should be assigned more bits and are able to contribute more to the quality of the whole sequence. Similar to the base layer, (33) can also be rewritten as the following form:

$$\frac{\lambda_{1,i}}{\lambda_{1,j}} = \frac{\frac{\partial \sum_{k=j}^{N_S} D_{1,k}}{\partial D_{1,j}}}{\frac{\partial \sum_{k=i}^{N_S} D_{1,k}}{\partial D_{1,i}}} = \frac{\omega_{1,i}}{\omega_{1,j}}. \tag{34}$$

TABLE III
DETAILED $\omega_{0,i}$ SETTING

hierarchical-B case			IBBB/IPPP case		
$\omega_{0,i}$	$\lambda_{0,L} < 90$	$\lambda_{0,L} \geq 90$	$\omega_{0,i}$	$\lambda_{0,L} < 120$	$\lambda_{0,L} \geq 120$
f_{8n}	1.0	1.0	f_{4n}	1.0	1.0
f_{8n+4}	$0.725\ln(\lambda_{0,L}) + 0.7963$	4.0	f_{4n+2}	$0.725\ln(\lambda_{0,L}) + 0.5793$	4.0
f_{4n+2}	$0.943\ln(\lambda_{0,L}) + 1.0352$	5.0	f_{2n+1}	$0.943\ln(\lambda_{0,L}) + 0.7531$	5.0
f_{2n+1}	$2.356\ln(\lambda_{0,L}) + 2.5880$	12.3	–	–	–

TABLE IV
DETAILED $\omega_{1,i}$ SETTING

hierarchical-B case			IBBB/IPPP case		
$\omega_{1,i}$	$\lambda_{1,L} < 90$	$\lambda_{1,L} \geq 90$	$\omega_{1,i}$	$\lambda_{1,L} < 120$	$\lambda_{1,L} \geq 120$
f_{8n}	1.0 * 1.1	1.0 * 1.1	f_{4n}	1.0 * 1.1	1.0 * 1.1
f_{8n+4}	$0.725\ln(\lambda_{1,L}) + 0.7963$	4.0	f_{4n+2}	$0.725\ln(\lambda_{1,L}) + 0.5793$	4.0
f_{4n+2}	$0.943\ln(\lambda_{1,L}) + 1.0352$	5.0	f_{2n+1}	$0.943\ln(\lambda_{1,L}) + 0.7531$	5.0
f_{2n+1}	$(2.356\ln(\lambda_{1,L}) + 2.5880) * \beta$	$12.3 * \beta$	–	–	–

Equation (34) obviously shows that the ratio of λ between different frames should be inversely proportional to their influence to the enhancement layer.

After determining the ratio of $\lambda_{l,i}$ between different frames in a GOP for both the base layer and enhancement layer, we can use the GOP level target bits as the constraint to solve the $\lambda_{l,i}$ for each frame

$$\sum_{i=1}^{N_G} R_{l,i} = R_{l,G}. \tag{35}$$

If we combine (4) and (35), for both the base layer and enhancement layer, we will have

$$\sum_{i=1}^{N_G} \alpha_{l,i}{\lambda_{l,i}}^{\beta_{l,i}} = R_{l,G}. \tag{36}$$

Through (36), (29), and (33), we can obtain the values of λ and μ. So are the values of $\lambda_{l,i}$ and $R_{l,i}$ for each frame. It should be noted that $\alpha_{l,i}$ and $\beta_{l,i}$ are parameters related to the video content. Therefore, the proposed algorithm can assign target bits to each frame according to the characteristic of the video content and thus can achieve quite good R-D performance.

As the rate control algorithm cannot guarantee that the actual bits cost by each frame are equal to its target bits without even one bit difference, the calculated $R_{l,i}$ is unable to be used as the target bits of each frame directly. The actual adaptive ratio $\Omega_{l,i}$ can be calculated as follows:

$$\Omega_{l,i} = \frac{R_{l,i}}{\sum_{j=i}^{N_G} R_{l,j}}. \tag{37}$$

After the above process, to calculate the target bits for each frame, how to measure the intra and inter layer dependency becomes the only problem. In terms of the detailed setting of $\omega_{0,i}$ and $\omega_{1,i}$, only the case in which the ω_0 and ω_1 are equal is considered. In hierarchical-B case, the method proposed in [43] which characterizes the dependency in terms of the better average performance is used in this paper. In IBBB/IPPP case, we set $\omega_{0,i}$ and $\omega_{1,i}$ in a similar way according to our experience. The detailed settings of $\omega_{0,i}$ and $\omega_{1,i}$ are shown in Tables III and IV, respectively. The $\lambda_{l,L}$ in the tables means the λ of the last frame in display order in the last GOP, and β is calculated as

$$\beta = \frac{height_1 \times width_1}{height_0 \times width_0} \times 2^{(QP_0 - QP_1)/6} \tag{38}$$

where $height_1$ and $width_1$ are the height and width for the enhancement layer frames, $height_0$ and $width_0$ are the height and width for the base layer frames, QP_1 and QP_0 are the initial QPs for the enhancement layer and base layer, respectively.

After the bit allocation process, the target bits should follow some constraints brought by the HRD. The HRD constraints used for spatial/quality scalability are the same as temporal scalability shown in formula (9) and (10).

2) How to Achieve the Target Bitrate: The $\lambda_{l,i}$ and QP are determined according to (13) and (20) to achieve the target bits. For the frames except for the first frame, α_l and β_l are initialized as 3.2003 and –1.367, and the R-λ model is updated using (17)–(19). Since the encoding structure becomes more flexible for HEVC spatial and quality scalability compared with HEVC, the values of δ_{α_l} and δ_{β_l} fixing at 0.1 and 0.05 are unsuitable for updating the R-λ model. The detailed settings of δ_{α_l} and δ_{β_l} will be discussed in Section III-C in details.

For the first frame of the base layer which is an intra frame, the same formula as temporal scalability, i.e., formula (16), is used to determine $\lambda_{l,R}$. For the first frame of the enhancement layer, we use a similar formula as (16) to calculate λ_l^R. Since the first frame of the enhancement layer is a P frame, we replace the $SATD_I$ using $SATD_P$ and fit the model parameters in a similar way as the temporal scalability does. The final formula used to calculate the λ_l^R of the first frame of the enhancement

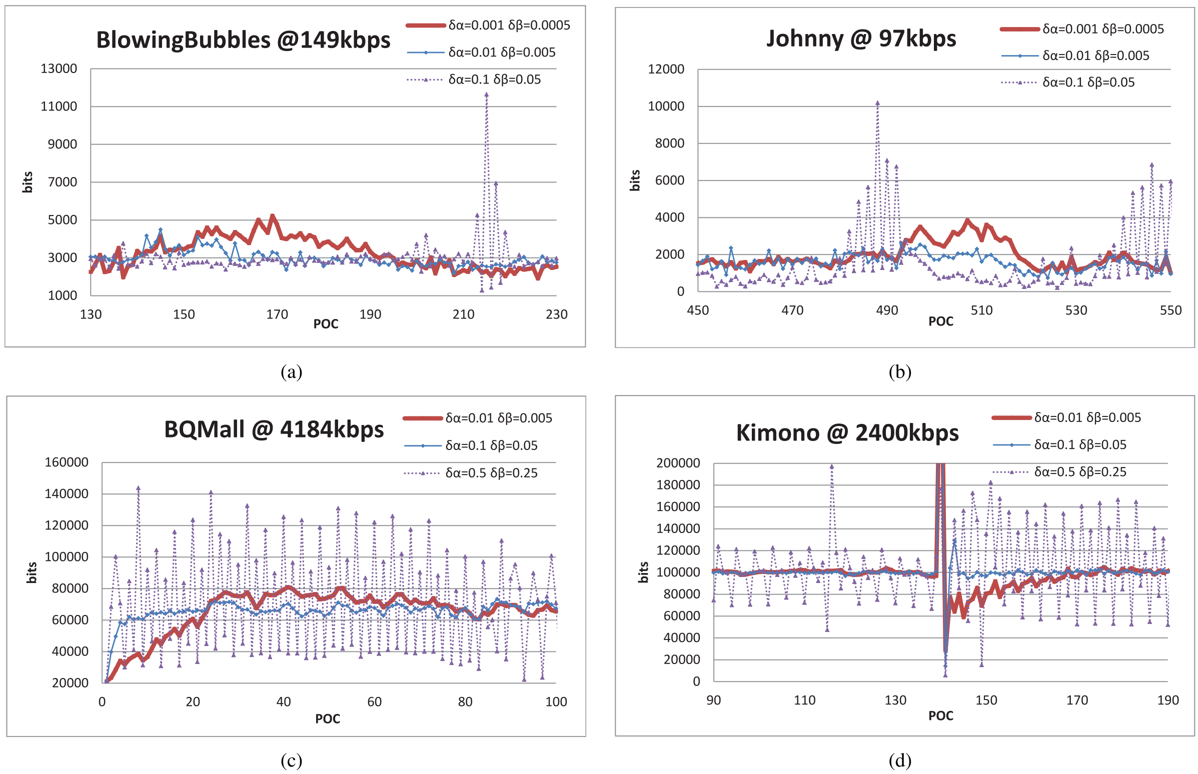


Fig. 4. Bits per frame using different updating parameters.

layer is shown as follows:

$$\lambda_l^R = (0.0482 \cdot SATD_P^{2.4345}) \cdot TBpp_l^{-1.0}. \tag{39}$$

C. Proposed Adaptive Updating Algorithm for R-λ Model

The whole rate control scheme for HEVC scalable extension has been described clearly in the former sections except how to determine the updating parameters δ_{α_l} and δ_{β_l} in (18) and (19). In this section, the determination of the updating parameters is discussed in details.

To find out the reasonable δ_{α_l} and δ_{β_l}, the experiments using λ-domain rate control algorithm for HEVC with equal bit allocation for each frame under IBBB coding structure are conducted. Different values of δ_{α_l} and δ_{β_l} are used in the experiments. The results are shown in Fig. 4 where POC is short for picture order count.

On one hand, if δ_{α_l} and δ_{β_l} are too large, when the characteristic of the video content happens to change a little, the large δ_{α_l} and δ_{β_l} will lead to the significant change of α_l and β_l. Then according to (13), the λ_l^R between consecutive frames will vary significantly along with the significant change of α_l and β_l. Finally, the serious change of λ_l^R may result in significant bits variation between consecutive frames. The dotted lines in Fig. 4(a) and (b) show such a situation. After the careful observation of the sequence BlowingBubbles [see Fig. 4(a)], we find that the motion between consecutive frames around the 210th frame is smaller compared with the other frames. Therefore, the updating method is trying to decrease λ_l^R to accommodate the new video content and keep the bits per frame steady. However, if the updating speed is too fast, λ_l^R will become so small that the bits cost by the current frame will be too much. The situation is similar around the 490th frame for the sequence Johnny [see Fig. 4(b)]. In another case, as the initialized values of α_l and β_l are determined without preprocessing, the model may be unable to adapt to the video content at first. Due to the fast updating speed, α_l and β_l will keep varying significantly. Therefore, the model will be unable to accommodate the specified video content for a long time. The dotted lines in Fig. 4(c) and (d) show such a case. Fig. 4(c) shows the case at the start of the sequence. Fig. 4(d) shows the situation when scene change happens around the 141th frame.

On the other hand, if δ_{α_l} and δ_{β_l} are too small, when a new object happens to appear in the scene, the characteristic of the video content will change a little, the model will then be unable to catch up with the change of the video content due to the slow updating speed. Such a situation can be seen from the thick solid lines in Fig. 4(a) and (b). For the sequence BlowingBubbles [see Fig. 4(a)], there happens to be a bubble appearing in the top-right of the video content at around the 150th frame. While for Johnny [see Fig. 4(b)], the right hand of the man happens to appear at around the 500th frame. In another case, as the initialized values of α_l and β_l are determined without preprocessing, they may be unsuitable for the video content at first. The convergence speed of α_l and β_l to adapt to the specified video content will be quite slow due to the slow updating speed. Such a situation can be seen from the thick solid lines in Fig. 4(c) and (d). For the sequence BQMall [see Fig. 4(c)], with the slow updating speed, α_l and β_l can only change a little after encoding a frame. Although the trend of changing α_l and β_l is right, it still costs

TABLE V
DETAILED SETTINGS OF δ_{α_l} AND δ_{β_l}

$TBpp_l$	δ_{α_l}	δ_{β_l}
$0.08 \leq TBpp_l$	0.1	0.05
$0.03 \leq TBpp_l < 0.08$	0.05	0.025
$TBpp_l < 0.03$	0.01	0.005

about 30 frames to adapt to the specified video content. For the sequence Kimono [see Fig. 4(d)], the situation is similar when scene change happens.

The situations with suitable values of δ_{α_l} and δ_{β_l} can be seen from the solid lines in Fig. 4. When the values of δ_{α_l} and δ_{β_l} are appropriate, the bits variation per frame is limited in a very narrow range and the convergence speed is within 5 frames. It should be emphasized that the model can quickly accommodate the specified video content again when scene change happens.

The previous part introduces the influence of the updating parameters in the rate control algorithm for a whole sequence with equal bit allocation. Since these problems may become more serious for the rate control scheme for HEVC scalable extension as pointed out in Section I, the suitable values of δ_{α_l} and δ_{β_l} should be carefully investigated. As can be seen from Fig. 4, the suitable values of δ_{α_l} and δ_{β_l} are quite different for different bitrates. δ_{α_l} equal to 0.01 and δ_{β_l} equal to 0.005 are more suitable for very low bitrate, while δ_{α_l} equal to 0.1 and δ_{β_l} equal to 0.05 are more suitable for very high bitrate. Therefore, δ_{α_l} and δ_{β_l} should be specified for each layer according to the $TBpp_l$. When $TBpp_l$ is large, δ_{α_l} and δ_{β_l} should be large and vice versa. In the following experiments, the detailed settings of δ_{α_l} and δ_{β_l} are shown in Table V. The threshold is selected according to a number of experiments using different values of δ_{α_l} and δ_{β_l}.

IV. EXPERIMENTAL RESULTS

Extensive experiments are conducted to verify the performance of the proposed rate control algorithm for HEVC scalable extension. The experiments can be divided into two parts. The first part aims to prove that the proposed rate control algorithm for temporal scalability can control the bitrate of each temporal layer very precisely and can achieve quite good R-D performance. The second part targets at showing the bitrate accuracy and R-D performance of the proposed rate control algorithm for spatial and quality scalability.

A. The Performance of the Proposed Rate Control Algorithm for Temporal Scalability

The proposed λ-domain rate control algorithm for temporal scalability is implemented in HM-8.0 to compare with HM-8.0 without rate control. The Q-domain rate control algorithm in [36] is also extended to support temporal scalability for comparison. The rate control algorithm in [36] is extended that each temporal layer is treated as an independent sequence and each temporal layer is with an independent R-Q model. The target bitrate of each temporal layer is specified by the users. It should be noted that the original Q-domain rate control algorithm does not apply R-Q model to the non-reference frames. However, for temporal scalability, if no R-Q model is used for the temporal layer with only non-reference frames, the rate control will be too coarse for the temporal layer. Therefore, we also apply an independent R-Q model for the temporal layer with only non-reference frames.

We adopt low delay main profile configuration [39] supporting temporal scalability for IBBB case and random access main profile configuration [39] supporting temporal scalability with only one intra frame at the beginning of each sequence for hierarchical-B case. To verify the bitrate accuracy of the proposed rate control algorithm for temporal scalability, HM-8.0 anchor with temporal scalability is firstly generated and the bits for each temporal layer are counted. The counted bits are then rounded and used as the target bits for each temporal layer of both Q-domain and λ-domain rate control algorithms. All the test sequences specified by the common test condition [39] are used in the experiment. The scene change detection threshold T is set as 30.

Because of the limited space, just the results of the bitrate accuracy and YPSNR of each temporal layer for the sequence Cactus in IBBB case are shown in Table VI. It should be noted that the bitrate for each temporal layer in the table only contains the bitrate for the current temporal layer, the bitrates of the temporal layers with smaller TID are not included. The YPSNR of each temporal layer is also only the average of the frames belonging to the current temporal layer. From the table, we can see that the proposed λ-domain rate control algorithm for temporal scalability outperforms the Q-domain rate control algorithm no matter in bitrate error or YPSNR.

The average bitrate errors for all the tested sequences of HEVC common test condition for both IBBB and hierarchical-B cases are shown in Table VII. For IBBB and hierarchical-B cases, the average bitrate errors of the Q-domain rate control algorithm are about 1.12% and 1.61%, while the average bitrate errors of the λ-domain rate control algorithm are only about 0.18% and 0.63%, accordingly. The proposed λ-domain rate control algorithm outperforms the Q-domain rate control algorithm obviously from this aspect.

Besides the small bitrate error, keeping the bits per frame of each temporal layer steady is also very important. Fig. 5 shows a typical example of bits per frame for each temporal layer in IBBB case. We can see very regular hierarchical structure when the λ-domain rate control algorithm is used while the bits per frame under the Q-domain rate control algorithm is irregular. The experimental results demonstrate that the proposed λ-domain rate control algorithm can achieve much steadier bits per temporal layer compared with the Q-domain rate control algorithm.

The accurate bitrate control is just part of the tasks of rate control, the R-D performance is another important factor to be considered. Both the HM-8.0 anchor without rate control and the Q-domain rate control algorithm are used to show the benefits of the proposed λ-domain rate control algorithm. The comparisons between the proposed λ-domain rate control algorithm with the Q-domain rate control algorithm and HM-8.0 anchor

TABLE VI
BITRATE ACCURACY FOR EACH TEMPORAL LAYER FOR IBBB CASE (BITRATE IN KBPS, PSNR IN DB)

	TID	target bitrate	R-Q model bitrate	R-Q model bitrate error	R-Q model YPSNR	R-λ model bitrate	R-λ model bitrate error	R-λ model YPSNR	delta YPSNR
Cactus	0	12 838	12796.84	0.32%	39.61	12832.68	0.04%	39.71	0.10
	1	4973	4906.49	1.34%	38.22	4974.44	0.03%	38.54	0.32
	2	4966	4967.29	0.03%	37.82	4966.93	0.02%	38.15	0.33
	0	3840	3841.40	0.04%	37.23	3841.15	0.03%	37.39	0.16
	1	1314	1314.65	0.05%	36.08	1314.66	0.05%	36.65	0.57
	2	1404	1404.68	0.05%	35.77	1404.32	0.02%	36.48	0.71
	0	1868	1876.45	0.45%	35.22	1868.92	0.05%	35.49	0.27
	1	503	503.12	0.02%	33.90	503.21	0.04%	34.59	0.69
	2	546	545.91	0.02%	33.63	546.08	0.02%	34.47	0.84
	0	939	939.22	0.02%	32.88	939.55	0.06%	33.41	0.53
	1	228	227.55	0.20%	31.59	228.06	0.03%	32.47	0.88
	2	258	258.21	0.08%	31.42	258.04	0.01%	32.37	0.95

TABLE VII
AVERAGE BITRATE ERROR FOR ALL SEQUENCES FOR TEMPORAL SCALABILITY

TID	R-Q model IBBB	R-Q model hierarchical-B	R-λ model IBBB	R-λ model hierarchical-B
0	1.48%	2.91%	0.22%	0.72%
1	1.03%	1.90%	0.24%	1.06%
2	0.83%	0.54%	0.08%	0.31%
3	–	1.09%	–	0.43%
average	1.12%	1.61%	0.18%	0.63%

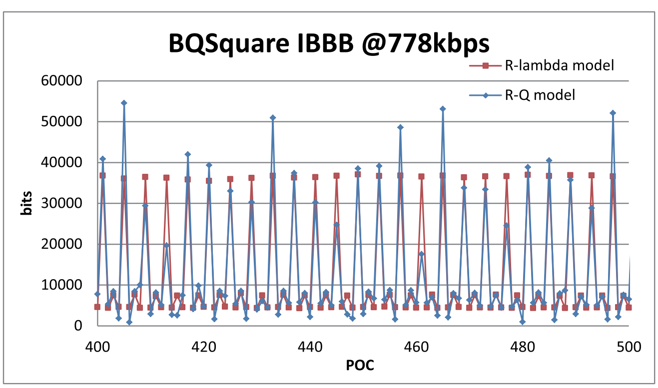


Fig. 5. Bits per frame for temporal scalability.

TABLE VIII
OVERALL R-D PERFORMANCE COMPARED WITH THE Q-DOMAIN RATE CONTROL ALGORITHM

	layer 1 IBBB	layer 1 hierarchical-B	highest layer IBBB	highest layer hierarchical-B
ClassA	–	–31.5%	–	–34.7%
ClassB	–13.4%	–22.9%	–18.6%	–25.3%
ClassC	–13.8%	–21.7%	–13.8%	–17.7%
ClassD	–7.7%	–11.9%	–8.0%	–11.0%
ClassE	–22.5%	–	–23.0%	–
ave	–13.8%	–22.1%	–15.6%	–22.4%

TABLE IX
OVERALL R-D PERFORMANCE COMPARED WITH THE HM-8.0 ANCHOR WITHOUT RATE CONTROL

	layer 1 IBBB	layer 1 hierarchical-B	highest layer IBBB	highest layer hierarchical-B
ClassA	–	6.6%	–	7.5%
ClassB	3.3%	5.7%	3.1%	6.3%
ClassC	0.2%	1.4%	–0.2%	1.5%
ClassD	1.1%	2.2%	0.9%	2.0%
ClassE	–1.6%	–	–0.7%	–
ave	1.1%	4.1%	1.0%	4.4%

without rate control are shown in Tables VIII and IX, respectively. The R-D performances of both layer 1 and the highest layer are shown to demonstrate the effectiveness of the proposed algorithm. It should be noted that the bitrates of the layer 1 and the highest layer include the bitrate of both the current layer and its lower layers. From these two tables, we can see that the λ-domain rate control algorithm can achieve an average of over 10% and 20% R-D performance improvement compared with the Q-domain rate control algorithm in IBBB and hierarchical-B cases, respectively. When compared with the HM-8.0 anchor without rate control, the proposed algorithm only suffers an average of about 1% and 4% R-D performance loss in IBBB and hierarchial-B cases for both layer 1 and the highest layer. Fig. 6 shows a typical example of the R-D curve for the highest layer. The R-D curve also demonstrates that the λ-domain rate control algorithm shows much better R-D performance than the Q-domain rate control algorithm. Besides, when compared with the HM-8.0 anchor without rate control, no obvious R-D curve difference can be observed in high bitrate case and slightly better R-D performance can be observed in low bitrate case.

Except for the R-D performance, the subjective quality is also quite important for a rate control scheme. We present some examples of the subjective quality comparisons between the proposed λ-domain rate control algorithm and the Q-domain rate

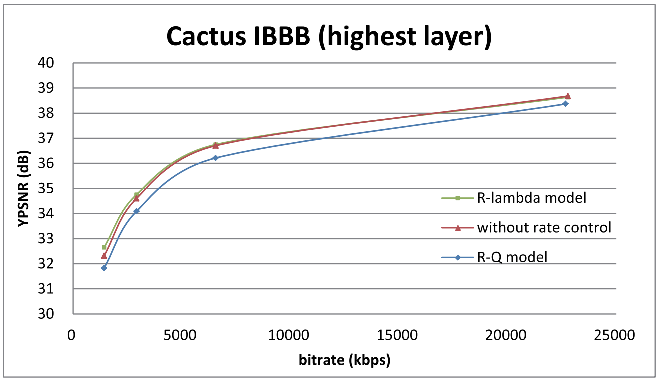


Fig. 6. Typical R-D curve example for temporal scalability.

(a)

(b)

Fig. 7. Subjective quality comparison for temporal scalability. (a) Hierarchical-B case. (b) IBBB case.

control algorithm in Fig. 7. The left two pictures are the reconstructed pictures resulting from the rate control algorithm proposed in this paper. The right two pictures are the reconstructed pictures resulting from the Q-domain rate control algorithm. The above two figures are 200×200 regions selected from the frame No. 356 in the sequence PartyScene with bitrate 349 kbps, 70 kbps, 81 kbps, and 40 kbps for temporal layer 0, 1, 2, and 3, respectively. The bottom two figures are 500×500 regions selected from the scene change frame No. 140 in the sequence Kimono with bitrate 2878 kbps, 1416 kbps, and 1665 kbps for temporal layer 0, 1, and 2 accordingly. From the two groups of pictures, we can observe significant inconsistency of visual quality in the right picture. But there are no such visual quality problem and artifacts in the left picture. The experimental results obviously show that the proposed λ-domain rate control algorithm can achieve better subjective quality compared with the Q-domain rate control algorithm.

B. The Performance of the Proposed Rate Control Algorithm for Spatial and Quality Scalability

The proposed adaptive ratio rate control algorithm for spatial and quality scalability is implemented in scalable HEVC (SHVC) reference software SHM-11.0 to compare with SHM-11.0 anchor without rate control. Besides, the fixed ratio λ-domain rate control algorithm in [13] is extended to support spatial and quality scalability for comparison. As its name implies, the fixed ratio λ-domain rate control algorithm means that the ratio of target bits between different pictures in a GOP is fixed for various sequences. Both hierarchical-B and IBBB/IPPP cases for spatial and quality scalability are tested to show the benefits of the proposed algorithm. In hierarchical-B and IBBB/IPPP cases, the random access and low delay B/P main profile configurations specified in SHVC test conditions [43] are used, respectively. For both the base layer and enhancement layer, the target bits are set the same as SHM-11.0 anchor without rate control. To be more specified, we first generate the anchor of SHM-11.0 and count the bits of both the base layer and enhancement layer. The counted bits of both the base layer and enhancement layer are then rounded and set as the target bits.

Due to the limited space, only the bitrate accuracy and YPSNR for the sequence PeopleOnStreet in IPPP SNR case is shown in Table X. Similar to temporal scalability, the target bitrate in the table only includes the bitrate for the base layer or the enhancement layer. The YPSNR is the average YPSNR of all the frames in the base layer or enhancement layer. From the table, we can see that the proposed adaptive ratio λ-domain rate control algorithm for spatial and quality scalability outperforms the fixed ratio λ-domain rate control algorithm no matter in bitrate error or YPSNR.

The average bitrate errors of all the test sequences for the base layer and enhancement layer are shown in Table XI. From the table, we can see that the proposed adaptive ratio λ-domain rate control algorithm achieves slightly better bitrate accuracy compared with the fixed ratio λ-domain rate control algorithm for both the base layer and enhancement layer. The reason is that the fixed ratio λ-domain rate control algorithm may assign target bits impossible to reach to some frames as the fixed ratio is not content-related, while the adaptive ratio λ-domain rate control algorithm can assign bits according to the characteristic of the current video content. Therefore, the situation where the target bits are impossible to reach will not happen and thus the proposed algorithm can achieve smaller bitrate error.

Since the sequence Kimono has a scene change around the 140th frame, we use the sequence Kimono as an example to explain the content adaptive feature of the proposed adaptive ratio λ-domain rate control algorithm. Fig. 8 shows the bits per frame for the sequence Kimono for both the fixed ratio and adaptive ratio λ-domain rate control algorithms. From the figure, we can see that the bits ratio before and after the scene change are the same under the fixed ratio λ-domain rate control

TABLE X
BITRATE ACCURACY FOR BOTH THE BASE LAYER AND ENHANCEMENT LAYER FOR IPPP SNR CASE (BITRATE IN KBPS, PSNR IN DB)

	layer	target bitrate	fixed R-λ model bitrate	fixed R-λ model bitrate error	fixed R-λ model YPSNR	adaptive R-λ model bitrate	adaptive R-λ model bitrate error	adaptive R-λ model YPSNR	delta YPSNR
PeopleOnStreet	base	20 206	20187.0	0.09%	37.70	20206.9	0.00%	38.01	0.31
	enhancement	41 333	41334.6	0.00%	41.89	41334.9	0.00%	42.07	0.18
	base	11 550	11517.1	0.20%	35.04	11550.3	0.00%	35.49	0.45
	enhancement	21 075	21056.9	0.09%	38.94	21076.0	0.00%	39.29	0.35
	base	7054	7039.7	0.20%	32.66	7053.1	0.01%	33.04	0.38
	enhancement	11 501	11466.4	0.30%	36.30	11501.5	0.00%	36.71	0.41
	base	4527	4516.8	0.23%	30.34	4525.3	0.04%	30.76	0.42
	enhancement	6597	6550.2	0.71%	33.78	6597.2	0.00%	34.09	0.31

TABLE XI
AVERAGE BITRATE ERROR FOR ALL SEQUENCES FOR SPATIAL AND QUALITY SCALABILITY

test cases	base layer fixed R-λ	base layer adaptive R-λ	enhancement layer fixed R-λ	enhancement layer adaptive R-λ
hier-B 2x	0.06%	0.05%	0.04%	0.03%
hier-B 1.5x	0.06%	0.07%	0.03%	0.03%
hier-B SNR	0.03%	0.04%	0.05%	0.03%
IPPP 2x	0.02%	0.02%	0.02%	0.02%
IPPP 1.5x	0.01%	0.01%	0.00%	0.01%
IPPP SNR	0.04%	0.03%	0.07%	0.02%
IBBB 2x	0.02%	0.02%	0.02%	0.02%
IBBB 1.5x	0.01%	0.01%	0.01%	0.01%
IBBB SNR	0.04%	0.03%	0.07%	0.03%
ave	0.032%	0.031%	0.034%	0.022%

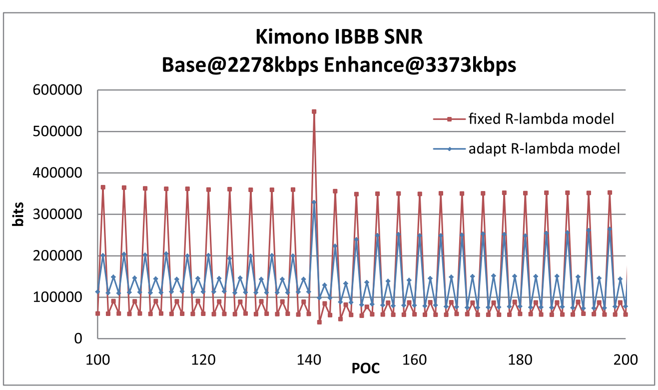


Fig. 8. Bits per frame for spatial and quality scalability.

algorithm. As the characteristic of video content before and after the scene change is totally different, the fixed ratio will be quite unreasonable and lead to serious quality degradation. Different from the fixed ratio λ-domain rate control algorithm, the proposed adaptive ratio λ-domain rate control algorithm can accommodate various video contents and thus can achieve different bits ratio before and after the scene change. Therefore, the proposed adaptive ratio λ-domain rate control algorithm can achieve quite good R-D performance.

As mentioned above, besides the accurate bitrate control, the R-D performance is another important factor to be considered. The R-D performance comparisons with the fixed ratio λ-domain rate control algorithm and the SHM-11.0 anchor without rate control in different test cases are shown in Tables XII and XIII, respectively. For the base layer, an average of 2.0% R-D performance improvement is achieved compared with the fixed ratio λ-domain rate control algorithm. When compared with the anchor without rate control, the adaptive ratio λ-domain rate control algorithm suffers about 2.1% R-D performance loss. For the enhancement layer, the average R-D performance improvement compared with the fixed ratio λ-domain rate control algorithm is about 2.5% and the R-D performance loss compared with SHM-11.0 anchor without rate control is about 3.1% in average. Figs. 9 and 10 show the typical R-D curves for enhancement layer in spatial and quality scalability, respectively. From these two figures, we can see that the adaptive ratio λ-domain rate control can always achieve better R-D performance than the fixed ratio λ-domain rate control algorithm.

TABLE XII
R-D PERFORMANCE COMPARISON WITH THE FIXED RATIO λ-DOMAIN RATE CONTROL ALGORITHM

(a) Hierarchical-B case	base layer 2x	base layer 1.5x	base layer SNR	enhancement layer 2x	enhancement layer 1.5x	enhancement layer SNR
ClassA	–1.7%	–	–2.5%	–4.1%	–	–3.9%
ClassB	–1.9%	–1.8%	–1.7%	–3.2%	–2.8%	–3.6%
ave	–1.9%	–1.8%	–2.0%	–3.5%	–2.8%	–3.7%

(b) IPPP case	base layer 2x	base layer 1.5x	base layer SNR	enhancement layer 2x	enhancement layer 1.5x	enhancement layer SNR
ClassA	–1.6%	–	–4.5%	–2.4%	–	–3.9%
ClassB	–1.6%	–1.8%	–2.0%	–1.5%	–1.5%	–2.2%
ave	–1.6%	–1.8%	–2.7%	–1.8%	–1.5%	–2.7%

(c) IBBB case	base layer 2x	base layer 1.5x	base layer SNR	enhancement layer 2x	enhancement layer 1.5x	enhancement layer SNR
ClassA	–1.7%	–	–4.6%	–2.4%	–	–3.9%
ClassB	–1.7%	–1.9%	–2.3%	–1.9%	–1.8%	–2.6%
ave	–1.7%	–1.9%	–3.0%	–2.1%	–1.8%	–3.0%

TABLE XIII
R-D PERFORMANCE COMPARISON WITH SHM-11.0 ANCHOR WITHOUT RATE CONTROL

(a) Hierarchical-B case

	base layer			enhancement layer		
	2x	1.5x	SNR	2x	1.5x	SNR
ClassA	4.5%	–	7.5%	5.4%	–	0.9%
ClassB	2.5%	3.3%	5.7%	6.9%	4.8%	5.4%
ave	3.1%	3.3%	6.2%	6.5%	4.8%	4.1%

(b) IPPP case

	base layer			enhancement layer		
	2x	1.5x	SNR	2x	1.5x	SNR
ClassA	0.9%	–	3.1%	0.4%	–	0.2%
ClassB	–2.3%	–0.7%	5.0%	2.5%	1.7%	2.4%
ave	–1.4%	–0.7%	4.4%	1.9%	1.7%	1.8%

(c) IBBB case

	base layer			enhancement layer		
	2x	1.5x	SNR	2x	1.5x	SNR
ClassA	1.7%	–	3.7%	1.5%	–	4.0%
ClassB	–1.8%	–0.1%	5.6%	3.2%	2.4%	2.9%
ave	–0.8%	–0.1%	5.1%	2.7%	2.4%	2.4%

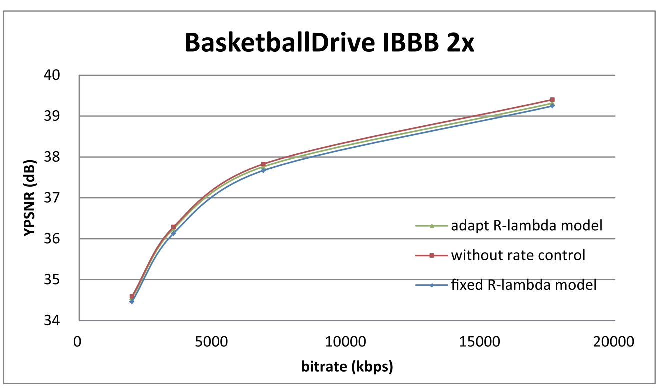


Fig. 9. Typical R-D curve example for spatial scalability.

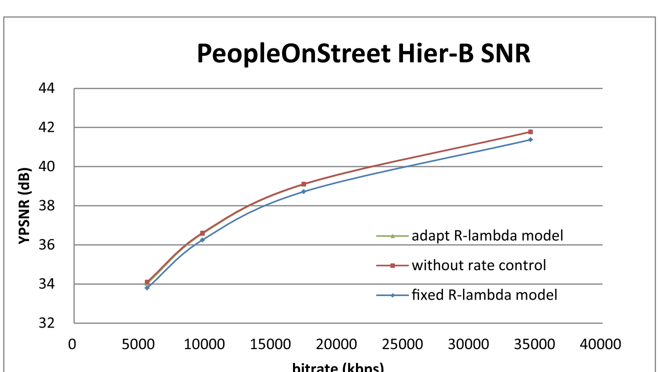


Fig. 10. Typical R-D curve example for quality scalability.

When compared with the SHM-11.0 anchor without rate control, the adaptive ratio λ-domain rate control algorithm sometimes shows similar R-D performance and sometimes presents some R-D performance loss. In term of subjective quality, in both the adaptive and fixed ratio rate control algorithms, we are unable to observe obvious spatial and temporal discontinuity.

C. *Summary of Experimental Results*

From the above experimental results and analysis, we can conclude that

1) The bitrate error between the actual bitrate and the target bitrate under the proposed rate control algorithm is rather small, which verifies that the R-λ model is sound in the HEVC scalable extension context and the proposed rate control algorithm has a very high bitrate control accuracy
2) Both the R-D performance and subjective quality of the proposed rate control algorithm for HEVC temporal scalability is much better than the existing Q-domain rate control algorithm.
3) The proposed rate control algorithm for HEVC spatial and quality scalability can adapt to the characteristics of various video content and thus can achieve better R-D performance than the original fixed ratio λ-domain rate control algorithm.

V. CONCLUSION

In this paper, we propose a λ-domain rate control algorithm for HEVC scalable extension. All the commonly used scalabilities including temporal, spatial and quality scalability are taken into consideration. To achieve the best R-D performance, an optimal initial target bits and encoding parameters determination algorithm is proposed for all kinds of salabilities. Besides, for HEVC spatial and quality scalability, an adaptive ratio bit allocation algorithm is proposed taking both the intra layer and inter layer dependency into consideration. Moreover, to achieve more accurate bitrate accuracy, an adaptive updating algorithm for R-λ model is proposed to control the bits per picture more precisely. The proposed λ-domain rate control algorithm for temporal scalability is implemented in HEVC. Both smaller bitrate error and much better R-D performance are achieved compared with the Q-domain rate control algorithm. The proposed adaptive ratio λ-domain rate control algorithm for spatial and quality scalability is implemented in SHVC. Compared with fixed ratio λ-domain rate control algorithm, both smaller bitrate error and better R-D performance are achieved. In summary, the proposed λ-domain rate control algorithm for HEVC scalable extension can make the bitstream more adaptable to the various characteristics of different devices and various network conditions of different users.

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, Sep. 2007.

[2] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 8, pp. 1374–1387, Dec. 2000.

[3] S. T. Hsiang and J. W. Woods, "Embedded video coding using invertible motion compensated 3-D subband/wavelet filter bank," *Signal Process., Image Commun.*, vol. 16, no. 8, pp. 705–724, May 2001.

[4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[5] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable extensions of the high efficiency video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20–34, Jan. 2016.

[6] D.-K. Kwon, M.-Y. Shen, and C.-C. 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.

[7] M. Jiang and N. Ling, "Low-delay rate control for real-time H.264/AVC video coding," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 467–477, Jun. 2006.

[8] J. Luo, I. Ahmad, and Y. Sun, "Controlling the bit rate of multi-object videos with noncooperative game theory," *IEEE Trans. Multimedia*, vol. 12, no. 2, pp. 97–107, Feb. 2010.

[9] Z. He, Y. K. Kim, and S. 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.

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

[11] Z. He and D. Wu, "Linear rate control and optimum statistical multiplexing for H.264 video broadcast," *IEEE Trans. Multimedia*, vol. 10, no. 7, pp. 1237–1249, Nov. 2008.

[12] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by R-lambda model for HEVC," doc. JCTVC-K0103, Shanghai, China, Oct. 10–19, 2012.

[13] B. Li, H. Li, L. Li, and J. Zhang, "λ domain rate control algorithm for high efficiency video coding," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841–3854, Sep. 2014.

[14] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, "Adaptive basic unit layer rate control for JVT," doc. JVT-G012, Pattaya, Thailand, Mar. 7–14, 2003.

[15] H. Wang and S. Kwong, "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, Jan. 2008.

[16] W. Wu and H. K. Kim, "A novel rate control initialization algorithm for H.264," *IEEE Trans. Consum. Electron.*, vol. 55, no. 2, pp. 665–669, May 2009.

[17] K.-P. Lim, G. J. Sullivan, and T. Wiegand, "Text description of joint model reference encoding methods and decoding concealment," doc. JVT-N046, Hong Kong, China, Jan. 2005.

[18] F. Shao, G. Jiang, W. Lin, M. Yu, and Q. Dai, "Joint bit allocation and rate control for coding multi-view video plus depth based 3D video," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1843–1854, Dec. 2013.

[19] S. Hu, H. Wang, S. Kwong, T. Zhao, and C. Kuo, "Rate control optimization for temporal-layer scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1152–1162, Aug. 2011.

[20] C.-W. Seo, J. W. Kang, J.-K. Han, and T. Nguyen, "Efficient bit allocation and rate control algorithms for hierarchical video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 9, pp. 1210–1223, Sep. 2010.

[21] J. Reichel, H. Schwarz, and M. Wien, "Joint scalable video model 11 (JSVM 11)," doc. JVT-X202, Jul. 2007.

[22] A. Leontaris and A. M. Tourapis, "Rate control for the joint scalable video model (JSVM)," doc. JVT-W043, San Jose, CA, USA, Apr. 21–27, 2007.

[23] J. Liu, Y. Cho, and Z. Guo, "Practical rate control algorithm for temporal scalability in scalable video coding," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1641–1644.

[24] V. P. Binh and S. H. Yang, "Initial quantization parameter determination for H.264/SVC," in *Proc. Int. Conf. Comput., Manage. Telecommun.*, pp. 225–229, Jan. 2013.

[25] L. Xu, W. Gao, X. Ji, D. Zhao, and S. Ma, "Rate control for spatial scalable coding in SVC," presented at the Picture Coding Symp., Lisbon, Portugal, Nov. 2007.

[26] Y. Liu, Y. Soh, and Z. Li, "Rate control for spatial/CGS scalable extension of H.264/AVC," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 1746–1750.

[27] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, "Rate and distortion modeling of CGS coded scalable video content," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 165–180, Apr. 2011.

[28] J. Liu, Y. Cho, Z. Guo, and C.-C. 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.

[29] 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.

[30] J. Yang, Y. Sun, Y. Zhou, and S. Sun, "Incremental rate control for H.264 scalable video coding," in *Proc. Global Telecommun. Conf.*, Dec. 2010, pp. 1–5.

[31] Y. Pitrey, Y. Serrand, M. Babel, and O. Deforges, "ρ-domain for low-complexity rate control on MPEG-4 scalable video coding," in *Proc. 10th IEEE Int. Symp. Multimedia*, Dec. 2008, pp. 89–96.

[32] Y. Pitrey, M. Babel, O. Deforges, and J. Vieron, "ρ-domain based rate control scheme for spatial, temporal, and quality scalable video coding," *Proc. SPIE, Vis. Commun. Image Process.*, vol. 7257, Jan. 2009, Art. no. 725705.

[33] M. Liu, Y. Guo, H. Li, and C. W. Chen, "Low-complexity rate control based on ρ-domain model for scalable video coding," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1277–1280.

[34] L. Li and H. Li, "λ domain based optimal bit allocation for scalable high efficiency video coding," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2015, pp. 2788–2791.

[35] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[36] H. Choi, J. Nam, J. Yoo, D. Sim, and I. Bajic, "Rate control based on unified R-Q model for HEVC," doc. JCTVC-H0213, San Jose, CA, USA, Feb. 2012.

[37] A. Ortega and K. Ramchandran, "Rate-distortion methods for image and video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 23–50, Nov. 1998.

[38] HM, *HEVC Test Model*. 2016, Berlin, Germany. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/

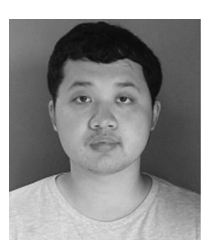
[39] F. Bossen, "Common test conditions and software reference configurations," doc. JCTVC-J1100, Stockholm, Sweden, Jul. 11–20, 2012.

[40] SHM, *Scalable HEVC Test Model*. 2016, Berlin, Germany. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_SHVCSoftware/

[41] Y.-J. Ahn, X. Wu, W. Lim, and D. Sim, "Target bits saturation to avoid CPB overflow and underflow under the constraint of HRD," doc. JCTVC-U0132, Warsaw, Poland, Jun. 19–26, 2015.

[42] B. Li, D. Zhang, H. Li, and J. Xu, "QP determination by lambda value," doc. JCTVC-I0426, Geneva, Switzerland, Apr. 27–7 May, 2012.

[43] X. Li, J. Boyce, P. Onno, and Y. Ye, "Common test conditions and software reference configurations for the scalable test model," doc. JCTVC-L1009, Geneva, Switzerland, Jan. 14–23, 2013.



Li Li received the B.S. degree in electronic engineering from the University of Science and Technology of China, Hefei, China, in 2011, where he is currently working toward the Ph.D. degree in electronic engineering and information science.

His research interests include image/video coding and processing.

Bin Li (M'13) received the B.S. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China, Hefei, China, in 2008 and 2013, respectively.

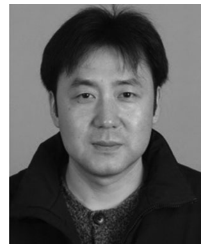
He joined Microsoft Research Asia, Beijing, China, in 2013, where he is currently a Researcher. He has authored or coauthored more than 20 papers. He holds more than ten granted or pending U.S. patents in the area of image and video coding. He has more than 30 technical proposals that have been adopted by Joint Collaborative Team on Video Coding. His current research interests include video coding, processing, and communication.

Dr. Li is currently the Co-Chair of the Ad Hoc Group of Screen Content Coding Extensions Software Development. He has been an active contributor to ISO/MPEG and ITU-T video coding standards. He was the recipient of the Best Paper Award for the International Conference on Mobile and Ubiquitous Multimedia from Association for Computing Machinery in 2011, as well as the Top 10% Paper Award of the 2014 IEEE International Conference on Image Processing.

Dong Liu (M'13) received the B.S. and Ph.D. degrees in electrical engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2004 and 2009, respectively.

He was a Member of Research Staff with the Nokia Research Center, Beijing, China, from 2009 to 2012. He joined USTC as an Associate Professor in 2012. He has authored or coauthored more than 20 papers in journals and international conferences. His research interests include image/video compression, multimedia signal processing, and data mining.

Prof. Liu was the recipient of the 2009 IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY Best Paper Award.



Houqiang Li (M'10–SM'12) received the B.S., M.Eng., and Ph.D. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 1992, 1997, and 2000, respectively, all in electronics engineering.

He is currently a Professor with the Department of Electronic Engineering and Information Science, USTC. He has authored or coauthored more than 100 papers in journals and conferences. His current research interests include video coding and communication, multimedia search, and image/video analysis.

Prof. Li served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from 2010 to 2013, and has been on the Editorial Board of the *Journal of Multimedia* since 2009. He has served on technical/program committees and organizing committees and as a Program Co-Chair and a Track or Session Chair for more than ten international conferences. He was the recipient of the Best Paper Award for Visual Communications and Image Processing in 2012, the International Conference on Internet Multimedia Computing and Service in 2012, and the International Conference on Mobile and Ubiquitous Multimedia from ACM in 2011, and is a Senior Author of the recipient of the Best Student Paper Award of the Fifth International Mobile Multimedia Communications Conference in 2009.