Rate Control Optimization for Temporal-Layer Scalable Video Coding

Sudeng Hu, Hanli Wang, *Member, IEEE,* Sam Kwong, *Senior Member, IEEE,*
Tiesong Zhao, *Student Member, IEEE,* and C.-C. Jay Kuo, *Fellow, IEEE*

***Abstract*—A novel frame-level rate control (RC) algorithm is presented in this paper for temporal scalability of scalable video coding. First, by introducing a linear quality dependency model, the quality dependency between a coding frame and its references is investigated for the hierarchical B-picture prediction structure. Second, linear rate-quantization (R-Q) and distortion-quantization (D-Q) models are introduced based on different characteristics of temporal layers. Third, according to the proposed quality dependency model and R-Q and D-Q models for each temporal layer, adaptive weighting factors are derived to allocate bits efficiently among temporal layers. Experimental results on not only traditional quarter common intermediate format/common intermediate format but also standard definition and high definition sequences demonstrate that the proposed algorithm achieves excellent coding efficiency as compared to other benchmark RC schemes.**

***Index Terms*—Hierarchical B-picture prediction, rate control, rate-distortion model, scalable video coding, temporal scalability.**

I. Introduction

SCALABLE video coding (SVC) [1], [2] has been developed by the Joint Video Team (JVT) as an extension of H.264/advanced video coding (AVC), aiming to encode the video signal once, but enable decoding from partial streams depending on the specific rate and resolution required by a specific application. In general, three types of scalability are designed, including temporal, spatial, and quality scalability. These types of scalability can be combined together or applied separately to provide a variety of benefits for various kinds of applications [1], [3]–[5]. Among them, the temporal scalability can be considered as a basis that the other two types of scalability can be built upon and it is used to provide variable temporal resolution sequences for different end-users who usually have specific temporal resolution preferences or transmission constraints.

Although scalable video bitstreams offer rich functionalities of rate adaptation in a wonderful fashion, it is not guaranteed that the generated scalable bitstreams are optimal in the rate-distortion (R-D) sense, which makes rate control (RC) also a necessity for SVC aiming to optimize the encoded video quality given the target bit rate (BR) (or bandwidth). Usually, video coding standards recommend their own non-normative RC algorithms during the standardization process, such as the Test Model 5 [6] for MPEG-2, Test Model Near-term 8 [7] for H.263, and Verification Model 18 [8] for MPEG-4. For H.264/AVC, JVT-W042 [9] is developed based on JVT-G012 [10] and meanwhile introduces several new features and has been adopted in Joint Model (JM).

SVC becomes much more complex than the previous standards and thus leads to new challenges in RC design. In this paper, we mainly focus on the RC for SVC temporal scalability. One of the problems arisen from the concept of multiple layers is how to allocate source bits efficiently among different temporal layers. To achieve temporal scalability, the hierarchical B-picture prediction (HBP) structure [11] is usually employed, which was developed for H.264/AVC at the very beginning and implemented in JM reference softwares. Due to its temporal scalable natures and good R-D performances, the HBP structure is also adopted in SVC Joint Scalable Video Model (JSVM) reference softwares to provide scalability for temporal layers. Unlike the traditional "IBBP" or "IPPP" group-of-picture (GoP) structure, multiple temporal layers are introduced by HBP in a more complex manner that the pictures of lower layers are referenced by the pictures of higher layers, resulting in complex quality dependency among the temporal layers. Several RC algorithms are proposed in H.264/AVC and SVC for HBP structure, aiming at allocating target bits reasonably to different layers. In [12], Liu *et al.* proposed a switchable mean absolute of difference (MAD) prediction scheme and a list of empirical weighting factors to allocate source bits to different temporal layers. Although the constant weighting factors improve the bit allocation strategy, it cannot reflect the changing content of video sequences and thus is unable to maximize coding efficiency. In [13], Xu *et al.*

Manuscript received April 1, 2011; revised August 26, 2010 and December 4, 2010; accepted March 5, 2011. Date of publication April 7, 2011; date of current version August 3, 2011. This work was supported in part by the Hong Kong Research Grants Council General Research Fund, under Project 9041353 (CityU 115408), and in part by the Program for Professors of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, the Program for New Century Excellent Talents in University of China, and the National Basic Research Program (973 Program) of China, under Grant 2010CB328101. This paper was recommended by Associate Editor G. Wen.

S. Hu, S. Kwong, and T. Zhao are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail: hsudeng2@student.cityu.edu.hk; cssamk@cityu.edu.hk; ztiesong2@student.cityu.edu.hk).

H. Wang is with the Department of Computer Science and Technology and Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China (e-mail: hanliwang@tongji.edu.cn).

C.-C. J. Kuo is with the Ming Hsieh Department of Electrical Engineering and Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA (e-mail: cckuo@sipi.usc.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.2011.2138810

proposed a scaling factor scheme to indicate the importance of different temporal layers in the bit allocation scheme. Li *et al.* [14] developed an initial quantization parameter (Q_p) determination scheme and design a frame-level bit allocation method. In [15], a Q_p selection tree framework is built to select proper Q_p values for different layers. In [16], bits and distortion dependency between temporal layers have been investigated for bit allocation, however, it is a multiple-pass scheme that a single GoP is encoded several times (depending on the number of temporal layers in the HBP structure) with different Q_p values in order to derive model parameters for RC. Thus, it is not suitable for real-time communication applications. Currently, although there are no comprehensive RC algorithms which are able to regulate BRs for all scalable layers, JVT-W043 [17] is adopted in the JSVM reference softwares as an initial version to control the output bitstream for temporal layer. It takes considerations of the quality dependency among the temporal layers by increasing Q_p from lower temporal layers to higher layers. Although quality dependency has been considered during the bit allocation process in these RC algorithms, the quality dependency relationship inside the HBP structure has not been fully investigated, and a more accurate dependency model is desired. In addition, the differences of R-D characteristics of each temporal layer are also important factors that should be considered for source bits allocation.

In this paper, a linear quality dependency model between a coding picture and its references is proposed. Based on this linear model, the importance of pictures in different temporal layers is exploited by analyzing its quality effect on the whole video sequence. Because of the complex reference relationship, the pictures in the HBP structure either directly or indirectly affect one another on quality, and therefore the pictures in different layers can have different effects on the quality of the whole sequence. In addition to quality dependency, the R-D characteristics of the temporal layers should be also considered in order to achieve better R-D performance when source bits are allocated. Although several R-D models have been proposed for RC in previous standards, those single layer R-D models are no longer accurate enough to model the complex R-D characteristics for SVC. Therefore, effective linear rate-quantization (R-Q) and distortion-quantization (D-Q) models are proposed for multiple temporal layers. Based on the proposed models, a novel R-D optimized RC scheme is developed. On the contrary, the basic coding unit for RC is a group of macroblocks (MBs), e.g., a frame or a MB or a segment of frame. Usually, the larger size of the basic unit, the better video quality can the RC algorithm achieve, but at the cost of degradation of BR accuracy. In this paper, we propose a frame level RC algorithm, which also can obtain excellent accuracy of BR achievement.

The rest of this paper is organized as follows. In Section II, the typical dyadic HBP structure is introduced, and the quality dependency and BR dependency inside the HBP structure are investigated. In Section III, a linear D-Q model and a linear R-Q model are proposed for temporal-layer SVC. In Section IV, adaptive weighting factors for BR allocation are derived to optimize the R-D performance. Then, a RC algorithm at both GoP and frame levels are presented. In Section V, experimental results are given to demonstrate the efficiency of the proposed RC algorithm. Finally, Section VI concludes this paper.

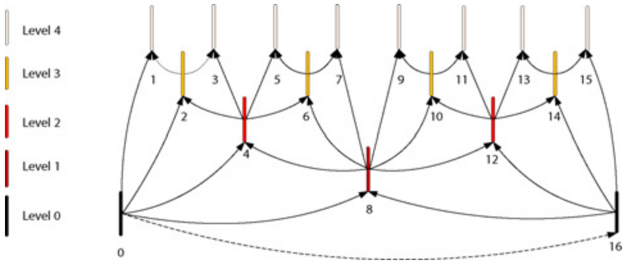


Fig. 1. Dyadic HBP structure with GoP size of 16.

II. R-D Dependency in Hierarchical Structure

A. HBP Structure

A typical dyadic HBP structure with GoP size of 16 is depicted in Fig. 1. The first picture in the video sequence is intra-coded as an instantaneous decoder refresh (IDR). The pictures located at the bottom of the structure are called key pictures, which can be either intra-coded as I frame or inter-coded as P frame. The rest of the pictures are coded as B frames and stored as references for pictures of higher temporal layers. For simplicity, the layers from the base to the highest layer are assigned with a level number which starts from 0 and is increased by 1 from a lower layer to the next higher layer as shown in Fig. 1.

In the hierarchical structure, the pictures in lower temporal layers are used as references by the pictures in higher temporal layers. Usually, the two nearby pictures are chosen for references as illustrated in Fig. 1. Although multiple reference picture concept of H.264/AVC can also be applied in the HBP structure, the most nearby pictures are more likely to provide efficient references. To verify this, in Table I, the number of 8 × 8 blocks using different reference indices is recorded. In the simulation, the reference number is set to 5, the GoP size is set to 16 and the number of GoPs is 5, the basis Q_p[1] is set to 32. As shown in Table I, almost only the first picture in either of the two reference lists is used as the reference. Meanwhile, the average peak signal-to-noise ratio (PSNR) and BR results are recorded when the reference number is set to 1 and 5, respectively. It is obvious in Table I that the setting with one reference picture has comparable coding efficiency to the setting with five reference pictures. This is because the closest pictures are most likely to be selected as the best reference even when more reference pictures are available. Therefore, we focus our analysis on the typical prediction structure of HBP as shown in Fig. 1.

B. Linear Quality Dependency

The video coding technology of motion-compensated prediction results in the quality dependency between a coding

[1] In order to improve the coding efficiency, cascading Q_p concept is applied in SVC. In the JSVM reference software [18], the actual Q_p values for different temporal layers are calculated according to the basis Q_p.

TABLE I
NO. OF 8 × 8 BLOCKS USING DIFFERENT REFERENCE INDICES

Reference No.			5								1	
	Sequence	Index:	1		2	3	4	5	BR (kb/s)	PSNR (dB)	BR (kb/s)	PSNR (dB)
CIF	*Coastguard*	List0: List1:	78 891 81 537	(98.6%) (98.7%)	1066 1041	45 21	16 0	6 0	626.24	32.78	625.31	32.79
	Container	List0: List1:	91 427 92 447	(99.4%) (99.1%)	345 747	32 112	36 20	110 0	133.79	36.94	131.69	36.90
	Football	List0: List1:	51 799 50 465	(93.0%) (96.1%)	2895 1721	669 268	210 39	128 0	1220.8	33.50	1221.6	33.52
QCIF	*Table Tennis*	List0: List1:	16 927 16 791	(95.3%) (97.0%)	610 491	174 20	44 8	4 0	100.92	35.25	101.35	35.25
	News	List0: List1:	22 247 22 369	(99.7%) (99.2%)	46 190	12 1	1 0	0 0	52.24	37.87	53.28	37.86
	Paris	List0: List1:	20 859 20 806	(99.2%) (98.9%)	157 227	15 15	5 0	0 0	110.22	35.47	111.01	35.46

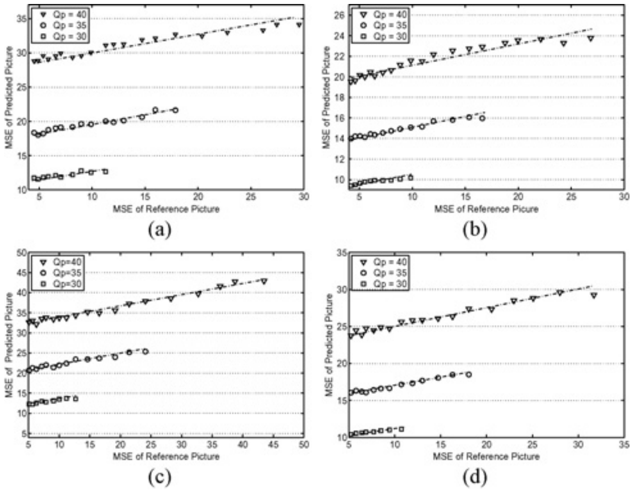


Fig. 2. Quality dependency in the sequence *Paris* with GoP size of 16. (a) Quality dependency relation between the second picture and its reference fourth picture for QCIF. (b) Quality dependency relation between the second picture and its reference fourth picture for CIF. (c) Quality dependency relation between the fourth picture and its reference eighth picture for QCIF. (d) Quality dependency relation between the fourth picture and its reference eighth picture for CIF.

picture and its reference pictures. It makes the problem of R-D performance optimization become more complicated in RC algorithms. In the HBP structure, the situation becomes even more complicated than the traditional "IBBP" or "IPPP" structure, because the reference relationship in the HBP structure is much more complex than the traditional structure. Moreover, the quality dependency is one of the most important factors affecting the bit allocation scheme in RC. In [16] and [19], the quality dependency relation is studied for MPEG-2 and H.264/SVC encoders, respectively. In this paper, we will first investigate the quality dependency in the HBP structure.

In Fig. 2, the dependency between the reference and its directly predicted picture is investigated for the video sequence *Paris* (of both QCIF and CIF) with the GoP size equal to 16. Similar observations are also drawn for other benchmark sequences. In Fig. 2(a) and (b), the Q_p values of the second picture in GoP are fixed with 30, 35, and 40, respectively, while the Q_p values of its reference (the fourth picture in GoP) are changed from 20 to 29, 20 to 34, and 20 to 39, respectively. In Fig. 2(c) and (d), the Q_p values of the fourth picture in GoP are fixed with 30, 35, and 40, and the Q_p values of its reference (the eighth picture in GoP) are changed from 20 to 29, 20 to 34, and 20 to 39, respectively. In Fig. 2, the quality is measured in terms of mean squared error (MSE). From statistical analysis, the dependency relationship can be approximately estimated by a linear model as expressed by

$$\Delta D = \alpha \cdot \Delta D_{ref} \tag{1}$$

where ΔD_{ref} is a change in the MSE of the reference picture and ΔD is the corresponding change in the predicted picture. α is the affecting factor varying in the range of [0.2, 0.5]. In this paper, α is set to 0.4 based on extensive experiments. Note that the linear quality dependency relation in (1) is consistent with the observation presented in [16].

C. Propagation of Quality Dependency

The predicted picture may be further referenced by other pictures in the HBP structure, therefore the quality change in the predicted picture may consequently cause quality changes of other pictures. This propagation of quality changes is illustrated in Fig. 3 for the sequence *Mobile* in QCIF format, where the quality change of the eighth picture has a direct effect on the fourth picture and indirect effects on the second and first pictures. From Fig. 3, it is observed that the propagation of quality change becomes weaker from close temporal layers to far temporal layers, e.g., the effect of the quality change of the eighth picture on the second picture is higher than that on the first picture.

Through the explicit and implicit dependency in the HBP structure, different pictures have different quality influences on the entire sequence. In order to evaluate this kind of influence of different temporal pictures, we investigate the relationship between the quality change of each picture and the total quality change that is consequently caused by this change. Let

$$\Delta D_{total} = f_n(\Delta D_n) \tag{2}$$

where ΔD_n is the quality change in picture of level n and ΔD_{total} is the corresponding total change in the sequence

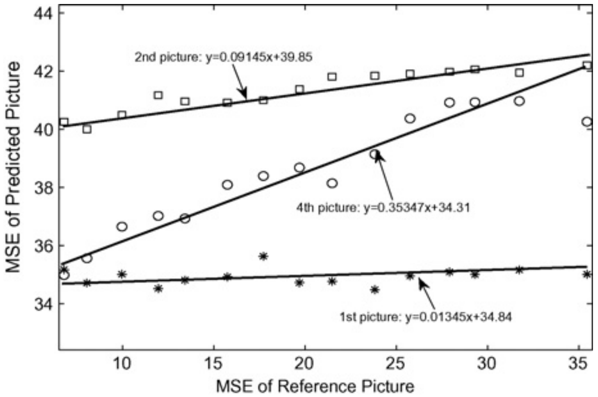


Fig. 3. Propagation of quality change effects on the sequence *Mobile* (QCIF). With the GoP size of 16, the Q_p values of the eighth picture are changed from 21 to 35, meanwhile the Q_p values of the first, second, and fourth pictures are set to 38, 37, and 36, respectively.

including both explicit and implicit changes; $f_n(\cdot)$ is their relationship function and the subscript n is the level index of the picture, because pictures in different positions have different functions. Our purpose is to find out an explicit expression of $f_n(\cdot)$.

For discussion convenience, let N be the level number of the highest temporal layer, e.g., $N = 4$ in Fig. 1. The reference structure of dyadic HBP structure is demonstrated in Fig. 1, where except the highest temporal layer, each B frame is directly referenced by two pictures of its every higher levels. For example, the fourth picture in level 2 is totally referenced by four pictures (the second and sixth pictures in level 3 and the third and fifth pictures in level 4). On the contrary, the key pictures in the base layer is not only referenced by higher layer pictures, but also referenced by the next key picture.

Based on the reference structure according to Fig. 1, the total quality change in the entire sequence caused by a picture of level i can be expressed as

$$f_i(\Delta D_i) = \Delta D_i + \sum_{k=i+1}^{N} 2 f_k(\alpha \cdot \Delta D_i). \tag{3}$$

In fact, (3) is a recursive expression of $f_n(\cdot)$. According to mathematical induction, by applying (3) recursively from the highest layer to the temporal layer of level 1, we can obtain

$$f_i(\Delta D_i) = (1 + 2\alpha)^{(N-i)} \Delta D_i, \qquad i = 1, \ldots, N. \tag{4}$$

On the contrary, for key pictures except the first IDR picture, depending on the number of successive P frames M in the base layer, the total change can be written as

$$\begin{aligned} f_0(\Delta D_0) &= \sum_{k=0}^{M} \alpha^k (1 + 2\alpha)^N \Delta D_0 \\ &= \frac{(1 - \alpha^{(M+1)}) \cdot (1 + 2\alpha)^N}{1 - \alpha} \Delta D_0. \end{aligned} \tag{5}$$

As far as the first IDR picture is concerned, an initial Q_p is usually given either by an empirical value or by initialization schemes [14], [20].

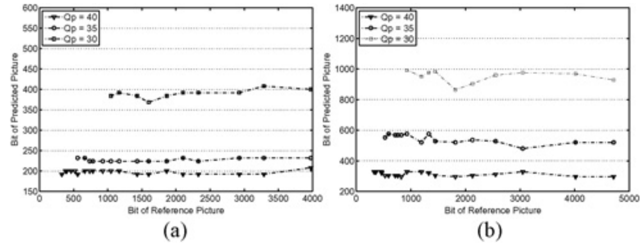


Fig. 4. Relation of BR dependency in sequence *Akiyo* (QCIF) with GoP size of 16. (a) Second picture and its reference fourth picture. (b) Fourth picture and its reference eighth picture.

Because $\alpha < 0.5$ as discussed before and M is usually larger than or equal to 2, after approximating (5), a general expression of $f_n(\cdot)$ is derived as

$$f_i(\Delta D_i) = \begin{cases} \frac{(1+2\alpha)^N}{1-\alpha} \Delta D_i, & i = 0 \\ (1 + 2\alpha)^{(N-i)} \Delta D_i, & 1 \le i \le N. \end{cases} \tag{6}$$

Equation (6) reveals the linear relationship with different impact factors between the quality change of each individual picture and the corresponding total quality changes of the entire video sequence.

D. BR Dependency

The BR dependency between a coding picture and its references is also investigated, aiming at a deeper exploration of R-D relations. In Fig. 4, taking the sequence *Akiyo* (QCIF) as an example, the Q_p value of a particular picture is fixed and the amount of its output bits is recorded by changing the Q_p values of its references. The Q_p values of the predicted pictures are set to 30, 35, and 40, respectively, while the Q_p values of the reference pictures are changed from 20 to 29, 34, and 39, respectively. It can be observed that the number of coded bits of reference has limited effects on the number of bits of the predicted picture.

For further investigation, in Fig. 5, we change the Q_p value of a certain temporal layer while fixing the Q_p values of the rest layers. The BRs of different layers are recorded in the figure. In the simulation, the GoP size is set to 16 and the number of GoPs is 5. Apparently, except the layer with changing Q_p value, the number of coded bits of other layers almost keeps constant, which indicates the BR independency among the temporal layers. A similar conclusion about the BR dependency relation between temporal layers is also drawn in [16]. Therefore, we assume that the number of output bits of a picture is not affected by their references.

III. R-D Model in SVC

A. R-Q Model

The relationship between BR (R) and quantization parameter (Q_p) or quantization step size (Q_{step}) has been studied extensively in previous papers [21]–[25], and a number of R-Q models have been developed based on observations and analyses. Before presentation of the R-Q model used in this paper, the difference between Q_p and Q_{step} is briefly

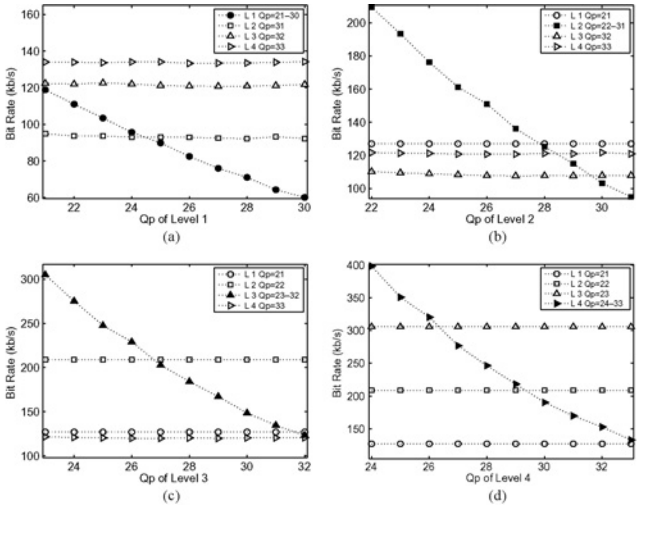


Fig. 5. BR dependency in sequence *Football* (QCIF). (a) Q_p of level 1 changes. (b) Q_p of level 2 changes. (c) Q_p of level 3 changes. (d) Q_p of level 4 changes.

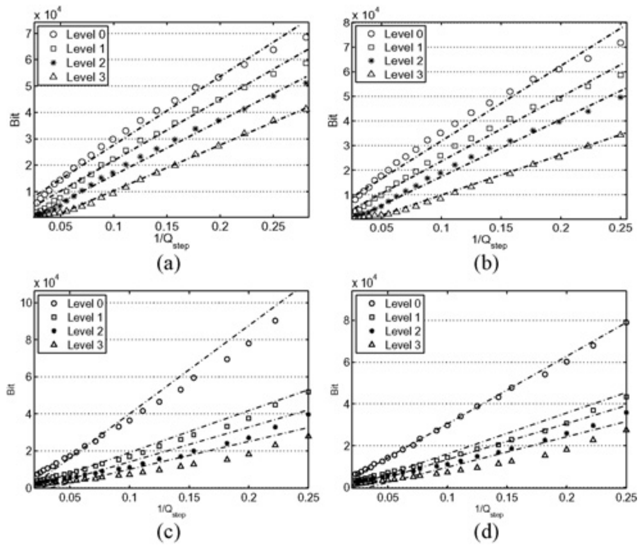


Fig. 6. Relationship between R and $1/Q_{step}$ for the 16th picture in level 0, the eighth picture in level 1, the fourth picture in level 2, and the second picture in level 3 within the same GoP. (a) First GoP in *Coastguard* (QCIF). (b) Third GoP in *Coastguard* (QCIF). (c) First GoP in *News* (CIF). (d) Third GoP in *News* (CIF).

introduced below. In scalar quantization, Q_{step} is the actual step size employed by a quantizer, while Q_p indicates the index of Q_{step}. In previous video coding standards such as H.263 and MPEG-4, Q_p is usually linear with Q_{step}, e.g., $Q_{step} = 2Q_p$ in H.263. However, in H.264/AVC and its SVC extension, there is a nonlinear relationship between Q_{step} and Q_p, i.e., Q_{step} doubles in size for every increment of 6 in Q_p, and according to [26], this nonlinear relationship can be approximated as

$$Q_{step} \approx c_1 \cdot c_2^{Q_p} \tag{7}$$

where c_1 and c_2 are two constant parameters.

With the assumption that the residual coefficients are Laplacian distributed, a quadratic R-Q model is proposed in [21] as written as

$$R = \frac{a \cdot m}{Q_{step}} + \frac{b \cdot m}{{Q_{step}}^2} + C \tag{8}$$

where m indicates the MAD of the residual between the original and reconstructed signal, a and b are the model parameters, and C is the number of bits used to code the header information. It is regarded as a classic model and has been adopted as a nonnormative RC tool in several standards such as MPEG-4 [25], [27], and H.264/AVC [28]. In [22] and [24], a linear model is proposed between R and $1/Q_{step}$ as

$$R = \frac{k \cdot m}{Q_{step}} + C \tag{9}$$

where k and C are the model parameters. As stated in [22], the first term on the right side of the formula can be considered as the number of texture bits and the second term indicates the number of header bits. In this paper, we apply the linear model in (9) for computational simplicity.

Regarding SVC, because multiple temporal layers are employed, a single R-Q model may be no longer accurate enough to describe the R-Q relation for each layer. For example, with similar MAD and same Q_{step} values, the number of bits generated by pictures of a lower layer is generally larger than that of a higher layer. In Fig. 6, the relationship between R and Q_{step} for the sequences *Coastguard* (QCIF) and *News* (CIF) are illustrated for the 16th picture in level 0, the eighth picture in level 1, the fourth picture in level 2, and the second picture in level 3, respectively. From the results, it can be seen that the parameter k in (9), which indicates the slop of the lines in Fig. 6, is quite different for different levels. Inspired by this observation, to model the different characteristics of each temporal layer properly and estimate the corresponding R-Q relation more accurately, the following linear model is applied to SVC as

$$R_i(Q^i_{step}) = \frac{k_i \cdot m_i}{Q^i_{step}} + C_i \tag{10}$$

where i refers to level i in a GoP.

B. D-Q Model

To study the D-Q relationship, a number of works have been proposed in the literature [22], [26], [29]–[32]. In [29], a classic D-Q model is formulated as

$$D = \chi \cdot {Q_{step}}^2 \tag{11}$$

where D represents the distortion of the reconstructed picture usually in terms of MSE and the parameter χ exhibits the typical value of $\frac{1}{12}$. Takagi *et al.* [30] proposed a linear D-Q model between the distortion measurement in PSNR and Q_p, which can be written as

$$PSNR = \rho \cdot Q_p + \zeta \tag{12}$$

where ρ and ζ are the model parameters, and this linear PSNR-Q_p model is applied in [22] and [26] for H.264/AVC

RC. In [31], an empirical linear D-Q model is proposed for H.264/AVC as

$$D = \gamma \cdot Q_{step} \tag{13}$$

where γ is the model parameter. In [32], Kamaci *et al.* designed a Cauchy distribution-based D-Q model as

$$D = \xi \cdot {Q_{step}}^{\beta} \tag{14}$$

where ξ and β are the model parameters. As compared to the Cauchy distribution-based D-Q model expressed in (14), the linear D-Q model in (13), and the classical D-Q model in (11) can be considered as two special cases of the Cauchy distribution-based model with the parameter β equal to 1 and 2, respectively. On the contrary, after mathematical analysis, the linear PSNR-Q_p model in (12) can be considered as a derivative of the Cauchy distribution-based D-Q model, by considering the definition of PSNR and the nonlinear $Q_{step} - Q_p$ relation in (7).

In order to decide which model to be used in this paper, extensive experiments are performed in SVC temporal layers for data collection and thus analysis of the D-Q models. Based on the statistics, it is found that the Cauchy distribution model with the parameter $\beta = 1$, or equally, the linear D-Q model in (13) is more accurate than other models for modeling temporal-layer SVC D-Q relations. More specifically, due to different coding characteristics in different temporal layers, the linear D-Q model with different model parameters can be applied to different temporal layers. For further demonstration, four typical fitting curves are plotted in Fig. 7, where it can be observed that the linear D-Q model can approximate the actual data very well, and for different temporal layers, the model parameter γ indicating the slope of the line in the plot is different. Therefore, due to the fitting accuracy and mathematical simplicity, the following linear model is employed to model D-Q relation in temporal-layer SVC as

$$D_i(Q^i_{step}) = \gamma_i \cdot Q^i_{step} \tag{15}$$

where i represents level i in a GoP.

IV. R-D Optimized Frame-Level RC

A. Adaptive Weighting Factor for Bit Allocation

In general, the bit allocation among temporal layers can be achieved in two ways.

1) The target BRs for each temporal layer are specified according to the requirement of end-users/applications. This kind of allocation scheme may not consider well the dependency among the temporal layers and thus the overall optimal R-D performance may not be reached.
2) The overall bit constraint for all temporal layers is given, and the target bits for each temporal layer are adaptively allocated by considering the dependency among the layers, which is wildly employed in the literature, such as [12] and [13].

In this way, the BR of every temporal layer is adaptively determined during the RC process rather than predefining. In this paper, we focus on the second case with the aim to optimize the visual quality under the overall BR constraint. To this end, adaptive weighting factors are derived to allocate target bits to each temporal layer based on the linear quality dependency analysis and the proposed R-Q and D-Q models.

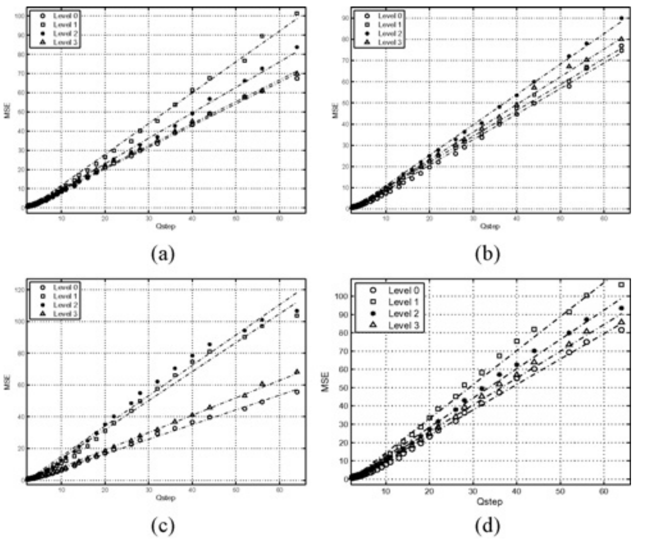


Fig. 7. Linear relationship between D and Q_{step} for the 16th picture in level 0, the eighth picture in level 1, the fourth picture in level 2, and the second picture in level 3 within the same GoP. (a) First GoP in *Soccer* (QCIF). (b) Third GoP in *Soccer* (QCIF). (c) First GoP in *Table* (CIF). (d) Third GoP in *Table* (CIF).

Let N_{rp} denote the total number of remaining pictures in a sequence, j be the picture index in the coding order such that $j = 1, 2, \ldots, N_{rp}$, N refer to the highest level in the HBP structure, and i be the level index such that $i = 0, 1, \ldots, N$. The optimization problem is to choose appropriate Q_{step} values to minimize the total distortion of the rest pictures in subject to the constraint that the number of total bits consumed should be equal to or less than the number of target bits. By applying the method of Lagrangian multiplies, the aim is to minimize the R-D cost J expressed as

$$J = \sum_{j=1}^{N_{rp}} D_j(\vec{Q}_{step}) + \lambda \left(\sum_{j=1}^{N_{rp}} R_j(Q^j_{step}) \right) \tag{16}$$

where λ is the Lagrangian multiplier, the vector parameter $\vec{Q}_{step} = [Q^1_{step}, \ldots, Q^{N_{rp}}_{step}]^T$. Due to the complex reference relation as discussed in Section II-B, the distortion may be also affected by other Q_{step}s, therefore the distortion of the jth frame D_j is a function of $\vec{Q}_{step}$, and we use $D_{total}(\vec{Q}_{step})$ to represent the total distortion of the remaining pictures. According to the Lagrangian method, we can obtain the formula as follows:

$$\frac{\partial D_{total}(\vec{Q}_{step})}{\partial Q^j_{step}} = -\lambda \frac{\partial R_j(Q^j_{step})}{\partial Q^j_{step}}, \ j = 1, \ldots, N_{rp}. \tag{17}$$

Assuming the jth picture belongs into level i, and after substituting (6) into (17), we have

$$\theta_i \frac{\partial D_j(\vec{Q}_{step})}{\partial Q^j_{step}} = -\lambda \frac{\partial R_j(Q^j_{step})}{\partial Q^j_{step}}, \ j = 1, \ldots, N_{rp} \tag{18}$$

where

$$\theta_i = \begin{cases} \frac{(1+2\alpha)^{(N-i)}}{1-\alpha}, & i = 0 \\ (1+2\alpha)^{(N-i)}, & 1 \le i \le N. \end{cases}$$

Based on the linear R-Q model in (10) and the linear D-Q model in (15), for the pictures in level i, (18) can be rewritten as

$$\frac{{Q^i_{step}}^2 \theta_i \gamma_i}{k_i m_i} = \lambda, \ i = 0, \ldots, N \tag{19}$$

where m_i are the predicted MAD values of the remaining pictures in level i. Because those MAD values are unavailable before motion-compensated prediction, they are predicted as

$$m_i = \kappa \cdot m_i^p + (1 - \kappa) \cdot \hat{m}_i \tag{20}$$

where $\hat{m}_i$ is the actual MAD of the last coded picture in level i and m_i^p is the previous MAD prediction in level i. κ is the weighting parameter which is set to 0.7 based on our experiments. Once a picture in level i is coded, its actual MAD value is used to update the MAD prediction for the rest pictures in level i.

Without loss of generality, the picture in the base layer can be regarded as the basis for bit allocation. For level i, considering (10) and (19), an adaptive weighting factor ω_i for texture bits allocation is derived as

$$\begin{aligned} \omega_i &= R_i^t / R_0^t \\ &= \sqrt{\frac{k_i m_i \theta_i \gamma_i}{k_0 m_0 \theta_0 \gamma_0}} \end{aligned} \tag{21}$$

where R_i^t is the number of texture bits assigned to a picture in level i, and R_0^t is the number of texture bits consumed by the key picture.

B. GoP Level Bit Allocation

The target bits for a GoP are allocated according to the bandwidth and buffer fullness after coding the previous GoP. Because the first picture in a GoP is a key picture, it usually generates more bits than the other pictures. To avoid buffer overflow that may be caused by the bit pulse of the key picture, the buffer fullness should remain at a secure level at the end of coding the previous GoP. In this paper, the number of target bits for a GoP is allocated as

$$R(n, 0) = \frac{u}{F_r} \times N_{GOP} - V(n-1) \tag{22}$$

where F_r is the frame rate, u is the bandwidth, $V(n-1)$ indicates the buffer size at the end of coding the $(n-1)$th GoP, N_{GoP} is the total number of pictures in the GoP, and $R(n, m)$ represents the remaining bits for the rest pictures after coding the $(m-1)$th picture in the nth GoP, which is updated as

$$R(n, m) = R(n, m-1) - B(n, m-1) \tag{23}$$

where $B(n, m-1)$ is the number of actual output bits of the $(m-1)$th picture in the nth GoP.

TABLE II

SUMMARY OF SIMULATION PARAMETERS

	QCIF	CIF	SD and HD
Base layer mode	AVC compatible		
Intra period	−1		
Reference no.	1		
GoP size	16		
GoP no.	16		
Symbol mode	CABAC		
Resolution	176*144	352*288	(SD) 720*576 (HD) 1280*720
Frame rate	30	30	(SD) 30 (HD) 24
Target BR	64 kb/s 128 kb/s 256 kb/s 512 kb/s	256 kb/s 512 kb/s 768 kb/s 1024 kb/s	0.8 Mb/s 1.6 Mb/s 3.2 Mb/s 6.4 Mb/s

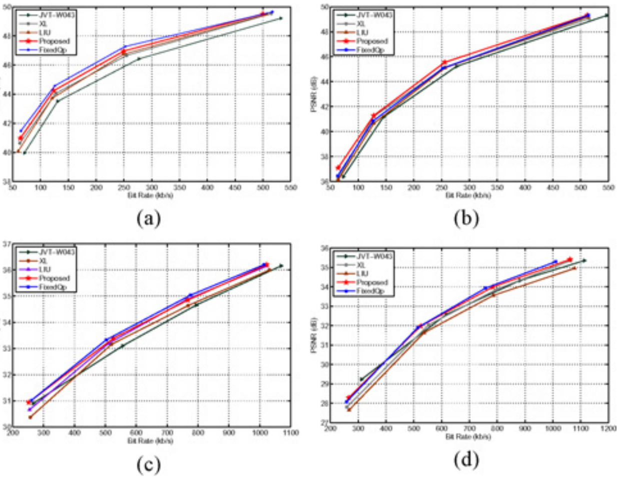


Fig. 8. R-D curves. (a) *Mother & Daughter* (QCIF). (b) *Silent* (QCIF). (c) *Coastguard* (CIF). (d) *Stefan* (CIF).

C. Frame Level RC

Suppose when encoding the mth picture in the nth GoP and this picture belongs to level i, there are N_i pictures left in level i of the nth GoP. The number of target texture bits T_m allocated to the picture m can be calculated based on the weighting factor in (21) and the number of remaining bits as

$$T_m = \left(R(n, m) - \sum_{k=0}^{N} N_k C_k \right) \times \frac{\omega_i}{\sum_{k=0}^{N} N_k \omega_k} \tag{24}$$

where C_k is the number of predicted header bits for level k. In H.264/AVC and its SVC extension, header bits take a larger part of total bits especially in very low BR. Several works such as [10] and [23] have proposed models to predict the header bits. In this paper, we employ the prediction scheme in [10] due to its mathematical simplicity. Other effective prediction methods of header bits are worthwhile studying in the future for potentially improving the proposed RC algorithm. Given the number of allocated target bits in (24) and the adopted R-Q model in (10), we can obtain the Q_{step} value Q_{step}^m for the picture m, and then the Q_p value Q_p^m according to the nonlinear relation between Q_{step} and Q_p.

As for the pictures in the highest level, the calculated Q_p values maybe vibrate severely. Considering that the pictures

TABLE III
RESULT SUMMARY OF FIVE RC ALGORITHMS ON QCIF SEQUENCES

		JVT-W043			XL			LIU			Proposed			FixedQp			
Sequence	R_t	R_o	PSNR	E	R_o	PSNR	E	R_o	PSNR	E	R_o	PSNR	E	R_o	PSNR	Iter.	E
	(kb/s)	(kb/s)	(dB)	(%)	(kb/s)	(dB)	(%)	(kb/s)	(dB)	(%)	(kb/s)	(dB)	(%)	(kb/s)	(dB)		(%)
Foreman (QCIF)	**64**	69	34.30	8.5	68	34.56	5.8	64	34.01	0.6	67	34.62	5.4	64	34.48	5	0.3
	128	136	37.64	6.6	133	37.69	4.1	128	37.34	0.2	129	37.71	0.7	125	37.59	10	2.2
	256	270	40.44	5.4	265	40.52	3.5	257	40.43	0.4	257	40.55	0.5	259	40.72	5	1.1
	512	552	43.94	7.9	518	43.43	1.2	513	43.69	0.2	514	43.65	0.3	507	43.74	5	1.0
News (QCIF)	**64**	73	38.11	13.8	65	38.09	1.2	64	37.80	0.1	64	38.68	0.3	64	38.39	4	0.6
	128	140	42.95	9.7	130	42.93	1.8	127	42.28	0.7	127	43.19	0.4	127	42.87	4	0.5
	256	381	48.42	48.8	257	46.91	0.4	255	46.73	0.5	255	47.19	0.3	252	47.03	6	1.6
	512	550	51.44	7.5	512	51.44	0.0	510	51.06	0.4	508	51.33	0.7	516	51.44	5	0.7
Paris (QCIF)	**64**	67	31.55	5.0	64	31.48	0.4	63	30.05	1.3	63	31.61	1.7	65	31.74	7	0.9
	128	139	36.08	8.3	128	35.82	0.2	126	35.07	1.3	127	35.94	1.0	127	35.8	4	0.6
	256	266	40.61	3.8	256	40.63	0.1	254	39.91	0.9	252	40.46	1.4	261	40.78	4	2.0
	512	527	45.68	3.0	512	45.17	0.1	508	45.24	0.7	507	45.17	0.9	517	45.6	7	1.0
Grandma (QCIF)	**64**	84	40.02	30.8	64	40.63	0.1	60	40.39	5.5	60	41.05	5.9	65	42.07	5	1.3
	128	148	43.90	15.4	128	44.27	0.0	124	43.89	3.2	128	44.54	0.3	130	44.87	5	1.6
	256	470	47.30	83.5	256	46.37	0.0	250	46.49	2.3	255	46.70	0.4	258	47.03	3	0.7
	512	529	48.55	3.4	516	48.47	0.8	510	48.58	0.4	510	48.75	0.4	503	48.9	4	1.8
Average				**16.3**			**1.2**			**1.2**			**1.3**				**1.1**

TABLE IV
RESULT SUMMARY OF FIVE RC ALGORITHMS ON CIF SEQUENCES

		JVT-W043			XL			LIU			Proposed			FixedQp			
Sequence	R_t	R_o	PSNR	E	R_o	PSNR	E	R_o	PSNR	E	R_o	PSNR	E	R_o	PSNR	Iter.	E
	(kb/s)	(kb/s)	(dB)	(%)	(kb/s)	(dB)	(%)	(kb/s)	(dB)	(%)	(kb/s)	(dB)	(%)	(kb/s)	(dB)		(%)
Akiyo (CIF)	**256**	310	42.92	21.1	256	44.75	0.1	254	44.58	0.6	255	45.28	0.2	259	46.14	5	1.0
	512	315	42.95	38.5	512	46.82	0.1	511	47.08	0.1	511	47.61	0.3	515	48.71	5	0.6
	768	1013	48.81	32.0	767	49.25	0.2	762	49.21	0.8	766	49.70	0.2	757	50.18	3	1.4
	1024	1143	49.08	11.6	1024	49.94	0.0	1022	49.95	0.2	1021	50.97	0.3	1018	51.50	3	0.6
Coastguard (CIF)	**256**	266	30.90	3.7	257	30.37	0.3	255	30.66	0.5	252	30.94	1.5	261	31.02	2	1.9
	512	556	33.10	8.6	520	33.16	1.6	510	33.19	0.3	526	33.39	2.7	503	33.33	1	1.7
	768	794	34.66	3.4	768	34.64	0.1	765	34.88	0.3	767	34.85	0.2	776	35.04	5	1.0
	1024	1069	36.16	4.4	1031	36.00	0.7	1023	36.15	0.1	1022	36.20	0.1	1013	36.20	4	1.1
Highway (CIF)	**256**	303	38.19	18.5	268	38.47	4.5	257	38.73	0.3	257	38.71	0.4	250	38.98	10	2.3
	512	730	39.15	42.6	512	39.35	0.0	512	39.43	0.1	513	39.44	0.2	503	39.99	4	1.7
	768	790	40.17	2.8	789	40.11	2.8	768	40.25	0.0	768	40.20	0.0	779	40.57	6	1.4
	1024	1198	40.83	17.0	1024	40.64	0.0	1023	40.66	0.1	1024	40.76	0.0	1036	41.00	3	1.2
Soccer (CIF)	**256**	273	32.58	6.6	268	32.43	4.9	271	32.35	5.7	259	32.37	1.1	259	32.44	3	1.3
	512	685	36.26	33.8	551	35.74	7.5	538	35.54	5.0	515	35.74	0.5	522	35.90	5	1.9
	768	790	37.80	2.8	803	37.92	4.5	797	37.75	3.8	769	37.89	0.2	756	37.79	3	1.6
	1024	1079	39.27	5.4	1065	39.59	4.0	1049	39.19	2.5	1173	39.43	14.5	1009	39.30	5	1.5
Average				**15.8**			**2.0**			**1.3**			**1.4**				**1.4**

in the highest level will not be referenced by other pictures and in order to maintain the smoothness of video quality in the highest level, a simple scheme is employed for deriving their Q_p values, i.e., the Q_p value of the pictures in the highest layer is set as

$$Q_p^N = \hat{Q}_p^{N-1} + 2 \tag{25}$$

where $\hat{Q}_p^{N-1}$ is the average Q_p value in level $N-1$.

V. Experimental Results

A. Simulation Setup

The proposed algorithm is implemented in the SVC reference software JSVM 9.17 [18]. In order to evaluate the RC performance of the proposed algorithm, various benchmark video sequences in QCIF, CIF, standard definition (SD) and high definition (HD) are tested. Except the first picture of the sequences is coded as I frame, all the other key pictures are coded as P frames. The typical simulation parameters are detailed in Table II.

The RC algorithms [13] (denoted as XL), [12] (denoted as LIU), JVT-W043 [17], and the FixedQp tool are utilized for comparison with the proposed algorithm. The FixedQp tool is a multiple-pass RC tool implemented in the JSVM softwares, where a logarithmic search algorithm is applied to find a proper Q_p. Different fixed Q_p values are tried to encode a sequence until the BR falls into an acceptable mismatch range or the encoding iteration exceeds the maximal number of iterations (in our experiments, the maximum negative and positive mismatch are set to 2% and the number of maximal iterations is set to 10). Although the FixedQp tool is unsuitable for real-time applications, its performance is still presented herein to evaluate the performance of the proposed algorithm.

TABLE V

Result Summary of Five RC Algorithms on SD and HD Sequences

		JVT-W043			XL			LIU			Proposed			FixedQp			
Sequence	R_t (Mb/s)	R_o (Mb/s)	PSNR (dB)	E (%)	R_o (Mb/s)	PSNR (dB)	E (%)	R_o (Mb/s)	PSNR (dB)	E (%)	R_o (Mb/s)	PSNR (dB)	E (%)	R_o (Mb/s)	PSNR (dB)	Iter.	E (%)
Autumn	**0.8**	1.02	23.81	28.0	0.72	22.9	9.8	0.87	23.24	9.2	0.79	23.95	1.5	0.81	24.43	3	1.4
(SD)	**1.6**	1.64	25.08	2.6	1.61	25.55	0.3	1.58	25.96	1.5	1.61	25.92	0.8	1.59	26.15	4	0.6
	3.2	3.26	27.20	1.7	3.25	27.97	1.5	3.20	27.66	0.0	3.32	28.04	3.9	3.23	28.22	3	1.1
	6.4	7.00	30.08	9.4	6.42	30.21	0.3	6.41	30.15	0.1	6.40	30.59	0.1	6.35	30.63	3	0.9
Crowds	**0.8**	0.85	28.77	5.8	0.83	28.85	3.9	0.82	28.62	2.4	0.80	28.61	0.1	0.79	28.85	3	0.7
(SD)	**1.6**	1.69	31.23	5.7	1.65	31.05	3.1	1.65	31.02	3.1	1.63	30.95	1.8	1.63	31.32	4	1.9
	3.2	3.74	33.67	17.0	3.43	33.54	7.1	3.34	33.30	4.5	3.21	33.45	0.3	3.21	33.73	4	0.4
	6.4	8.63	36.80	34.8	6.87	36.21	7.3	6.79	36.12	6.1	6.46	36.35	1.0	6.36	36.41	4	0.6
Stockholm	**0.8**	0.82	34.18	2.0	0.82	34.24	2.9	0.79	33.94	1.1	0.82	34.26	2.9	0.79	34.50	3	0.7
(HD)	**1.6**	1.75	34.49	9.2	1.66	34.85	3.6	1.63	34.68	1.7	1.63	35.11	2.1	1.61	35.26	3	0.8
	3.2	5.56	35.68	73.7	3.30	35.48	3.0	3.14	35.40	1.7	3.43	35.81	7.3	3.21	35.81	4	0.4
	6.4	8.81	36.41	37.6	6.59	36.31	3.0	6.40	36.08	0.0	6.20	36.60	3.1	6.52	36.74	3	1.9
Harbor	**0.8**	0.83	30.75	3.5	0.81	30.75	1.8	0.80	30.77	0.5	0.80	30.62	0.1	0.79	30.70	4	1.8
(HD)	**1.6**	1.68	32.79	5.2	1.61	32.82	0.8	1.58	32.81	1.2	1.60	33.00	0.1	1.62	33.11	5	1.0
	3.2	3.45	34.62	7.8	3.20	34.73	0.1	3.20	35.08	0.0	3.20	35.04	0.0	3.19	35.21	2	0.3
	6.4	8.60	37.14	34.4	6.75	36.96	5.4	6.39	37.16	0.2	6.35	37.38	0.8	6.52	37.57	4	1.9
Spincalendar	**0.8**	0.82	34.35	2.2	0.80	34.26	0.2	0.81	33.42	1.8	0.83	34.50	3.2	0.75	34.42	10	6.6
(HD)	**1.6**	1.64	35.04	2.7	1.67	35.54	4.6	1.62	35.26	1.4	1.60	35.83	0.1	1.64	36.02	10	2.7
	3.2	4.63	36.18	44.6	3.23	36.11	0.9	3.20	36.15	0.0	3.20	36.75	0.1	3.21	36.91	5	0.4
	6.4	11.01	37.60	72.0	6.46	37.22	0.9	6.38	37.09	0.3	6.41	37.71	0.1	6.41	37.92	4	0.1
Average				**20.0**			**3.0**			**1.8**			**1.5**				**1.3**

B. Accuracy of BR Achievement

The mismatch between the target BR and the actual output BR is studied. In order to evaluate the accuracy of BR achievement, the following measurement is used as

$$E = \frac{|R_t - R_o|}{R_t} \times 100\% \tag{26}$$

where R_t and R_o are the target BR and actual output BR, respectively. The results of BR mismatch E (%) for the comparative algorithms are presented in Tables III–V.

As shown in Tables III–V, the mismatch of JVT-W043 is quite large that it is more than 15% in average. As we know, JVT-W043 is the reorganization of the classic H.264/AVC RC algorithm JVT-G012 [10]. Although it has been implemented in JSVM reference softwares and begins to support the HBP structure, JVT-W043 is still not very effective for RC when the GoP size is larger than 4, e.g., GoP size is equal to 8 or 16. The main reason is that most frames in the HBP structure are coded as B frames, while mainly the coding characteristics of P frames are considered for designing JVT-W043. Regarding the FixeQp tool, the BR accuracy mainly relies on the preset parameters such as the maximum mismatch and the number of maximum iterations. Usually, a configuration of less mismatch may result in more iterations and thus more encoding computations. As far as the other three comparative RC algorithms are concerned, including XL, LIU, and the proposed algorithm, similar BR mismatch results are obtained as shown in the tables, within the range from 1.2% to 3.0% in average.

C. R-D Performance

As given in Tables III–V, the R-D performances in terms of PSNR and BR results are presented for comparison, where different target BRs are set for various sequences. Since the actual output BRs of different algorithms are not matched exactly, the BDPSNR (dB) and BDBR (%) [33] are employed in our experiments for fair comparison. The performances of JVT-W043 are set as the benchmark among the five algorithms, and the performances of XL, LIU, the proposed algorithm, and the FixedQp tool are compared with JVT-W043 in terms of BDPSNR and BDBR. The results are summarized in Table VI according to the results in Tables III–V.

In Table VI, the positive BDPSNR result or the negative BDBR result obtained by the four algorithms, including XL, LIU, the proposed algorithm, and the FixedQp tool, indicate the corresponding algorithm achieves better R-D performance than JVT-W043. Among the comparative RC algorithms, the FixedQp tool achieves the best performance, in average, about 1.20 dB better than JVT-W043. The better performance of the FixedQp tool mainly comes from the cascading Q_p setting in the hierarchical structure and constant Q_p values for different temporal layers throughout the sequences. In general, a strategy of applying stable Q_p values is able to achieve better R-D performances than the one that uses disturbing Q_p values. However, the usage of constant Q_p values may not be adaptive to varying contents of video sequences, thus leading to buffer overflow or underflow. Regarding the other three algorithms, including XL, LIU, and the proposed algorithm, the Q_p values are calculated based on the status of output bits and thus they can maintain relatively stable output BRs. As observed from the results, the proposed algorithm obtains 1.01 dB gain in the average BDPSNR as compared to JVT-W043, and is better than the other two algorithms XL and LIU.

For further illustration, four typical R-D curves are shown in Fig. 8, where it can be seen that the proposed algorithm is comparable or a little worse in R-D performances as compared to the FixedQp tool and better than the other three algorithms. Due to the space limit, the frame-to-frame PSNR fluctuation curves about one test scenario (*Akiyo* in CIF at 256 kb/s) are presented in Fig. 9 to demonstrate the comparison between the proposed algorithm and the other four algorithms. From the

TABLE VI
RESULTS ON BDPSNR AND BDBR

Format	Sequence	XL		LIU		Proposed		FixedQp	
		BDPSNR (dB)	BDBR (%)	BDPSNR (dB)	BDBR (%)	BDPSNR (dB)	BDBR (%)	BDPSNR (dB)	BDBR (%)
QCIF	*Foreman*	0.14	−3.02	0.22	−4.68	0.30	−6.48	0.40	−8.43
	News	0.70	−9.51	0.65	−8.76	1.03	−14.32	0.87	−11.87
	Paris	0.18	−2.57	−0.06	1.06	0.27	−3.84	0.28	−4.07
	Grandma	1.60	−20.90	1.66	−17.29	2.09	−27.62	2.26	−33.59
CIF	*Akiyo*	1.56	−45.81	1.71	−43.42	2.39	−54.13	3.30	−60.89
	Coastguard	0.14	−1.95	0.29	−6.71	0.41	−9.75	0.49	−11.53
	Highway	1.91	−23.77	1.78	−35.60	1.83	−34.01	2.12	−45.90
	Soccer	1.54	−10.29	1.41	−8.07	1.58	−12.74	1.80	−15.81
SD	*Autumn*	0.52	−14.83	0.59	−16.97	0.84	−22.63	1.09	−29.42
	Crowds	0.11	−3.22	0.03	−0.71	0.23	−5.59	0.50	−12.69
HD	*Stockholm*	0.37	−32.74	0.21	−17.61	0.66	−46.56	0.77	−51.82
	Harbor	0.28	−9.13	0.50	−15.24	0.60	−16.86	0.71	−20.48
	Spincalendar	0.50	−29.34	0.28	−7.48	0.90	−46.71	1.06	−52.03
Average		**0.73**	**−15.93**	**0.71**	**−13.96**	**1.01**	**−23.17**	**1.20**	**−27.58**

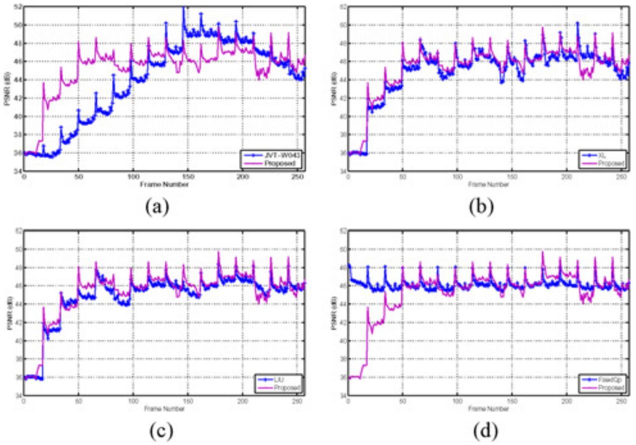


Fig. 9. Comparison of PSNR fluctuation of the proposed algorithm versus the other four algorithms with the target BR 256 kb/s on the sequence *Akiyo* (CIF). (a) Proposed versus JVT-W043. (b) Proposed versus XL. (c) Proposed versus LIU. (d) Proposed versus FixedQp.

results, it can be seen that the FixedQp tool can achieve the most stable quality among the five algorithms, since constant Q_p values are used by the FixedQp tool to encode frames in each temporal layer. Regarding the proposed RC algorithm, it can achieve better quality results than JVT-W043, XL, and LIU.

VI. Conclusion

In this paper, a linear quality dependency model was introduced between a picture and its references. By investigating the HBP structure, the video quality effects of individual pictures in different temporal layers on the entire sequence are studied. Then, effective linear R-Q and D-Q models were proposed for modeling the R-D characteristics of multiple temporal layers. In order to achieve better R-D performances for SVC, adaptive weighting factors are developed for efficient bit allocation among different temporal layers. The proposed algorithm was implemented in the SVC reference software JSVM 9.17. The experimental results demonstrated that the proposed algorithm can achieve better R-D performances than the one-pass RC algorithms JVT-W043 [17], XL [13], and LIU [12], and is comparable to or a little worse than the multiple-pass RC FixedQp tool.

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] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, *Joint Draft 11: Scalable Video Coding*, document JVT-X201, Joint Video Team, Geneva, Switzerland, Jul. 2007.

[3] T. Schierl, M. R. Civanlar, and O. Shapiro, "Multipoint videoconferencing with scalable video coding," *J. Zhejiang Univ. Sci. A*, vol. 7, no. 5, pp. 696–705, May 2006.

[4] T. Schierl, T. Stockhammer, and T. Wiegand, "Mobile video transmission using scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1204–1217, Sep. 2007.

[5] M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon, "Realtime system for adaptive video streaming based on SVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1227–1237, Sep. 2007.

[6] *Test Model 5* [Online]. Available: http://www.mpeg.org/MPEG/MSSG/tm5

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

[8] W. Li, J. R. Ohm, M. V. D. Schaar, H. Jiang, and S. Li, *MPEG-4 Video Verification Model Version 18.0*, document N3908, ISO/IEC JTC1/SC29/WG11, Jan. 2001, ch. 15, Appendix I: Rate Control, pp. 299–311.

[9] A. Leontaris and A. M. Tourapis, *Rate Control Reorganization in the Joint Model (JM) Reference Software*, document JVT-W042, 23rd JVT Meeting, San Jose, CA, Apr. 2007.

[10] Z. Li, F. Pan, K. P. Lim, G. Feng, X. Lin, and S. Rahardja, *Adaptive Basic Unit Layer Rate Control for JVT*, document JVT-G012-r1, 7th JVT meeting, Pattaya, Thailand, Mar. 2003.

[11] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of hierarchical B-pictures and MCTF," in *Proc. IEEE ICME*, Jul. 2006, pp. 1929–1932.

[12] 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.

[13] L. Xu, W. Gao, X. Ji, and D. Zhao, "Rate control for hierarchical B-picture coding with scaling-factors," in *Proc. IEEE ISCAS*, May 2007, pp. 49–52.

[14] M. Li, Y. Chang, F. Yang, S. Wan, S. Lin, and L. Xiong, "Frame layer rate control for H.264/AVC with hierarchical B-frames," *Signal Process.: Image Commun.*, vol. 24, no. 3, pp. 177–199, Mar. 2009.

[15] Y. Cho, C.-C. J. Kuo, and D.-K. Kwon, "GOP-based rate control for H.264/SVC with hierarchical B-pictures," in *Proc. IEEE IIHMSP*, Nov. 2007, pp. 387–390.

[16] Y. Cho, J. Liu, D.-K. Kwon, and C.-C. J. Kuo, "H.264/SVC temporal bit allocation with dependent distortion model," in *Proc. IEEE ICASSP*, Apr. 2009, pp. 641–644.

[17] A. Leontaris and A. M. Tourapis, *Rate Control for the Joint Scalable Video Model (JSVM)*, document JVT-W043, Joint Video Team, San Jose, CA, Apr. 2007.

[18] *Joint Scalable Video Model JSVM 9.17 Software Package*, CVS Server for the JSVM Software, Mar. 2009.

[19] L.-J. Lin and A. Ortega, "Bit-rate control using piecewise approximated rate-distortion characteristics," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 4, pp. 446–459, Aug. 1998.

[20] 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.

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

[22] 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.

[23] D. Kwon, M. 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.

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

[25] 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. 1999.

[26] 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.

[27] A. Vetro, H. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 186–199, Feb. 1999.

[28] Z. G. Li, F. Pan, K. P. Lim, and S. Rahardja, "Adaptive rate control for H.264," in *Proc. IEEE ICIP*, Oct. 2004, pp. 745–748.

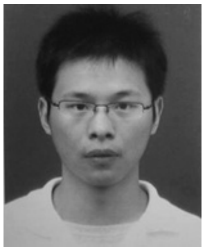
[29] T. Berger, *Rate-Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[30] K. Takagi, Y. Takishima, and Y. Nakajima, "A study on rate distortion optimization scheme for JVT coder," *Proc. SPIE*, vol. 5150, pp. 914–923, Jul. 2003.

[31] H. Wang and S. Kwong, "A rate-distortion optimization algorithm for rate control in H.264," in *Proc. IEEE ICASSP*, Apr. 2007, pp. 1149–1152.

[32] N. Kamaci, Y. Altinbasak, 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.

[33] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33, VCEG Meeting ITU-T SG16/Q6, Austin, TX, Apr. 2001.



Sudeng Hu received the B.E. degree from Zhejiang University, Hangzhou, China, in 2007, and the M.Phil. degree from the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong, in 2010.

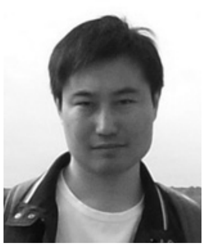
His current research interests include the fields of image and video compression, rate control, and scalable video coding.

Hanli Wang (M'08) received the B.S. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2001 and 2004, respectively, and the Ph.D. degree in computer science from the City University of Hong Kong, Kowloon, Hong Kong, in 2007.

From 2007 to 2008, he was a Research Fellow with the Department of Computer Science, City University of Hong Kong. From 2007 to 2008, he was a Visiting Scholar with Stanford University, Palo Alto, CA, invited by Prof. C. K. Chui. From 2008 to 2009, he was a Research Engineer with Precoad, Inc., Menlo Park, CA. From 2009 to 2010, he was an Alexander von Humboldt Research Fellow with the University of Hagen, Hagen, Germany. In 2010, he joined the Department of Computer Science and Technology, Tongji University, Shanghai, China, as a Professor. His current research interests include digital video coding, image processing, pattern recognition, and video analysis.

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

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



Tiesong Zhao (S'08) received the B.S.E.E. degree from the University of Science and Technology of China, Hefei, China, in 2006. He is currently pursuing the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong.

His current research interests include digital image and video coding, and fast algorithms.

C.-C. Jay Kuo (F'99) received the B.S. degree from the National Taiwan University, Taipei, Taiwan, in 1980, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1985 and 1987, respectively, all in electrical engineering.

He is currently the Director of the Signal and Image Processing Institute and a Professor of electrical engineering, computer science, and mathematics with the Department of Electrical Engineering and Integrated Media Systems Center, University of Southern California, Los Angeles. His current research interests include digital image/video analysis and modeling, multimedia data compression, communication and networking, and biological signal/image processing. He is the co-author of about 190 journal papers, 810 conference papers, and 10 books.

Dr. Kuo is a fellow of the American Association for the Advancement of Science and the International Society for Optical Engineers.