Low-Delay Rate Control for Consistent Quality Using Distortion-Based Lagrange Multiplier

Miaohui Wang, *Member, IEEE*, King Ngi Ngan, *Fellow, IEEE*, and Hongliang Li, *Senior Member, IEEE*

***Abstract*— Video quality fluctuation plays a significant role in human visual perception, and hence, many rate control approaches have been widely developed to maintain consistent quality for video communication. This paper presents a novel rate control framework based on the Lagrange multiplier in high-efficiency video coding. With the assumption of constant quality control, a new relationship between the distortion and the Lagrange multiplier is established. Based on the proposed distortion model and buffer status, we obtain a computationally feasible solution to the problem of minimizing the distortion variation across video frames at the coding tree unit level. Extensive simulation results show that our method outperforms the rate control used in HEVC Test Model (HM) by providing a more accurate rate regulation, lower video quality fluctuation, and stabler buffer fullness. The average peak signal-to-noise ratio (PSNR) and PSNR deviation improvements are about 0.37 dB and 57.14% in the low-delay (P and B) video communication, where the complexity overhead is ∼4.44%.**

***Index Terms*— Rate control, distortion model, Lagrange multiplier, HEVC.**

I. Introduction

With the rapid development of visual communication in recent years, efficient video compression techniques have been considered by the multimedia community. To meet this demand, ITU-T/SG16/Q6 (VCEG) and ISO/IEC JTC1/SC29/WG11 (MPEG) have established the Joint Collaborative Team on Video Coding (JCTVC) to develop the latest video coding standard – *High Efficiency Video Coding* (HEVC) [1]. As HEVC involves many sophisticated coding features and techniques, it has significantly improved the compression performance in comparison with the previous video coding standards, such as MPEG-2, H.263, MPEG-4 and H.264/Advanced Video Coding (AVC) [2]. For example, HEVC supports diverse block sizes, flexible quad-tree structure and efficient filters. However, it is worth noting that the above video standards only specify the syntax of decoding the bitstream. Meanwhile, the compressed video qualities depend considerably on implementing the rate control scheme on the encoder side. Thus, computational rate control algorithms are widely investigated after the international video standards have been developed.

Normally, various video applications, like video broadcasting and video surveillance, are transmitted via a constant bit rate (CBR) channel. To maintain a short-term constant output bit rate in the CBR channel, traditional codecs adopt a uniform bandwidth allocation scheme in a group of pictures (GOP). However, the number of encoding bits changes from frame to frame owing to the time-varying video complexity. In this case, it is infeasible to adjust encoding parameters to achieve the exact fixed bandwidth for each frame in a GOP. Consequently, an associate encoder buffer is usually employed to regulate the output bits before transmitting. If the channel bandwidth is less than the output bit rate, the encoded bits will accumulate in the encoder buffer. When the size of the accumulated bits is too large, the encoder needs to skip some frames to alleviate the buffer delay and avoid the buffer overflow. Conversely, if the channel bandwidth is larger than the output bit rate, this indicates that some channel is wasted, and it may cause buffer underflow. Since the buffer overflow and underflow result in undesirable effects on the video quality fluctuation, it is essential to control the bit rate to maintain a consistent quality over the entire video sequence.

In real-time video communications, rate control becomes more challenging as it needs to satisfy low-latency of transmitting video data. In such a case, the encoder buffer must maintain a very small size, and therefore, the encoder requires a more accurate bit allocation and coding settings to reduce the fluctuations of buffer fullness as well as to avoid the undesirable buffer overflow and underflow. The conventional rate control schemes usually involve two steps: (1) the target bit is allocated to each basic unit according to its relative complexity and the buffer status, and (2) quantization parameter (QP) is computed by rate-quantization (R-Q) models, such as quadratic model [3] and ρ-domain model [4]. It should be pointed out that with traditional R-Q models, QP can determine the bits for residue information (i.e., quantized transformed coefficient (QTC)) but not for non-residue information (i.e., partition mode, motion and other header information). Although the overhead bits can be predicted from the previous frames, the accuracy of prediction is still not well addressed in both H.264/AVC and HEVC

Manuscript received December 28, 2014; revised May 13, 2015, January 10, 2016, and February 22, 2016; accepted April 5, 2016. Date of publication April 11, 2016; date of current version May 11, 2016. This work was supported in part by the Research Grants Council, Hong Kong, under Grant CUHK415712, in part by the National Natural Science Foundation of China under Grant 61525102, and in part by the Program for Science and Technology Innovative Research Team for Young Scholars in Sichuan Province, China, under Grant 2014TD0006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ricardo L. De Queiroz.

M. Wang and K. N. Ngan are with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: wang.miaohui@gmail.com; knngan@ee.cuhk.edu.hk).

H. Li is with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 610051, China (e-mail: hlli@uestc.edu.cn).

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

Digital Object Identifier 10.1109/TIP.2016.2552646

1057-7149 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

owing to time-varying video complexity. On the other hand, similar to H.264/AVC, the reference software of HEVC utilizes rate distortion optimization (RDO) to determine the encoding settings so as to obtain the best rate-distortion (R-D) performance. To solve this RDO problem, the Lagrangian method is used to achieve the optimal trade-off between rate and distortion on the encoder side. Additionally, it can be seen that the Lagrange multiplier affects the total bits, including both the residue and non-residue bits. In the HEVC reference software, simulation results indicate that the Lagrange multiplier is more sensitive, simple and efficient in controlling the output target bits in comparison with QP, and not surprisingly, an interesting relationship between the target bits and Lagrange multiplier λ (also called Rate-lambda or R-λ model) was developed for HEVC in JCTVC-K0103 [5].

In the R-λ model, λ is first determined by the target bits. Then, QP is computed by a logarithmic function of λ in the HEVC reference software. Although the R-λ method has shown a significant improvement in comparison with conventional methods, there are two main difficulties.

- Inaccurate bit allocation: The number of bits for a coding tree unit (CTU) relies heavily on the frame budget, the previous overhead bits and the weights of CTUs in the current frame, which is explicitly determined one-by-one in a raster-scanning order. In practice, the frame bandwidth can be easily consumed by the first several CTUs due to inaccurate estimation of the model parameters. Hence, the buffer state greatly affects the target bits for the subsequent CTUs in a frame, and the raster-scanning bit allocation scheme can cause inaccurate bit rates and adverse effects on the overall quality control.
- Inaccurate λ estimation: As mentioned earlier, since there is the bit allocation problem, λ adjustment is frequently applied for CTUs to achieve the frame budget. Specifically, an extremely large (or small) λ is usually computed according to the target bits and buffer fullness. To maintain λ in a reasonable range, many "bound" and "crop" operations are employed on the encoder side. For example, the HEVC reference software allows the maximum λ variation in terms of QP up to 10 between two successive frames which results in a large quality fluctuation. Thus, inaccurate λ greatly degrades the final rate control result.

Although some improvements [6], [7] in the R-λ models have been reported recently, the two above-mentioned limitations are still not being addressed satisfactorily. This paper proposes a new distortion based Lagrange multiplier method to improve the HEVC rate control in low-delay communication. The major innovations of this paper can be summarized in three aspects. Firstly, with the assumption of consistent video quality coding, a new relationship between the distortion and λ is established, which can be used to control the video quality fluctuations via λ. Secondly, a computationally feasible solution to the problem of minimizing distortion across the video frames at the CTU level is obtained based on the above distortion model, which can avoid the raster-scanning bit allocation. Thirdly, according to buffer fullness, the CTU level λ is adjusted such that the target bits satisfy the overall bandwidth in low-delay video communications.

The rest of the paper is organized as follows. We introduce the related work in Section II. The new distortion model and the associated rate model are described in Section III. The proposed rate control algorithm is then implemented in Section IV. The simulation results are presented in Section V, and Section VI contains the concluding remarks.

II. Related Work in HEVC

Recently, many rate control methods have been considered in the HEVC video coding system. In this section, we give a brief review of the different rate models to facilitate the study of the rate control algorithms. In HEVC, the rate control models can be roughly categorized into R-Q and R-λ models.

A. R-Q Models

Choi *et al.* proposed a pixel-wise unified R-Q model (URQ) [8] that is a direct extension of the quadratic R-Q rate control method in the early HEVC reference software (HM6.0). The quadratic R-Q model is first developed for hybrid video coding by Chiang and Zhang [3] which has been widely studied in H.264/AVC [9]. Similar to the quadratic R-Q method in H.264/AVC, in URQ, the target bit is allocated based on the mean absolute difference (MAD) - based complexity, and the quantization step is computed on the encoder side. A detailed description of URQ can be found in their recent work [10]. Although URQ shows a better performance to control the output bits than the anchor HM6.0, two major limitations of the quadratic model are not avoided. Firstly, the quadratic model only determines the bits for the residue information but not for the non-residue information. Secondly, the quadratic model contains an inter-dependency relationship between the RDO process and the computation of quantization step, which results in a well-known "chicken and egg" dilemma [11].

Additionally, a general R-Q model is studied in HEVC, which is based on the number of QTCs. He and Mitra [4] proposed the ρ-domain rate model for simple block structure based video coding, where ρ is the percentage of zeros among QTCs. The ρ-domain model is based on the observation that there is an approximated linear relationship between bits and ρ in a single transform block size. Recently, Wang *et al.* developed a quadratic ρ-domain based Rate-GOP method [12] in HEVC. With Rate-GOP algorithm, QP can first be determined by the picture order count (POC), and then the bits of non-zero QTCs are simulated by a quadratic function of quantization step. Although simulation results have shown its superiority compared to traditional rate control methods, further investigation is necessary to verify the ρ-domain model in HEVC. Specifically, in [13] and [14], experiments show the facts that signal characteristics with respect to the coding unit depth levels are considerably different. In addition, the variances of the transform coefficients of intra frames are significantly different from that of inter frames. Thus, in the HEVC encoder, the ρ-domain methods should be modeled separately due to the depth of the coding units and the types of

encoded frames. Unfortunately, this is still not well addressed in the Rate-GOP method.

B. R-λ Models

To avoid the aforementioned problems, the Lagrange multiplier based rate control methods have been studied in H.264/AVC and HEVC. With the RDO scheme on the encoder side, larger λ is related to larger distortion and fewer bit, and the converse holds for smaller λ. Furthermore, since QP adjustments will cause additional overhead bits, λ is used to adjust the rate and distortion during the actual encoding stage. Taking into account this observation, Jiang and Ling [15], [16] proposed to adjust λ adaptively according to the rate cost. In [17], Wang and Yan established a new relationship between λ and MAD. Since the MAD has been integrated into the quadratic R-Q model in H.264/AVC, it indicates that the relationship between rate and λ has been implicitly established.

Nowadays, Li *et al.* [5] (see a detailed description in [18]) developed a new relationship between rate and λ. Compared to the R-Q method, the R-λ model considers the overall bit rate, including both the residue and non-residue bits. Additionally, as the R-λ method outperforms the URQ method in terms of bit estimation accuracy and video quality control, it has been recommended for HEVC by the JCTVC. However, it should be pointed out that the R-λ model only considers the target bit but ignores the characteristics of video data in frame level rate control. Thus, the frame-content complexity has been considered to improve the R-λ model. For example, Wang and Karczewicz [6] proposed to use summation of absolute transformed differences (SATD) to measure the complexity of an intra frame. In [7], Wang and Ngan proposed a gradient based R-lambda (GRL) model for intra frame rate control. Simulation results show that the SATD and gradient can be used to effectively measure the frame-content complexity and enhance the performance of intra-frame rate control.

III. Modeling Rate and Distortion in High Efficiency Video Coding

Rate-distortion optimization has been widely studied in block-based video coding systems, such as in the reference software of H.264/AVC and HEVC. The R-D relationship indicates that the higher the rate R is, the lower is distortion D, and vice versa. Thus, the fundamental problem in rate control is to minimize the distortion subject to a given rate constraint $R_{\max}$ [19].

$$\min D, \quad s.t.\ R \le R_{\max}, \tag{1}$$

Usually, this constrained problem can be solved by the Lagrangian optimization method [20] in hybrid video coding. The Lagrangian cost function is

$$J = D + \lambda \times (R - R_{\max}), \tag{2}$$

where λ is the Lagrange multiplier. Additionally, it is noted that the R-D curve is convex in video coding, and if we assume that both R and D are differentiable everywhere, λ can be expressed by

$$\lambda = -\frac{\partial D}{\partial R}. \tag{3}$$

As shown in equation (2), R and D for a given encoding block or frame depend on the scaling factor λ. Conventionally, λ is obtained from the input QP [21]. The relationship between R-Q and D-Q (i.e., distortion-quantization) models has been studied and used for rate control in H.263, H.264/AVC and MPEG-4 codecs. However, it is interesting to note that QP only determines the residue bits, but λ determines the overall rate cost, including both residue and non-residue bits. Furthermore, λ also affects the encoding mode selection, such as partitions and motions. Thus, in this section, the essential goal is to model the effect of λ on distortion and rate. We concentrate on modeling distortion and rate for motion-compensated prediction (MCP) frames. It should be noted that similar models can be established for intra frames, but the related models will not be derived here because the intra frames are not frequently used in low-delay video communications.

A. Distortion Modeling

In the past decades, many R-D models [22]–[24] have been proposed for rate control. For example, in H.264/AVC and MPEG-4, the R-D curves can be modeled by a logarithmic expression. Specifically, this logarithmic model has been proved assuming a high-rate environment (e.g., above 0.5 bits/pixel) [25]. However, in the HEVC reference software, since the encoding efficiency is greatly improved, the traditional R-D model is not appropriate for the simulation of the encoder in the low-rate case. In the meantime, as shown in equation (3), the more accurate the R-D model is, the better λ can be obtained. Therefore, in HEVC, a more interesting hyperbolic R-D model is proposed in equation (4), where this model was introduced by Mallat in 1998 [26].

$$D(R) = \mathrm{K} \times R^{-\mathrm{C}}, \tag{4}$$

where C and K are the model parameters related to the characteristic of the video source. Putting equation (4) into equation (3), we know that λ can be re-written as

$$\begin{aligned} \lambda &= -\frac{\partial D}{\partial R} = \mathrm{C} \times \mathrm{K} \times R^{-\mathrm{C}-1} \\ &= \mathrm{C} \times \mathrm{K}^{-\frac{1}{C}} \times D^{\frac{\mathrm{C}+1}{C}} = \gamma D^{\tau}, \end{aligned} \tag{5}$$

where γ and τ are both coding constants. The distortion measure D is the mean squared error (MSE) between the original and reconstructed CTUs or frames. MSE is a mathematically tractable and fast-to-compute full-reference (FR) quality metric, which has been widely used in modern block-based video compression.

To validate the relationship between D and λ, we have compressed several video sequences in the low-delay case using the HEVC reference software HM10.0 and the simulation results are shown in Fig. 1. For several values of λ, we plot $D(\lambda)$ as a function of λ and fit the data via the formula in (5), where the average values are used. This experiment

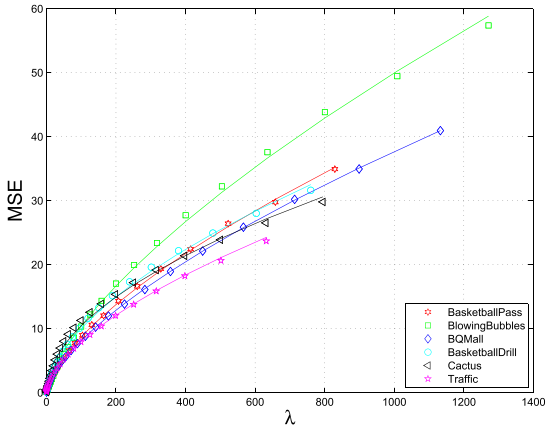

Fig. 1. Relationship between D and λ. The color curves are the fitting results of $\lambda = \gamma D^{\tau}$. Each sample point represents one test video, and the data points with the same color denote the encoding results associated to the same video sequence.

confirms that in the HEVC video coding system, the distortion curves can be approximated by a power formula.

From now on, we will focus on the consistent video quality modeling in which the more interesting $D(\lambda)$ relationship between two consecutive frames is established. The distribution of distortion between two consecutive frames is crucial in maintaining consistent quality control in video coding. For consistent quality videos, it is observed that the distortion of the current CTU is distributed similarly to that of the co-located position in the previous frame. Hence, we incorporate the history of coding information into the distortion model to avoid the complexity of parameter estimation in equation (5). In addition, to maintain a consistent quality, we assume $D_i^{curr} \approx D_i^{prev}$. Putting equation (5) into this assumption, we can get $\frac{D_i^{curr}}{\lambda_i^{curr}} \approx \frac{D_i^{prev}}{\lambda_i^{prev}}$, where D_i^{curr} and λ_i^{curr} are the distortion and the Lagrange multiplier for the ith CTU in the current frame. Similarly, D_i^{prev} and λ_i^{prev} are for the co-located CTU in the previous frame. With this assumption, we can construct a heuristic linear distortion model $D_i^{curr} \approx \frac{D_i^{prev}}{\lambda_i^{prev}} \times \lambda_i^{curr}$.

In order to remove large fluctuation during encoding, both λ_i^{prev} and D_i^{prev} can be computed as the weighted averages from the previous frames [14] such as $\lambda_i^{prev} = \frac{1}{N_l}\sum_{k=1}^{N_l} \omega_k \lambda_k^{prev}$ and $D_i^{prev} = \frac{1}{N_l}\sum_{k=1}^{N_l} \omega_k D_k^{prev}$, where N_l is the total number of the most recent encoded frames, and ω_k is the associated weight. On the other hand, from the implementation point of view, we need to reduce the computational complexity and save the storage of the encoder. Thus, we make a trade-off between model accuracy and implementation complexity, i.e., $\lambda_i^{prev} = \frac{1}{N}\sum_{k=1}^{N} \lambda_k^{prev}$, and D_i^{prev} is the actual distortion of the ith CTU in the previous frame. Consequently, a linear distortion model is completed as,

$$D_i^{curr} = \frac{\sigma}{\frac{1}{N}\sum_{k=1}^{N} \lambda_k^{prev}} \times D_i^{prev} \times \lambda_i^{curr}, \tag{6}$$

where σ is a scaling factor that reduces the speed of distortion changes between the co-located CTUs.

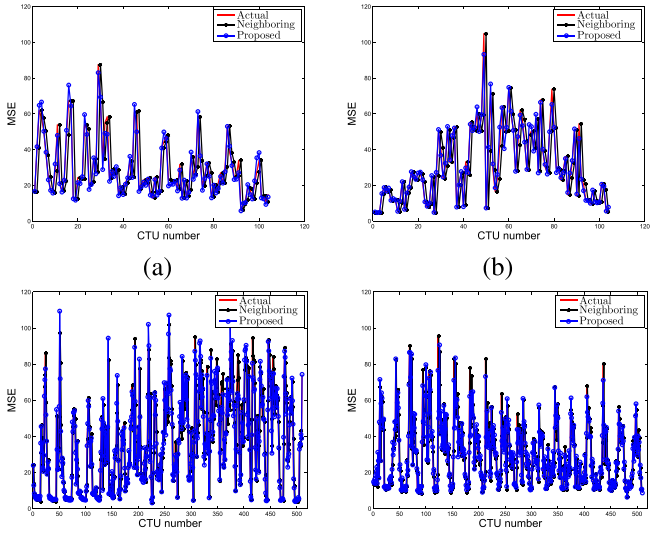

Fig. 2. Difference between the actual distortion and the estimated distortion per CTU of the *11*th frame for (a) *BasketballDrill*, (b) *BQMall*, (c) *BQTerrace* and (d) *Cactus* (QP=32).

TABLE I
COMPARISONS OF DIFFERENT DISTORTION MODELS

Sequences	Bit rate	PCC		NRMSE	
		Proposed (optimal σ)	Neighb-oring	Proposed (optimal σ)	Neighb-oring
BasketballPass	220 Kbps	0.9772	-0.1570	0.0704	0.3976
BlowingBubbles	474 Kbps	0.9176	0.2438	0.0923	0.3390
BQSquare	372 Kbps	0.9594	0.3872	0.0852	0.3068
BasketballDrill	871 Kbps	0.9126	0.4184	0.0776	0.2127
BQMall	1.33 Mbps	0.9540	0.5167	0.0671	0.2190
RaceHorses	1.27 Mbps	0.9115	0.3880	0.0933	0.2530
FourPeople	540 Kbps	0.9794	0.7236	0.0437	0.1663
Johnny	350 Kbps	0.9845	0.7706	0.0364	0.1465
BQTerrace	3.81 Mbps	0.9670	0.7915	0.0612	0.1571
Cactus	3.21 Mbps	0.9584	0.7147	0.0587	0.1537
ParkScene	1.62 Mbps	0.9599	0.6412	0.0553	0.1635
Average		0.9529	0.4944	0.0674	0.2287

Extensive experiments have been performed to verify the proposed distortion model, and four typical results are presented in Fig. 2 that shows the distortion estimation results for each CTU in low-delay video application. The encoding frame is the *11*th frame. The distortion of each CTU in the current frame is estimated by the co-located distortion in the *10*th encoded frame. The scaling factor σ is chosen to reduce the estimation error. For example, $\sigma =$ 1.30, 1.34, 1.26 and 1.33 are the optimal values (in MSE sense) for (a), (b), (c) and (d), respectively. However, in this paper, we do not explore the use of scaling factor for a better estimation. Instead, as we will see later in Section IV-A, σ is not required in the computation of λ at the CTU level. More detailed numerical experiments are tabulated in Table I. Pearson's correlation coefficient (PCC) and normalized root mean square error (NRMSE) are computed to measure the estimation accuracy and estimation error, respectively. The NRMSE is defined as in

$$NRMSE = \frac{1}{(D_{act,\max} - D_{act,\min})} \times \sqrt{\frac{\sum_{i=1}^{N} \left(D_{act,i} - D_{est,i}\right)^2}{N_{ctu}}}, \tag{7}$$

where N_{ctu} is the number of encoded CTUs in a frame, D_{act} and D_{est} are the actual distortion and estimated distortion of the ith CTU, $D_{act,\max}$ and $D_{act,\min}$ are the maximum and minimum values of D_{act}, respectively. Table I shows the comparison between the proposed method and a typical distortion model that uses the neighboring information to estimate the current one. Typically, the neighboring reconstructed error D_{i-1}^{curr} is considered as the predicted distortion of the current D_i^{curr}, i.e., $D_i^{curr} = D_{i-1}^{curr}$. The PCCs of the proposed linear distortion method (6) range from 0.91 to 0.99. In addition, the average NRMSE of our method is 0.0674, while the neighboring method is 0.2287. It can be seen that our distortion model is able to achieve higher estimation accuracy than the traditional neighboring estimation method.

B. Rate Modeling

In HEVC, the rate of the ith CTU in a frame is a function of λ as shown in equation (5). For convenience, equation (5) is re-written as

$$R_i = \alpha_i \times \lambda_i^{\beta_i}, \tag{8}$$

where R_i is the target budget of a CTU, and α_i and β_i are the corresponding model parameters.

Extensive simulation results in JCTVC-K0103 have shown that $R(\lambda)$ in equation (8) is sufficient to represent the relationship between R and λ on the HEVC reference software. In HEVC, given a target budget R_i for a CTU, the associated λ_i is computed using equation (8). Consequently, the quantization parameter used for quantizing the transform coefficients is computed by a natural logarithmic formula

$$QP_i = a \times ln(\lambda_i) + b, \tag{9}$$

where QP_i is the quantization parameter for the ith CTU, and both a and b are constants, which are empirically set as 4.2 and 13.71 in the reference software of HEVC HM10.0, respectively.

IV. Rate Control With Distortion-Based Lagrange Multiplier

A. Proposed Distortion-Based Lagrange Multiplier

Quality variation in a compressed video signal has an essential impact on the human visual perception [27]–[30], so the goal of the consistent quality control is to minimize the distortion variation across video frames within the constraints of frame rate, bandwidth, and delay requirement. Based on the aforementioned distortion and rate models, we investigate the problem of optimal bit allocation from the viewpoint of minimizing average (MINAVE) distortion [31] for the HEVC encoder system. Specifically, we want to find an expression for the Lagrange multipliers that minimizes the average distortion $\frac{1}{N}\sum_{k=1}^{N} D_k$ subject to the summation of the target budgets $\sum_{k=1}^{N} R_k$ which is not larger than the frame target budget $R_{\max}$. In addition, replacing D_i with the proposed distortion model in equation (6) and R_i with equation (8), respectively, the constrained MINAVE problem can be formulated as:

$$\begin{aligned}\lambda_1^*, \ldots, \lambda_N^* &= \underset{\substack{\lambda_1^{curr},\lambda_2^{curr},\ldots,\lambda_N^{curr},\\ \sum_{k=1}^{N} R_k\left(\lambda_k^{curr}\right)\le R_{\max}}}{\arg\min} \frac{1}{N}\sum_{k=1}^{N} D_k^{curr}\left(\lambda_k^{curr}\right) \\ &= \underset{\substack{\lambda_1^{curr},\lambda_2^{curr},\ldots,\lambda_N^{curr},\\ \sum_{k=1}^{N} R_k\left(\lambda_k^{curr}\right)\le R_{\max}}}{\arg\min} \sum_{k=1}^{N} \frac{\sigma}{\lambda_k^{prev}} \times D_k^{prev} \times \lambda_k^{curr},\end{aligned} \tag{10}$$

where λ_i^* is the optimal value of λ_i^{curr}.

Equation (10) is the key formulation for consistent quality control. In equation (10), the average distortion and the target budget constraint are convex function and convex set of the variable λ_i^{curr}, respectively. According to the Lagrange theory [32], there is a unique solution $\lambda^* = [\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*]$ that can be obtained from equation (10). One way to solve it is to use the Karush-Tuhn-Tucker (KKT) condition. Unfortunately, it is difficult to obtain a closed-form solution of equation (10) from its Lagrangian cost function, since the complex rate model is involved. Another way is to search the possible set of λ^* that satisfies the KKT condition of equation (10). One should find if there is a certain λ^* that leads to a solution satisfying the KKT condition. This exhaustive search method is infeasible for practical low-delay video applications. Clearly, a more efficient solution is needed.

Based on the observation that the budget is a power function with respect to the variable λ_i^{curr}, we propose to relax the constraint $\sum_{k=1}^{N} R_k\left(\lambda_k^{curr}\right) \le R_{\max}$ to $\prod_{k=1}^{N} R_k\left(\lambda_k^{curr}\right) \le \left(\frac{R_{\max}}{N}\right)^N$ with the inequality of arithmetic and geometric means. Then, we rewrite the problem in equation (10) as in equation (11). In equation (11), we can find a closed-form solution of λ^*. However, it is noted that we should consider the buffer state and frame budget, and hence we need to adjust λ^* to guarantee that $\sum_{k=1}^{N} R_k\left(\lambda_k^{curr}\right) \le R_{\max}$ in the practical encoder HM10.0. The proposed method is formulated as follows.

$$\begin{aligned}\lambda_1^*, \ldots, \lambda_N^* = \underset{\substack{\lambda_1^{curr},\lambda_2^{curr},\ldots,\lambda_N^{curr},\\ \sum_{k=1}^{N}\ln(R_k)\le N\ln\left(\frac{R_{\max}}{N}\right)}}{\operatorname{argmin}} & \sum_{k=1}^{N} \frac{\sigma}{\lambda_k^{prev}} \\ & \times D_k^{prev} \times \lambda_k^{curr}.\end{aligned} \tag{11}$$

The Lagrangian cost function of equation (11) can be expressed in equation (12), which is an unconstrained problem. In addition, if R_i is replaced with equation (8), then we have

$$\begin{aligned}&\underset{\lambda_i^{curr}>0, i=1,\ldots,N, u\ge 0}{\operatorname{argmin}} L\left(\lambda_1^{curr}, \lambda_2^{curr}, \ldots, \lambda_N^{curr}, u\right) \\ &= \underset{\lambda_i^{curr}>0, i=1,\ldots,N, u\ge 0}{\operatorname{argmin}} \sum_{k=1}^{N} \frac{\sigma}{\lambda_k^{prev}} \times D_k^{prev} \times \lambda_k^{curr} \\ &\quad + u\left(\sum_{k=1}^{N} \ln\left(R_k\right) - N\ln\left(\frac{R_{\max}}{N}\right)\right)\end{aligned}$$

$$= \underset{\substack{\lambda_i^{curr}>0,\alpha_i>0,\beta_i<0,\\ i=1,\ldots,N,u\geq 0}}{\arg\min} \sum_{k=1}^{N} \frac{\sigma}{\lambda_k^{prev}} \times D_k^{prev} \times \lambda_k^{curr} + u\left(\sum_{k=1}^{N}\left(\ln(\alpha_k) + \beta_k \ln\left(\lambda_k^{curr}\right)\right) - N\ln\left(\frac{R_{\max}}{N}\right)\right), \tag{12}$$

where u is the Lagrange multiplier. In Lagrange's theory, it has been shown that if there is a u^* such that equation (12) achieves the minimum value at $\lambda^* = [\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*]$, then λ^* is also an optimal solution to equation (11).

In addition, since equation (12) is minimizing a convex and differentiable function on a convex set, the KKT condition guarantees that the KKT point λ^* is an optimum solution. Consequently, after some straightforward manipulations (see Appendix), we obtain the optimal λ^* in equation (13).

$$\lambda_i^* = \frac{-\beta_i}{D_i^{prev}} \times e^{\frac{N\ln\left(\frac{R_{\max}}{N}\right) - \sum_{k=1}^{N}\ln\alpha_k\left(\frac{-\beta_k}{D_k^{prev}}\right)^{\beta_k}}{\sum_{k=1}^{N}\beta_k}}. \tag{13}$$

As mentioned before, the summation of the CTU bit budgets $R\left(\lambda_i^*\right)$ in equation (11) should satisfy the frame budget constraint. Thus, we can adjust λ_i^* to a proper $\lambda_{CTU,i}^*$ such that $\sum_{k=1}^{N} R_{CTU,k}\left(\lambda_{CTU,k}^*\right) \leq R_{\max}$. In general, adjacent frames in a video sequence have very high correlations that guarantee that we can predict the associated model parameters for the current frame from the previous encoded one. Furthermore, equal bit allocation scheme is used for each frame in the proposed method. To maintain a proper bit budget for the current frame, we use a scaling factor χ of the previous frame to update λ_i^* in equation (14). The scaling factor χ is obtained from the actual budget R_{act}^{prev} and target budget R_{tar}^{prev} of the previous frame.

$$\chi = R_{tar}^{prev} / R_{act}^{prev}. \tag{14}$$

To satisfy the frame budget, we adjust the CTU level target bit budget as $R_{CTU,i} = \chi R_i$. As a result, when replacing $R_{CTU,i}$ and R_i with equation (8), we can obtain the CTU level $\lambda_{CUT,i}^*$ as

$$\lambda_{CUT,i}^* = \chi^{\frac{1}{\beta_i}} \lambda_i^*, \tag{15}$$

where $\lambda_{CUT,i}^*$ is used to encode the current CTU.

We conducted experiments to evaluate the proposed method (15) in low-delay video communications. The standard video sequences with different resolutions and frame rates are encoded. Tables V and VI summarize the simulation results of the actual bit rates and the target ones. Experimental results show that the proposed method can achieve a higher average bits estimation accuracy.

B. Proposed Rate Control Algorithm

The traditional three-level bit allocation scheme has been very successful in the H.264/AVC rate control. As a result, JCTVC-K0103 and other rate control proposals in the HEVC codec follow the same structure, such as GOP level, frame level and CTU level. Inspired by these pioneering works, we adopted the same approach of JCTVC-K0103 with the detailed algorithm as described as Algorithm 1. To easily understand the following steps, the reader is referred to the JCTVC-K0103 implementation in the HM10.0 software. The constant exponents (e.g., in Step 2.2 and 3.3) are empirically set in our paper. Symbols used in the proposed method are tabulated in Table II.

TABLE II
SYMBOLS USED IN THE DISTORTION-BASED LAGRANGE MULTIPLIER ALGORITHM

Symbol	Description
λ_{curF}	Frame level Lagrange multiplier
α, β	Frame level Rate model parameters
$R_{bpp,tar}$, $R_{bpp,real}$	Frame level target and actual bits, respectively
e	Euler's number
a, b	Scaling factor used to compute the quantization parameter
QP_{curF}	Frame level quantization parameter
N	The number of CTUs in a frame
$QP_{ctu,k}$	Quantization parameter of the kth CTU
$\lambda_{ctu,k}^*$	Lagrange multiplier of the kth CTU
$\alpha_{ctu,k}^*$, $\beta_{ctu,k}^*$	Proposed CTU level model parameters of the kth CTU
$D_{ctu,k}^*$	Distortion of the kth CTU
M	Size of the current CTU.
$R_{ctu,bpp,real,k}$	Actual bits of the kth CTU
$\lambda_{ctu,real,k}$	Actual Lagrange multiplier of the kth CTU
$\lambda_{curF,real}$	Actual Lagrange multiplier of the current frame
χ	Model parameter used to adjust the output bits

V. SIMULATION RESULTS

To evaluate the performance of the proposed consistent quality control approach, we compare it with the state-of-the-art techniques on HM10.0 platform [33]. The experiments are conducted on a dual-core (i3-2100@3.10G Hz) workstation with RAM 4GB that is also used to measure the computational complexity of our method. In the simulation, the rate control based encoder parameters are set as follows: RateControl (enabled), NumLCUInUnit (enabled), LCULevelRateControl (enabled), RCLCUSeparateModel (enabled), InitialQP (disabled), KeepHierarchicalBit (disabled) and RCForceIntraQP (disabled). All the other encoder settings are set identically for all methods. We have performed the following four methods.

- HM10.0: JCTVC-K0103 rate control algorithm has been implemented and enabled for all tests in this section.
- Choi *et al.* [10]: Choi's method [10] (i.e., JCTVC-H0213) has been implemented and compared in HM10.0.
- Lee *et al.* [14]: a frame-level rate control method is implemented and tested in HM10.0.
- Proposed method: the proposed distortion-based Lagrange multiplier method is implemented in HM10.0.

In the experiment, the first two frames (i.e., the first intra and inter frame) are encoded by the HM10.0 scheme, which are used to collect the corresponding model parameters. The proposed method was evaluated with two bandwidths (i.e., low and high bit rate) under the low-delay configurations (i.e., P and B Main coding profile). In both coding structures, all the representative standard sequences with

Algorithm 1 Proposed distortion-based rate control method

Step 1: GOP level rate control.

Initialize the encoding parameters, such as the target bit budget, buffer status, etc.

Step 2: Frame level rate control.

2.1 Compute the current frame $\lambda_{curF} = \alpha \times R_{bpp,tar}^{\beta}$;

2.2 Adjust the frame level λ_{curF}. If λ_{curF} is bigger than that of the previous frame λ_{preF}, λ_{curF} is clipped by

$$\min\left(\lambda_{curF}, \lambda_{preF} \times e^{\frac{4.0}{4.0}}\right).$$

Otherwise, λ_{curF} is clipped by

$$\max\left(\lambda_{curF}, \lambda_{preF} \times e^{\frac{-4.0}{2.0}}\right);$$

2.3 Compute the frame level

$$QP_{curF} = \lfloor a \times \ln(\lambda_{frame}) + b + 0.5 \rfloor.$$

Step 3: CTU level rate control.

3.1 Let $k = 0$;

3.2 If the current CTU is intra encoded, then $QP_{ctu,k} = QP_{curF}$.

Otherwise,

$$\lambda_{ctu,k} = \frac{-\beta_{ctu,k}}{D_{ctu,k}^{prev}} \times e^{\frac{N \ln\left(\frac{R_{\max}}{N}\right) - \sum_{k=1}^{N} \ln \alpha_{ctu,k} \left(\frac{-\beta_{ctu,k}}{D_k^{prev}}\right)^{\beta_{ctu,k}}}{\sum_{k=1}^{N} \beta_{ctu,k}}},$$

and $\lambda_{ctu,k}^{*} = \chi^{\frac{1}{\beta_i}} \lambda_{ctu,k}$;

3.3 Adjust the CTU level $\lambda_{ctu,k}^{*}$.

If $\lambda_{ctu,k}^{*}$ is bigger than the previous CTU $\lambda_{ctu,k-1}^{*}$, $\lambda_{ctu,k}^{*}$ is clipped by

$$\min\left(\max\left(\lambda_{ctu,k}, \lambda_{ctu,k-1} \times e^{\frac{-2.0}{4.0}}\right), \lambda_{ctu,k-1} \times e^{\frac{2.0}{4.0}}\right).$$

Otherwise, $\lambda_{ctu,k}^{*}$ is clipped by

$$\min\left(\max\left(\lambda_{ctu,k}, \lambda_{ctu,k-1} \times e^{\frac{-1.0}{4.0}}\right), \lambda_{ctu,k-1} \times e^{\frac{1.0}{4.0}}\right);$$

3.4 Compute the CTU level

$$QP_{ctu,k} = \lfloor a \times \ln(\lambda_{ctu,k}) + b + 0.5 \rfloor;$$

3.5 Encode the kth CTU;

3.6 Update the CTU level model parameters: $\alpha_{ctu,k}^{*}$, $\beta_{ctu,k}^{*}$, and $D_{ctu,k}^{*}$;

Compute the real $\lambda_{real,ctu,k}$, according to the actual bits $R_{ctu,bpp,real,k}$.

$$\lambda_{ctu,real,k} = \alpha_{ctu,k} \times R_{ctu,bpp,real,k}^{\beta_{ctu,k}};$$

Then, update the CTU level model parameters:

$$\alpha_{ctu,k}^{*} = \alpha_{ctu,k} + \delta_{\alpha}\left(\ln(\lambda_{ctu,k}) - \ln(\lambda_{ctu,real,k})\right) \times \alpha_{ctu,k};$$

$$\beta_{ctu,k}^{*} = \beta_{ctu,k} + \delta_{\beta}\left(\ln(\lambda_{ctu,k}) - \ln(\lambda_{ctu,real,k})\right) \times \beta_{ctu,k};$$

$$D_{ctu,k}^{*} = \frac{1}{M \times M} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} |org_{i,j} - rec_{i,j}|^2.$$

3.7 $k = k + 1$;

If $k > N - 1$, jump to **Step 4**;

Otherwise, jump to **Step 3.2**.

Step 4: Update the frame level parameters: α^{*}, β^{*}, and χ.

$$\lambda_{curF,real} = \alpha \times R_{bpp,real}^{\beta},\ \chi = R_{bpp,tar}/R_{bpp,real};$$

$$\alpha^{*} = \alpha + \delta_{\alpha}\left(\ln(\lambda_{curF}) - \ln(\lambda_{curF,real})\right) \times \alpha;$$

$$\beta^{*} = \beta + \delta_{\beta}\left(\ln(\lambda_{curF}) - \ln(\lambda_{curF,real})\right) \times \beta.$$

Step 5: Check the loop condition.

5.1 If the current frame belongs to the current GOP, jump to **Step 2**;

Otherwise, jump to **Step 1**.

unique characteristics in the format of 4:2:0 YUV were used to simulate low-delay communications. The intra frame period (IFP) is set about 0.5 fps.

In the results, the standard deviation of Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) [34] are the measures for the smoothness of video quality. Besides the above measures, the quality change between adjacent frames is also employed, i.e.,

$$V_{avg} = \frac{1}{L-1} \sum_{k=2}^{L} \left| D_{Y,k} - D_{Y,k-1} \right|, \tag{16}$$

where L is the length of the coded video frames, and $D_{Y,k}$ can be the luminance Y-PSNR value or Y-SSIM of the kth frame. Since the Y-PSNR is frequently used in comparison, we do not explicitly distinguish PSNR and Y-PSNR in this paper. The additional complexity is measured by

$$T_{avg} = |T_{method} - T_{anchor}| / T_{anchor} \times 100\%, \tag{17}$$

where T_{method} and T_{anchor} are the total computational complexities of the candidate method and HM10.0, respectively. The buffer size is set as

$$B_{uffer} = D_{elay} \times T_{arget}, \tag{18}$$

where D_{elay} is the delay time for the real-time video bitstream, and T_{arget} is the channel bandwidth. D_{elay} is set at about 0.3 seconds in the simulation.

Obviously, high quality and low buffer occupancy cannot be achieved concurrently due to a contradiction between them in low-delay communications. In order to achieve low-latency, the buffer size B_{uffer} is set as small as possible. According to equation (18), the buffer occupancy is mainly determined by the target bits, which can be adjusted by changing the target bits as well as the video quality. Fig. 3 shows four typical buffer occupancy results under the low-delay P configuration, which demonstrates the superiority of the proposed approach that can achieve lower buffer occupancy and less buffer fluctuation in comparison with HM10.0. Such an improvement is beneficial to enhance both the video quality and the buffer occupancy level. The buffer fullness curves in Fig. 3 show that the proposed method has no buffer overflow whereas HM10.0 can lead to the buffer overflow. Meanwhile, our method can provide less buffer variation (e.g., no buffer underflow) in comparison with Choi *et al.* [10] and Lee *et al.* [14]. It is noted that our work here is not to study the effect of buffer overflow (or underflow) and frame-skip, but to examine the role of the low buffer occupancy in the low-delay transmission. We believe that if the frame-skip is enabled, HM10.0 would give a worse quality fluctuation due to the loss of high-quality reference frames.

Besides controlling the buffer occupancy, high quality is also desirable in low-latency communication. In Fig. 4, we compare the overall rate-distortion performance of the proposed algorithm with that of HM10.0, Choi *et al.* [10] and Lee *et al.* [14] under the low-delay P configuration. Experiments show that R-lambda based methods (i.e., HM10.0 and our method) generally gives a better rate-distortion performance compared to that of R-Q based methods (i.e., Choi *et al.* [10] and Lee *et al.* [14]). Furthermore, it can be seen that our method outperforms the other methods for all of bit rates. In Fig. 4 (b), our method can achieve a significant video quality improvement. The reason is that the

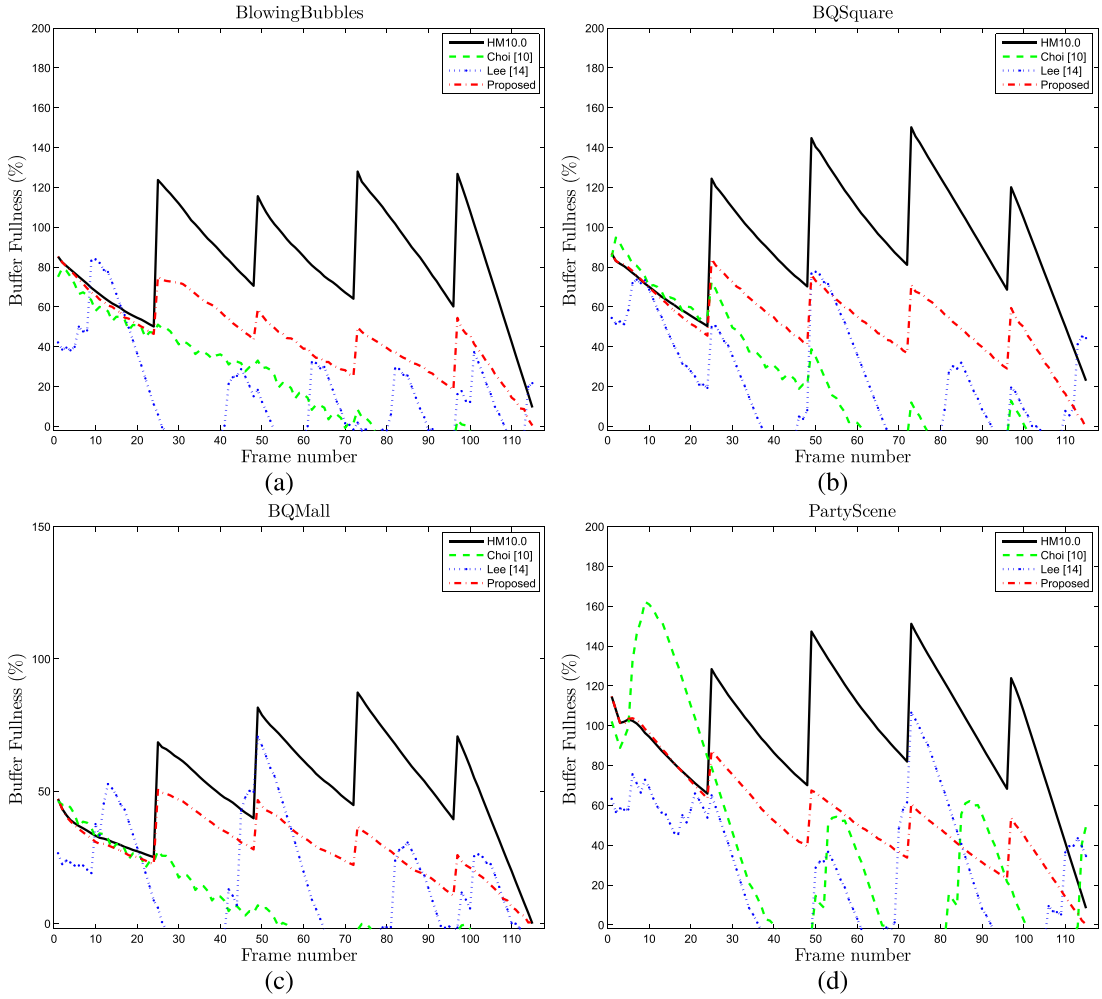

Fig. 3. Comparison of frame-by-frame buffer occupancy: (a) *BlowingBubbles* (0.8 Mbps), (b) *BQSquare* (0.8 Mbps), (c) *BQMall* (3.2 Mbps) and (d) *PartyScene* (3.2 Mbps).

BQSquare sequence contains zoom motion, and HM10.0 has poor estimation of encoding parameters in the raster-scanning bit allocation, λ adjustment, and quality control. On the other hand, our method establishes a novel relationship between distortion and λ. Based on the new distortion model, we obtain a computationally feasible solution to the problem of obtaining optimal λ for consistent video quality that can avoid the use of raster-scanning bit allocation. Consequently, the proposed method can achieve a better compromise between quality variation and buffer occupancy in comparison with HM10.0. Other simulation results also verify similar performance.

Fig. 5 shows the subjective comparison of video quality for *Baskballpass* (i.e., from frame 106 to 115) under the low-delay P configuration. The comparisons of the reconstructed frames in Fig. 5 validate the visual quality performance of the proposed method. As shown in the top two rows of Fig. 5, it can be seen that our method has a better quality in the texture area, where the numbers on the scoreboard clearly demonstrate the superiority of our method. In addition, distortion maps show a subjective comparison, and small distortion indicates higher video quality. We can observe that the fourth row of Fig. 5 is more homogeneous compared to the third row, where the brightness is the error measure. Fig. 6 shows the case for *BQSquare* from frame 106 to 115. In the results of *BQSquare*, the maximum SSIM variation is 0.0376 in HM10.0 while the maximum SSIM variation is 0.0230 in our method. Simulations have presented similar performance for all other sequences that we have tested.

Since the SSIM score is considered to be closer to human evaluation than the PSNR value, simulation results of subjective comparison in terms of SSIM are tabulated in Tables III and IV. In the simulation, we first compute the frame-by-frame SSIM score for each video sequence, and then the standard deviation of SSIM and the average SSIM change between adjacent frames are used as the measures of video quality variation. In Tables III and IV, "Avg.", "Std. dev." and "V_{avg}" represents the average SSIM, the standard deviation of SSIM and the average quality change between adjacent frames, respectively. It can be seen that compared to HM10.0 and Choi *et al.* [10], our method can achieve the highest average SSIM scores and the lowest SSIM variations in both the low-delay P and B Main coding structures. Meanwhile, the average SSIM score of Choi *et al.* [10] is worse than that

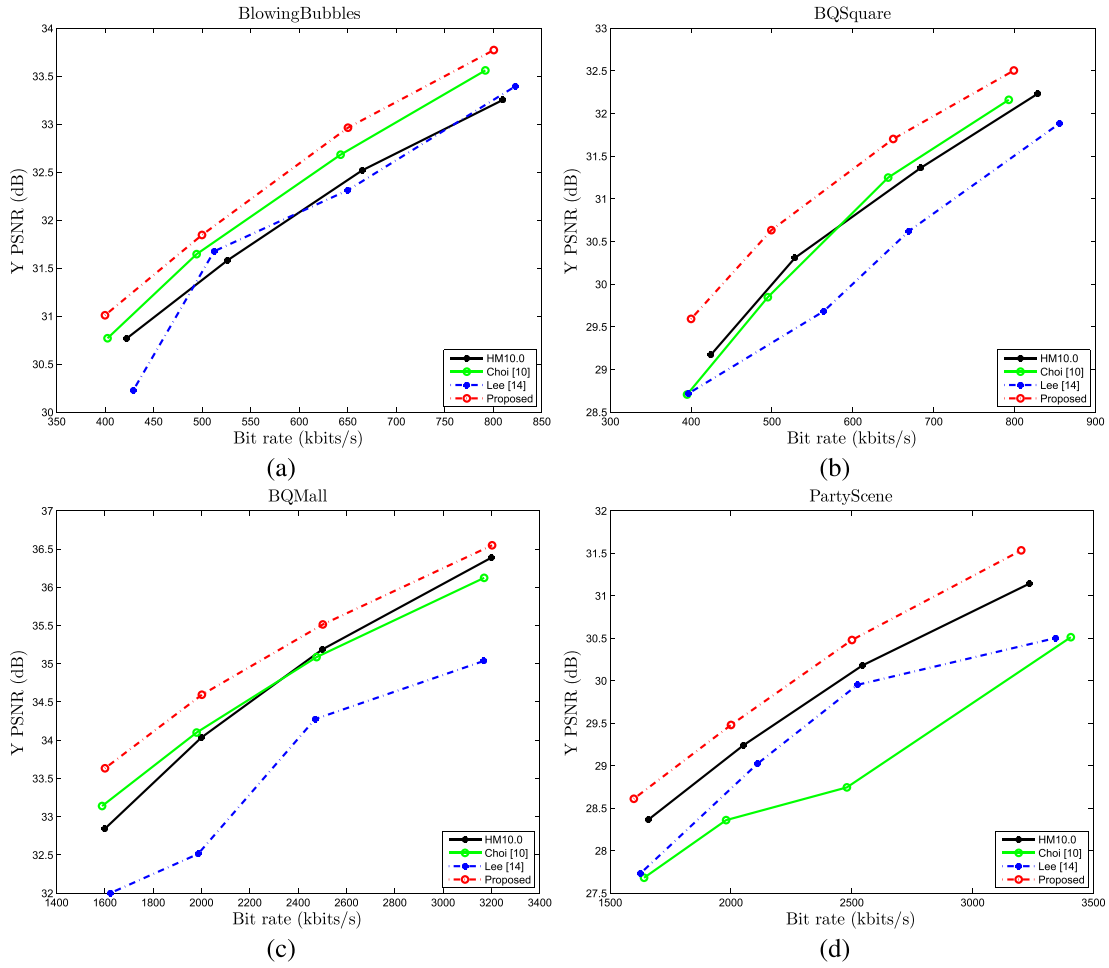

Fig. 4. Rate-distortion curves: (a) *BlowingBubbles*, (b) *BQSquare*, (c) *BQMall* and (d) *PartyScene*.

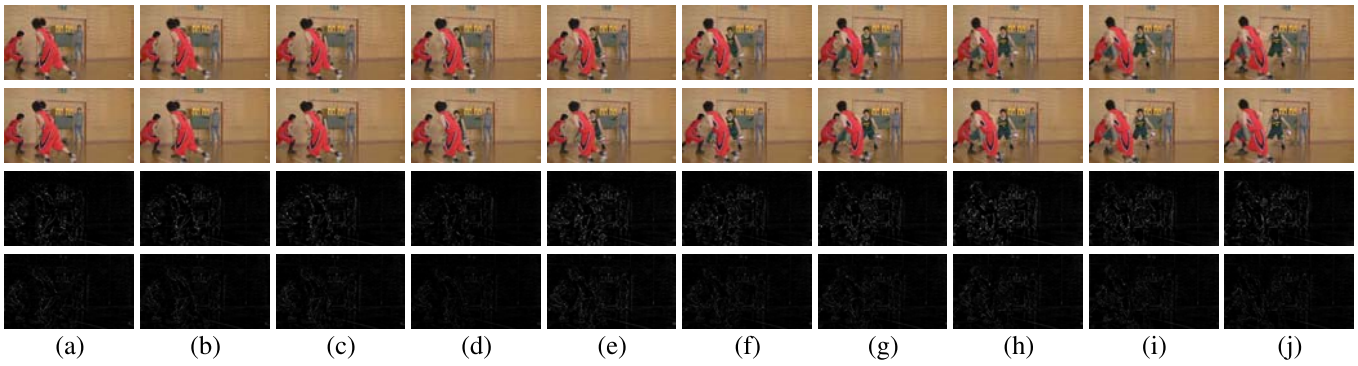

Fig. 5. Subjective visual quality comparisons of *BasketballPass* at bit rate 0.4 Mbps: (a) ∼ (j) HM10.0, proposed method, distortion map of HM10.0, and distortion map of the proposed method (from top to bottom).

of the HM10.0 algorithm. The detailed simulation results of SSIM can be found in Table III and Table IV.

We conducted the subjective experiment to evaluate visual quality as suggested in [30]. There are total fifteen subjects who participated in the test. The mean opinion score (MOS) takes the five-point scale rule (i.e., 5-excellent, 4-good, 3-fair, 2-poor, and 1-bad). Five representative videos are selected as the test source, including the following content characteristics, such as fast and slow motion, high and low resolution. The distorted videos are the low bit-rate reconstructed results from Table V. The results are shown in Fig. 7, where the bigger MOS value means the better visual quality of the decoded video stream. It can be observed that the proposed method outperforms the other methods.

For easy comparison, the simulation results of PSNR and the variations thereof are tabulated in Table V. In the low-delay P coding structure, the average PSNR improvement of our method is about 0.35 dB and 0.59 dB in comparison

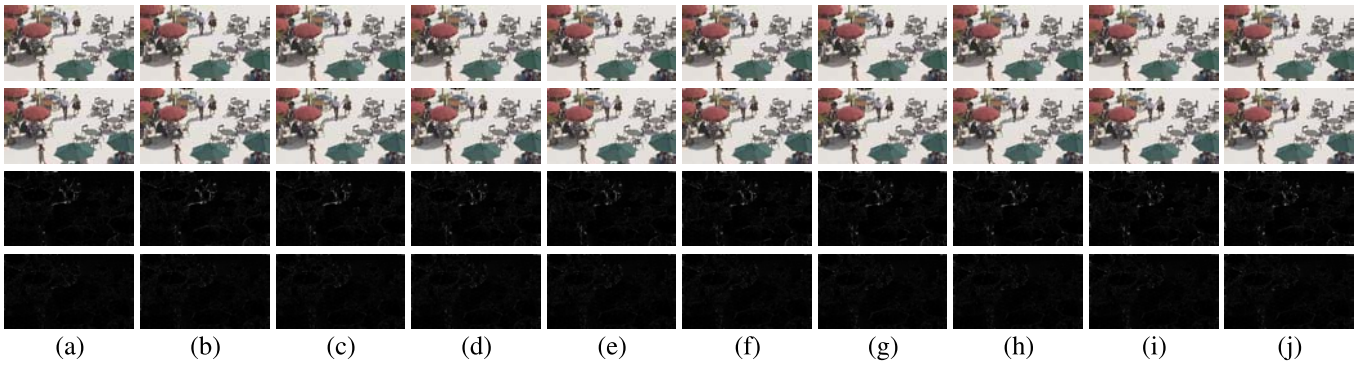

Fig. 6. Subjective visual quality comparisons of *BQSquare* at bit rate 0.4 Mbps: (a) ∼ (j) HM10.0, proposed method, distortion map of HM10.0, and distortion map of the proposed method (from top to bottom).

TABLE III

COMPARISONS OF SSIM VALUES IN THE LOW-DELAY P CODING STRUCTURE

Sequence	Target rate	HM10.0			Choi [10]			Proposed		
		Avg.	Std. dev.	V_{avg}	Avg.	Std. dev.	V_{avg}	Avg.	Std. dev.	V_{avg}
Class A	1600	0.8987	0.0103	0.0281	0.9004	0.0226	0.0403	0.8988	0.0101	0.0205
	3200	0.9351	0.0159	0.0243	0.9205	0.0228	0.0386	0.9355	0.0139	0.0139
Class B	5000	0.8965	0.0117	0.0154	0.8916	0.0240	0.0216	0.8989	0.0064	0.0051
	10000	0.8951	0.0202	0.0096	0.8889	0.0271	0.0207	0.8990	0.0084	0.0032
Class C	1600	0.8788	0.0443	0.0100	0.8693	0.0478	0.0252	0.8814	0.0211	0.0075
	3200	0.9216	0.0286	0.0089	0.8983	0.0484	0.0354	0.9194	0.0165	0.0079
Class D	400	0.8375	0.0468	0.0247	0.8214	0.0411	0.0410	0.8355	0.0324	0.0179
	800	0.8962	0.0339	0.0206	0.8901	0.0274	0.0215	0.8949	0.0203	0.0120
Class E	400	0.9612	0.0015	0.0016	0.9610	0.0015	0.0025	0.9661	0.0014	0.0013
	800	0.9664	0.0014	0.0018	0.9660	0.0014	0.0014	0.9710	0.0014	0.0011
Class F	400	0.8934	0.0104	0.0318	0.8665	0.0161	0.0530	0.8928	0.0074	0.0246
	800	0.9519	0.0035	0.0215	0.9469	0.0113	0.0319	0.9539	0.0024	0.0129
Average		0.9110	0.0190	0.0165	0.9017	0.0243	0.0278	0.9123	0.0118	0.0107

TABLE IV

COMPARISONS OF SSIM VALUES IN THE LOW-DELAY B CODING STRUCTURE

Sequence	Target rate	HM10.0			Choi [10]			Proposed		
		Avg.	Std. dev.	V_{avg}	Avg.	Std. dev.	V_{avg}	Avg.	Std. dev.	V_{avg}
Class A	1600	0.8308	0.0256	0.0469	0.8316	0.0302	0.0412	0.8314	0.0229	0.0389
	3200	0.8809	0.0203	0.0354	0.8758	0.0258	0.0294	0.8791	0.0175	0.0287
Class B	5000	0.9314	0.0197	0.0090	0.9243	0.0290	0.0177	0.9310	0.0163	0.0084
	10000	0.9207	0.0270	0.0081	0.9145	0.0284	0.0166	0.9231	0.0162	0.0068
Class C	1600	0.8683	0.0472	0.0200	0.8639	0.0449	0.0278	0.8754	0.0202	0.0093
	3200	0.8873	0.0286	0.0245	0.8813	0.0379	0.0278	0.8896	0.0126	0.0131
Class D	400	0.8387	0.0472	0.0250	0.8307	0.0294	0.0230	0.8364	0.0335	0.0179
	800	0.8977	0.0360	0.0202	0.8878	0.0220	0.0164	0.8969	0.0203	0.0118
Class E	400	0.9612	0.0031	0.0034	0.9600	0.0031	0.0049	0.9615	0.0031	0.0033
	800	0.9659	0.0049	0.0039	0.9654	0.0044	0.0045	0.9661	0.0045	0.0041
Class F	400	0.9272	0.0055	0.0276	0.9359	0.0128	0.0298	0.9258	0.0052	0.0224
	800	0.9489	0.0039	0.0271	0.9518	0.0080	0.0203	0.9511	0.0028	0.0180
Average		0.9049	0.0224	0.0209	0.9019	0.0230	0.0216	0.9056	0.0146	0.0153

with HM10.0 and Choi *et al.* [10], respectively. The average rate control errors of HM10.0, Choi *et al.* [10] and our method are 1.54%, 1.40% and 0.08%, respectively. For the quality variation measurement in terms of standard deviation, HM10.0, Choi *et al.* [10] and our method are 2.52, 2.21 and 1.31, respectively. In addition, for the quality variation measurement in terms of V_{avg}, HM10.0, Choi *et al.* [10] and our method are 0.50, 1.03 and 0.21, respectively. As can be observed from the simulation results, the proposed method generally outperforms HM10.0 and Choi *et al.* [10] in terms of the average PSNR and PSNR variations in the low-delay P coding structure.

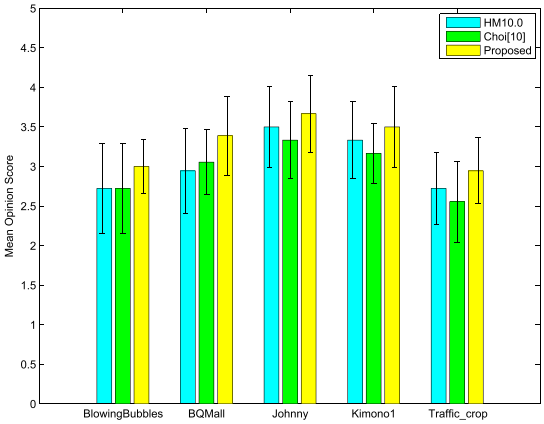

Fig. 7. Subjective quality comparisons between the proposed method and the traditional algorithms.

In order to obtain a better coding efficiency, HEVC supports the low-delay B coding structure. The associated simulation results are shown in Table VI. In our method, model parameters updates in equation (13) only rely on the nearest previous frame. The average PSNR improvement of our method is about 0.39 dB and 1.24 dB, compared to HM10.0 and Choi *et al.* [10], respectively. Meanwhile, the average standard deviations of PSNR are 2.57, 1.74 and 1.34 for HM10.0, Choi *et al.* [10] and our method, respectively, while the quality variation V_{avg}s of PSNR are 0.48, 0.95 and 0.21, respectively. It can be seen that the proposed method can provide a more consistent video quality compared to HM10.0 and Choi *et al.* [10] in terms of the PSNR variations under the low-delay B coding structure.

To get a consistent quality control, we employ additional operations to compute a distortion-based λ for each CTU. T_{avg} is considered as a factor to measure its complexity. Compared to HM10.0 in the low-delay P coding structure, the T_{avg} value of our method is about 3.90%, while that of Choi *et al.* [10] is about 27.57% as shown in Table V. In the low-delay B coding structure, the T_{avg} value of our method is about 4.98%, while Choi *et al.* [10] is about 22.52%. It can be seen that the complexity of our method is lower than that of Choi *et al.* [10] in both the low-delay P and B coding structures. There are two reasons. One is that the model parameters in the

TABLE V
SIMULATION RESULTS IN THE LOW-DELAY P CODING STRUCTURE

Sequence	Target rates (Kbps)	HM10.0 Actual rates	HM10.0 PSNR(dB) Avg.	HM10.0 PSNR(dB) Std. dev.	HM10.0 V_{avg}	Choi [10] Actual rates	Choi [10] PSNR(dB) Avg.	Choi [10] PSNR(dB) Std. dev.	Choi [10] V_{avg}	Choi [10] T_{avg} (%)	Proposed Actual rates	Proposed PSNR(dB) Avg.	Proposed PSNR(dB) Std. dev.	Proposed V_{avg}	Proposed T_{avg} (%)
Classs A	1600	1713.1	28.74	2.4542	0.5481	2582.5	28.36	3.2900	0.9985	22.1	1599.3	29.27	1.4278	0.3379	3.77
	3200	3248.9	31.89	2.4504	0.4469	3319.3	31.43	2.9049	1.1435	27.17	3200.9	32.31	1.0591	0.1705	5.13
Classs B	5000	5064.9	35.36	2.1654	0.4097	4966.4	34.90	1.8259	1.0161	18.75	5000.3	35.94	0.7756	0.1399	4.51
	10000	10000	37.51	1.3195	0.3620	9897.6	37.15	1.6782	1.0694	24.34	10002	37.75	0.6789	0.1519	7.70
Classs C	1600	1614.7	31.85	2.9184	0.5883	1599.8	31.49	2.7578	0.9939	19.30	1599.3	32.26	1.4983	0.2508	4.21
	3200	3208.6	34.97	2.2451	0.5463	3231.5	34.44	2.4724	1.2662	34.67	3199.4	35.08	1.3328	0.2347	8.20
Classs D	400	411.62	30.89	2.8042	0.5811	397.8	30.53	2.7460	0.8562	26.49	399.9	31.24	1.8128	0.2700	6.55
	800	809.71	33.96	2.8818	0.6259	793.9	33.54	2.0568	0.8910	30.04	800.3	34.21	1.5509	0.2704	10.21
Classs E	400	415.02	37.55	1.2548	0.1697	465.6	37.24	2.0035	0.9280	16.37	396.8	37.79	0.4733	0.1321	3.17
	800	814.45	39.04	1.3924	0.1655	804.5	38.61	1.9105	0.9891	24.12	797.8	40.24	0.4431	0.1154	5.59
Classs F	400	407.18	34.33	3.5457	0.7099	371.7	34.09	6.1225	1.5257	15.18	396.2	34.69	3.3558	0.5973	2.20
	800	796.62	37.33	3.2495	0.6929	779.9	36.83	5.8019	1.8814	23.62	801.1	37.24	3.5629	0.5496	4.16
Average			34.45	2.3901	0.4872		34.05	2.9642	1.1299	23.51		34.68	1.4976	0.2684	5.45

TABLE VI
SIMULATION RESULTS IN THE LOW-DELAY B CODING STRUCTURE

Sequence	Target rates (Kbps)	HM10.0 Actual rates	HM10.0 PSNR(dB) Avg.	HM10.0 PSNR(dB) Std. dev.	HM10.0 V_{avg}	Choi [10] Actual rates	Choi [10] PSNR(dB) Avg.	Choi [10] PSNR(dB) Std. dev.	Choi [10] V_{avg}	Choi [10] T_{avg} (%)	Proposed Actual rates	Proposed PSNR(dB) Avg.	Proposed PSNR(dB) Std. dev.	Proposed V_{avg}	Proposed T_{avg} (%)
Classs A	1600	1709.4	28.89	2.4598	0.5329	1588.6	28.28	1.8114	0.9738	18.15	1600.3	29.39	1.4143	0.3316	4.73
	3200	3249.1	32.06	2.4740	0.4350	3173.4	31.52	1.3702	0.8343	26.21	3200.8	32.47	1.0580	0.1712	7.21
Classs B	5000	5065.3	35.63	2.1259	0.3881	4943.9	34.79	1.4088	0.9367	9.61	4998.9	36.19	0.7576	0.1389	4.98
	10000	10001	37.81	1.2862	0.3378	9895.7	37.27	1.5508	1.0219	17.58	10002	38.03	0.6572	0.1488	9.88
Classs C	1600	1616.1	31.79	3.0225	0.5705	1605.8	31.05	1.8236	0.9519	21.59	1599.0	32.25	1.4853	0.2329	6.64
	3200	3211.7	35.03	2.2365	0.5283	3213.0	34.21	2.0227	1.1537	30.17	3199.4	35.17	1.3002	0.2257	11.51
Classs D	400	411.42	30.98	2.8674	0.5864	396.2	30.16	1.4272	0.7137	18.12	400.1	31.37	1.9237	0.2883	4.94
	800	811.16	34.18	3.0051	0.6179	793.8	33.97	1.4399	0.7876	26.41	800.2	34.51	1.6083	0.2716	7.92
Classs E	400	415.69	36.87	1.2885	0.1955	380.4	36.39	1.5248	0.7819	13.27	407.8	37.23	1.1827	0.1719	5.61
	800	814.49	39.38	1.4404	0.1865	774.7	38.95	1.5762	0.7851	19.75	803.8	39.31	1.2603	0.1302	8.64
Classs F	400	408.33	34.23	3.7927	0.7067	427.6	34.01	3.7861	1.5118	17.36	395.7	33.99	3.4276	0.6180	2.23
	800	797.54	37.49	3.4699	0.7346	779.9	37.15	4.7091	1.6160	24.71	791.0	37.73	3.7267	0.6022	3.89
Average			34.53	2.4557	0.4850		33.98	2.0376	1.0057	20.24		34.80	1.6502	0.2776	6.52

proposed algorithm can be easily computed. The other being the computational complexity of the R-lambda based rate control is marginal when compared to that of the whole encoding system. Furthermore, it should be pointed out that the real-time concept is from the algorithm design instead of the computer-based implementation here. Therefore, considering the improvement in video quality, one can conclude that the proposed algorithm outperforms HM10.0 in low-delay video communications.

VI. CONCLUSION

This paper focuses on consistent quality control for HEVC and introduces an efficient distortion-based Lagrange multiplier approach in low-latency video communications. Using the distortion of co-located CTU in the previous frame, a new relationship between distortion and λ is established and employed to control the video quality fluctuations. Based on the proposed distortion model, we obtain a computationally feasible λ in the minimization of the total distortion subject to a given bit rate. When considered jointly with the buffer state, the CTU level λ is further adjusted such that the target bits satisfy the overall bandwidth of low-delay video communication. As demonstrated in the simulation experiments, the proposed rate control method outperforms state-of-the-art techniques in terms of bit rate regulation, video quality fluctuation and encoder buffer fullness.

APPENDIX

The Karush-Tuhn-Tucker (KKT) condition of equation (12) is (A.1), as shown at the top of the next page.

It is easy to check that $\sum_{k=1}^{N} \frac{\sigma}{\lambda_k^{prev}} \times D_k^{prev} \times \lambda_k^{curr}$ and $\sum_{k=1}^{N} \left(\ln(\alpha_k) + \beta_k \ln\left(\lambda_k^{curr}\right)\right) - N \ln\left(\frac{R_{\max}}{N}\right)$ are convex functions with respect to the variable λ_k^{curr}. In this case, we know that if the point $\left(\lambda_1^*, \lambda_2^*, \ldots, \lambda_N^*\right)$ satisfies the above KKT conditions, it is a globe minimum solution for the optimization problem (12).

In practice, the distortion and the previous $\lambda_i^{prev} (i = 1, \ldots, N)$ is a non-negative value, and hence we know that $\frac{\sigma}{\lambda_i^{prev}} \times D_i^{prev} \neq 0$ for any $i \in [1, 2, \ldots, N]$. In this case,

$$\begin{cases} \dfrac{\partial L}{\partial \lambda_i^{curr}} = \dfrac{\sigma}{\lambda_i^{prev}} \times D_i^{prev} + u\dfrac{\beta_i}{\lambda_i^{curr}} = 0, \quad i = 1, \ldots, N. & (i) \\ u\left(\sum\limits_{k=1}^{N}\left(\ln(\alpha_k) + \beta_k \ln\left(\lambda_k^{curr}\right)\right) - N\ln\left(\dfrac{R_{\max}}{N}\right)\right) = 0, u \geq 0. & (ii) \\ \lambda_i^{curr} > 0, \quad \lambda_i^{prev} > 0, \quad D_i^{prev} > 0, & (iii) \\ \sigma > 0, \quad \alpha_i > 0, \quad \beta_i < 0, \quad i = 1, \ldots, N. \end{cases} \tag{A.1}$$

we know $u \neq 0$ in the condition (i). As a result, λ_i^{curr} can be expressed as

$$\lambda_i^{curr} = \frac{(-\beta_i)\, u}{\frac{\sigma}{\lambda_i^{prev}} \times D_i^{prev}} = \frac{u}{\frac{\sigma}{\lambda_i^{prev}}} \times \frac{-\beta_i}{D_i^{prev}}. \tag{A.2}$$

Substituting equation (A.2) into the condition (ii), we can get

$$\sum_{k=1}^{N}\left(\ln(\alpha_k) + \beta_k \ln\left(\frac{u}{\frac{\sigma}{\lambda_k^{prev}}} \times \frac{-\beta_k}{D_k^{prev}}\right)\right) = N\ln\left(\frac{R_{\max}}{N}\right). \tag{A.3}$$

Equation (A.3) can be further rewrote as

$$\frac{u}{\frac{\sigma}{\lambda_i^{prev}}} = e^{\frac{N\ln\left(\frac{R_{\max}}{N}\right) - \sum\limits_{k=1}^{N}\left(\ln\alpha_k\left(\frac{-\beta_k}{D_k^{prev}}\right)^{\beta_k}\right)}{\sum\limits_{k=1}^{N}\beta_k}}. \tag{A.4}$$

Using equations (A.2) and (A.4), λ_i^{curr} is formulated as

$$\lambda_i^{curr} = \frac{u}{\frac{\sigma}{\lambda_i^{prev}}} \times \frac{-\beta_i}{D_i^{prev}} = \frac{-\beta_i}{D_i^{prev}} \times e^{\frac{N\ln\left(\frac{R_{\max}}{N}\right) - \sum\limits_{k=1}^{N}\left(\ln\alpha_k\left(\frac{-\beta_k}{D_k^{prev}}\right)^{\beta_k}\right)}{\sum\limits_{k=1}^{N}\beta_k}}. \tag{A.5}$$

ACKNOWLEDGMENT

The author would like to thank the helpful comments given by the anonymous reviewers.

REFERENCES

[1] 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.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[3] 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.

[4] Z. He and S. K. 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.

[5] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by R-lambda model for HEVC," document JCTVC-K0103, Shanghai, China, 2012.

[6] X. Wang and M. Karczewicz, "Intra frame rate control based on SATD," document JCTVC-M0257, Incheon, South Korea, 2013.

[7] M. Wang, K. N. Ngan, and H. Li, "An efficient frame-content based intra frame rate control for high efficiency video coding," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 896–900, Jul. 2015.

[8] H. Choi, J. Nam, J. Yoo, D. Sim, and I. Bajić, "Rate control based on unified RQ model for HEVC," document JCTVC-H0213, San Jose, CA, USA, 2012.

[9] 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.

[10] H. Choi, J. Yoo, J. Nam, D. Sim, and I. V. Bajić, "Pixel-wise unified rate-quantization model for multi-level rate control," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1112–1123, Dec. 2013.

[11] X. Jing, L.-P. Chau, and W.-C. Siu, "Frame complexity-based rate-quantization model for H.264/AVC intraframe rate control," *IEEE Signal Process. Lett.*, vol. 15, pp. 373–376, 2008, doi: 10.1109/LSP.2008.920010.

[12] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP based rate control for high efficiency video coding," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1101–1111, Dec. 2013.

[13] B. Lee and M. Kim, "Modeling rates and distortions based on a mixture of Laplacian distributions for inter-predicted residues in quadtree coding of HEVC," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 571–574, Oct. 2011.

[14] B. Lee, M. Kim, and T. Q. Nguyen, "A frame-level rate control scheme based on texture and nontexture rate models for high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 3, pp. 465–479, Mar. 2014.

[15] 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.

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

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

[18] 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.

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

[20] 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.

[21] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.

[22] 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.

[23] M. Wang and M. van der Schaar, "Operational rate-distortion modeling for wavelet video coders," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3505–3517, Sep. 2006.

[24] Y.-K. Tu, J.-F. Yang, and M.-T. Sun, "Rate-distortion modeling for efficient H.264/AVC encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 5, pp. 530–543, May 2007.

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

[26] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1027–1042, Apr. 1998.

[27] Z. Chen and K. N. Ngan, "Towards rate-distortion tradeoff in real-time color video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 2, pp. 158–167, Feb. 2007.

[28] L. Xu, D. Zhao, X. Ji, L. Deng, S. Kwong, and W. Gao, "Window-level rate control for smooth picture quality and smooth buffer occupancy," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 723–734, Mar. 2011.
[29] J. Hou, S. Wan, Z. Ma, and L.-P. Chau, "Consistent video quality control in scalable video coding using dependent distortion quantization model," *IEEE Trans. Broadcast.*, vol. 59, no. 4, pp. 717–724, Dec. 2013.
[30] L. Xu, S. Li, K. N. Ngan, and L. Ma, "Consistent visual quality control in video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 6, pp. 975–989, Jun. 2013.
[31] G. M. Schuster, G. Melnikov, and A. K. Katsaggelos, "A review of the minimum maximum criterion for optimal bit allocation among dependent quantizers," *IEEE Trans. Multimedia*, vol. 1, no. 1, pp. 3–17, Mar. 1999.
[32] D. A. Pierre, *Optimization Theory With Applications*. New York, NY, USA: Dover, 2012.
[33] (2014). *HEVC Reference Software*. [Online]. Available: http://hevc.kw.bbc.co.uk/trac/browser/jctvc-hm/tags
[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

Miaohui Wang (S'13–M'16) received the Ph.D. degree from the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, in 2015. From 2014 to 2015, he was a Visiting Scholar with the Innovation Laboratory, InterDigital Inc., San Diego, CA, USA. He received the Best Master Thesis Award in Shanghai (2011) and Fudan University (2012), China.

He has authored or co-authored numerous technical papers in international journals and conferences. His current research interests cover a wide range of topics related with video compression and transmission, computer vision and machine learning, including transform coding, rate control, image restoration, and denoising, and deepneuron-network-based applications. He is a member of the IEEE Circuits and Systems Society.

King Ngi Ngan (M'79–SM'91–F'00) received the Ph.D. degree in electrical engineering from Loughborough University, Loughborough, U.K.

He was a Full Professor with Nanyang Technological University, Singapore, and with the University of Western Australia, Perth, Australia. He is currently a Chair Professor with the Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong. He holds honorary and visiting professorships with numerous universities in China, Australia, and South East Asia. He has published extensively, including three authored books, six edited volumes, over 300 refereed technical papers, and has edited nine special issues in journals. He holds ten patents in image or video coding and communications.

Dr. Ngan is a fellow of IET, U.K., and IEAust, Australia, and was an IEEE Distinguished Lecturer from 2006 to 2007. He has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the *Journal on Visual Communications and Image Representation*, the *EURASIP Journal of Signal Processing: Image Communication*, and the *Journal of Applied Signal Processing*. He has chaired a number of prestigious international conferences on video signal processing and communications, and has served on the advisory and technical committees of numerous professional organizations. He co-chaired the IEEE International Conference on Image Processing, Hong Kong, in 2010.

Hongliang Li (SM'12) received the Ph.D. degree in electronics and information engineering from Xi'an Jiaotong University, Xi'an, China, in 2005.

He was a Research Associate with the Visual Signal Processing and Communication Laboratory (VSPC), Chinese University of Hong Kong (CUHK), Hong Kong, from 2005 to 2006. From 2006 to 2008, he was a Post-Doctoral Fellow with VSPC, CUHK. He is currently a Professor with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China. He has authored or co-authored numerous technical articles in international journals and conferences. He is a Co-Editor of the book entitled *Video Segmentation and its Applications* (Springer, 2011). His research interests include image segmentation, object detection, image and video coding, visual attention, and multimedia communication system.

Dr. Li is a member of the Editorial Board of the *Journal on Visual Communications and Image Representation*, and the Area Editor of *Signal Processing: Image Communication*. He served as a Technical Program Co-Chair in ISPACS 2009, the General Co-Chair of the ISPACS 2010, the Publicity Co-Chair of the IEEE VCIP 2013, the Local Chair of the IEEE ICME 2014, and a TPC Member in a number of international conferences, such as ICME 2013, ICME 2012, ISCAS 2013, PCM 2007, PCM 2009, and VCIP 2010. He serves as a Technical Program Co-Chair of the IEEE VCIP 2016. He was selected as the New Century Excellent Talent in University, Chinese Ministry of Education, China, in 2008.