Perceptual-Based Global Optimization for CTU-Level Rate Control in HEVC

Mingliang Zhou, Xuekai Wei, Wei Gao, Chi-Keung Fong, Peter H. W. Wong, Wilson Y. F. Yuen, Shiqi Wang, *Member, IEEE*, Sam Kwong, *Fellow, IEEE*

***Abstract*—We propose a coding tree unit (CTU) level rate control scheme from the perspective of perceptual rate-distortion optimization to improve the coding efficiency. Firstly, the perceptual rate-distortion model is established based on the divisive normalization scheme, which characterizes the relationship between local visual quality and coding bits. Subsequently, the established model is applied to the CTU-level rate control, which is then transformed into a global optimization problem solved with convex optimization. Our algorithm can optimally achieve CTU level bit allocation given the bit rate budget. According to the experimental results, our algorithm can substantially enhance the coding performance and consistently outperform the rate control scheme in HEVC reference software and the existing algorithms for different test configurations in terms of rate-perceptual distortion performance.**

***Index Terms*—Human visual perception, rate control, rate-distortion optimization, global optimization**

I. Introduction

HIGH-EFFICIENCY video coding (HEVC) [1], [2], which substantially improves the coding performance compared to the preceding video coding standards, plays important roles in various video-relevant applications. Rate control (RC) is of paramount significance to video compression and transmission, which performs efficient bit allocation and encoding given the constraint of bit rate budget. In particular, RC aims to achieve high bit rate control accuracy and improve rate-distortion (R-D) performance [3], [4]. Bit allocation in RC targets for realizing the optimal R-D performances through efficient allocation at different coding levels, including group of pictures (GOPs), frame and block [1]. Generally speaking, RC algorithms are developed in accordance with the specific video coding standards [5]–[11]. For example, TM5 [5] algorithm is applied in MPEG-2 and VM8 is utilized by MPEG-4 [6], etc.

Recently, several studies have been conducted to improve RC optimization or HEVC [12]–[22]. There are three categories of RC algorithms for HEVC: quadratic model [12], ρ-domain model [13] and R-λ model. More specifically, Li *et al.* [14] first proposed the λ domain RC based on the relationship between coding bits and the Lagrangian multiplier. Due to the low complexity and high efficiency, R-λ model has been adopted in HEVC reference software as the default RC algorithm. Lee *et al.* investigated the Laplacian probability distribution function (PDF) in [22] to model the residue, and proposed independent R-Q models to establish the relationship between the quantization parameters and coding bits, including texture and non-texture bits. Moreover, intra frame RC algorithms have also been studied. Li *et al.* [23] proposed an adaptive bit allocation algorithm to improve the R–λ model RC algorithm on intra frame. In [24], sum of absolute transformed differences (SATD) was used to measure the complexity for intra-frame, which further improves the performance. Wang *et al.* proposed an intra R-λ model in [25], and the gradient was used to characterize the picture complexity.

CTU level RC is also playing an important role in regulating the bit rate and improving the coding performance [16]–[20]. However, as the ultimate receiver of the video streams is the human visual system, the perceptual characteristics should be fully considered in CTU-Level bit allocation. In this paper, we investigate the CTU level rate control based on perceptual rate-distortion optimization, and formulates the CTU level rate control. The contributions of this study are as follows:

- We establish the relationship between rate and perceptual distortion with the divisive normalization scheme, which is further applied in the formulation of CTU level bit allocation.
- We transform the CTU-level bit allocation into a global optimization problem, which is subsequently solved by convex optimization.
- Extensive experimental results demonstrate that the proposed scheme achieves efficient bit allocation and rate control with sufficient control accuracy and better perceptual rate-distortion performance.

The paper is structured as follows. Section II reviews the related works. The problem formulation of perceptual rate control is presented in Section III. Section IV proposes the CTU-level rate control scheme and the experimental results are shown in Section IV. Section V draws the conclusions.

II. Related Work

A. Perceptual Video Coding

The peak signal noise ratio (PSNR) and mean squared error (MSE), which have been widely adopted in the video coding optimization process, are widely criticized due to the low correlation with human visual perception. The Structural Similarity (SSIM) index was proposed based on the comparison of the image structure [26], and has been proved to better measure the distortion of perceived by human visual system (HVS). Therefore, it has been widely adopted in the

Mingliang Zhou, Xuekai Wei, Wei Gao, Shiqi Wang and Sam Kwong are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong.

Chi-Keung Fong, Peter H. W. Wong and Wilson Y. F. Yuen are with Senior Software Engineer TFI Digital Media

Corresponding author: Sam Kwong, e-mail: cssamk@cityu.edu.hk.

optimization process of image and video processing. Wang *et al.* proposed to optimize video coding performance with SSIM as the evaluation criterion, and the quantization, rate-distortion optimization (RDO) and two-pass encoding based on SSIM were developed [27]–[29]. However, the CTU level bit allocation was not considered in these methods. In [30], [31], SSIM-based RC schemes were also proposed to achieve better rate control performance of H.264/AVC. The SSIM-based schemes were developed for inter frame coding in [30]. In [21], a game theory (GT) approach based on SSIM was proposed, which achieves satisfactory coding performance for intra coding. However, this method has not been extended to the inter coding case. The Just Noticeable Distortion (JND) model that quantitatively measures the maximum allowed distortion in visual perception receives significant interest in perceptual video coding. In [32], Yang *et al.* proposed a JND-based residual suppression method by a nonlinear additive model. Luo *et al.* [33] developed a JND-based compliant perceptual video coding by tuning the levels of quantized transform coefficients.

B. Block-Level Rate Control

In H.264/AVC, macroblock (MB) is the basic processing unit and MB-level RC algorithms have been proposed for H.264/AVC. In particular, the RC algorithm with the quadratic rate quantization model was proposed [7], [11]. In order to improve the coding performance, Jiang *et al.* proposed a new RC method in [15] with a linear R-Q model based on RC at MB level to improve the model parameter estimation accuracy. However, there are still errors in bitrate estimation due to the inaccuracy of MB level RC model.

In Fig. 1, the basic process of RC in HEVC is illustrated. In the GOP level [34], [35] , the bit allocation is performed given the information of buffer status and total bit-rate budget. In analogous to the GOP level, the frame [36]–[38] and CTU level bit allocation is achieved based on the weighting factors of each picture and CTU, respectively. Due to the importance of CTU-level RC, which can greatly influence the Rate-Distortion (R-D) performances, various RC algorithms at the CTU level have been proposed for HEVC. In [16], an optimized CTU-level RC strategy was proposed by Li *et al.*. Wang *et al.* [17] proposed a RC scheme based on Lagrange multiplier, which greatly improved the coding efficiency. In [20], Zhou *et al.* proposed a novel CTU-level RC method based on content complexity correlation for HEVC. However, these methods are optimized based on MSE, which may not be optimal in terms of perceptual quality.

III. Perceptual Rate Distortion Modelling

In this Section, we will discuss the relationship between rate and perceptual distortion at the CTU level. The divisive normalization framework [27] is introduced to characterize the mean square error (MSE) and perceptual distortion, and a divisive normalized factor f is introduced to establish the correspondence,

$$D'(R) = D(R)/f^2, \tag{1}$$

where $D(R)$ refers to CTU level distortion in terms of the mean square error (MSE), and $D'(R)$ is the normalized distortion. To obtain the divisive normalization factors, each CTU can be divided into *l* subblocks for DCT transform, and the factor f is obtained from the SSIM index in DCT domain [27]:

$$f = \frac{\frac{1}{l}\sum_{i=1}^{l}\sqrt{\frac{\sum_{j=1}^{N_L-1}(U_i(j)^2+V_i(j)^2)}{N_L-1}+C_1}}{E\left(\sqrt{\frac{\sum_{j=1}^{N_L-1}(U(j)^2+V(j)^2)}{N_L-1}+C_1}\right)}. \tag{2}$$

Here, $E(\cdot)$ is the expectation operation in the whole frame. $U(j)$ and $V(j)$ denote the DCT coefficients of the input and reconstructed signals, and $U_i(j)$ and $V_i(j)$ are the corresponding *j*-th DCT coefficient in the *k*th subblock. Here, the DCT coefficients of the reconstructed signals are approximated by the original input signals as the frame has not been encoded when deriving the normalization factors. N_L is the subblock size and it is set to be 16. C_1 is the constant in accordance with the definition of SSIM index.

After the divisive normalization process, a perceptual distortion model is established, which can be further applied to the rate-distortion optimization and rate control process. The correlation between R and D is characterized in several *R-D* models. Among them, the famous ones are the exponential model [14], linear model [15], logarithmic model [6] and ρ model [13], etc. It is shown by analysis and experiments that the logarithmic model can well depict the relationship between R and D. To characterize the relationship between R and D, we adopt a typical logarithmic function

$$D'(R) = ln(c \times R^{-k}). \tag{3}$$

where *c* and *k* are model parameters depending on the video content.

We also demonstrate the effectiveness of the model based on experimental validations. In particular, the Low Delay B coding structure with a single reference image in HM-16.8 is used. The $R-D$ curve is fitted through the model, and we denote $Dn = 1/D'(R)$ to express the distortion of the luma component. Fig. 2 shows the relationships between rate and perceptual distortion for several typical sequences, where different coding QPs are employed. Here, the values of c and k are obtained by fitting the actual values with the model. It is obvious that the prediction accuracy is relatively high for different test sequences. Moreover, as shown in Table I, a series of representative sequences under different QPs are tested, and for different QPs the average Pearson correlation coefficient is around 0.94 between the predicted and actual values using the model. As such, it is adopted in our scheme.

According to (2), the CTU-level rate control is achieved by optimizing the overall distortion given the available bit rate allocated to the frame. In particular, the perceptual rate distortion cost *J* should be optimization,

$$J = \sum_{i=1}^{N} D'(R_i) + \lambda_c R_i, \tag{4}$$

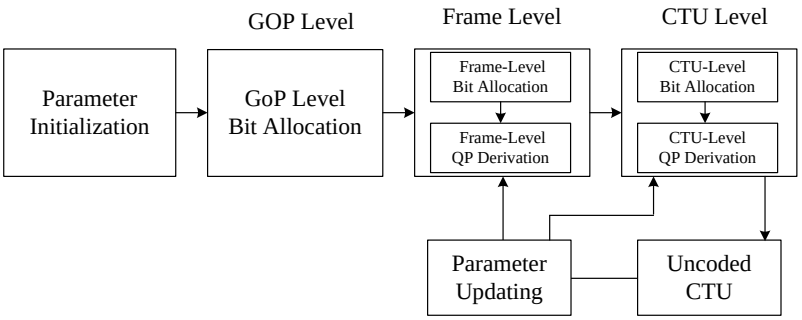


Fig. 1. The flowchart of the RC schemes in HEVC.

where λ_c is the language multiplier in HEVC, which is also used when the distortion is normalized with the divisive normalization strategy. $D'(R_i)$ is the perceptual distortion of the ith CTU, and *N* is the number of CTUs in one frame.

IV. Perceptual Rate Distortion Optimization for CTU Level Rate Control

We formulate the CTU-level bit allocation as a global optimization problem, and each CTU aims to improve the reconstruction quality by competing for the resources under the constraint of the target frame-level coding bits. In particular, the CTUs are coded dependently due to the constraint of the frame level coding bits, such that the allocation of each CTU can be performed by solving the optimization problem. The proposed scheme not only improves the coding efficiency in terms of perceptual rate-distortion, but also benefits the future R-D modelling with high accuracy.

Therefore, a global optimization approach with different CTUs has similar motivations to collaborate on bit allocation. However, the features of cooperation approach are different [18]. The global optimization approach aims to optimally allocate bits to each CTU given available resources. Given the allocation strategy, CTU_1, CTU_2, ..., CTU_N will receive the utility $R_1, R_2, \ldots, R_N$ respectively. Assuming the utility vector is one of the possible utility combination sets, and it is denoted as $U = (R_1^m, R_2^m, ..., R_N^m), m \in [0, \Omega]$, where Ω is the quantity of the possible combinations of utility. We define the minimum utilities of all CTUs as $U^* = (R_1^*, R_2^*, \ldots, R_N^*)$.

A. Global Optimization Approach for CTU-Level Bit Allocation

Based on the proposed rate and distortion models, the optimal bit allocation is investigated by minimizing average distortion. As such, bit allocation at CTU level is formulated as follows,

$$\begin{aligned} \{R_1^*, R_2^*, \ldots, R_N^*\} &= \text{argmin} \sum_{i=1}^{N} D'(R_i) \\ s.t. \sum_{i=1}^{N} R_i &\le R_c. \end{aligned} \tag{5}$$

where *N* and R_c are the number of CTU of one frame and frame-level bit rate respectively.

The constrained optimization problem can be converted into an unconstrained optimization problem as follows:

$$J = \sum_{i=1}^{N} D'(R_i) + \lambda \left(R_c - \sum_{i=1}^{N} R_i \right). \tag{6}$$

$$\begin{cases} \dfrac{\partial J}{\partial R_j} = -k_j \dfrac{1}{R_j} - \lambda = 0, \\ R_c - \displaystyle\sum_{i=1}^{N} R_i = 0. \end{cases} \tag{7}$$

Here, we will prove that the utility set *U* is non-empty and bounded, and the set of feasible utility *U* is convex such that the CTU-level rate control can be achieved by optimally bit rate allocation.

Theorem 1: *U* is a convex set.

Appendix I proves this theorem. The convexity of the utility function is satisfied.

B. Optimization Solution

To obtain the optimal coding bits for each CTU, we derive the appropriate Lagrange multiplier such that the perceptual distortion within a frame is minimized. Typically, Eqn. (6) is the minimal value of different function and convex function on convex set. Therefore, KKT condition ensures that the optimal solution is KKT point R^*. The optimal bits in equation (7) is obtained as follows,

$$R_j^* = \frac{k_j}{\sum_{i=1}^{N} k_i} R_c. \tag{8}$$

Proof:

Given Eqn.(7), we have,

$$\sum_{i=1}^{N} \frac{-k_i}{\lambda} = R_c. \tag{9}$$

Then the following relationship can be derived,

$$\lambda = \frac{\sum_{i=1}^{N} -k_i}{R_c}. \tag{10}$$

Subsequently, by taking Eqn.(7) into account, we have

TABLE I
CORRELATION COEFFICIENTS BETWEEN ACTUAL AND ESTIMATED VALUES FOR DIFFERENT CTUS.

Sequence	CTU index=3	CTU index=5	CTU index=7	CTU index=10	Avg
PeopleOnstreet(1600p)	0.9275	0.9865	0.9937	0.9916	0.9748
ParkScene(1080p)	0.9033	0.9477	0.9543	0.9689	0.9436
FourPeople(720p)	0.9126	0.9237	0.9379	0.9730	0.9368
BQMall(832x480)	0.9877	0.8194	0.9266	0.9301	0.9160
BQsquare(416x240)	0.9654	0.9538	0.9544	0.9466	0.9551
Avg	0.9393	0.9262	0.9534	0.9620	0.9452

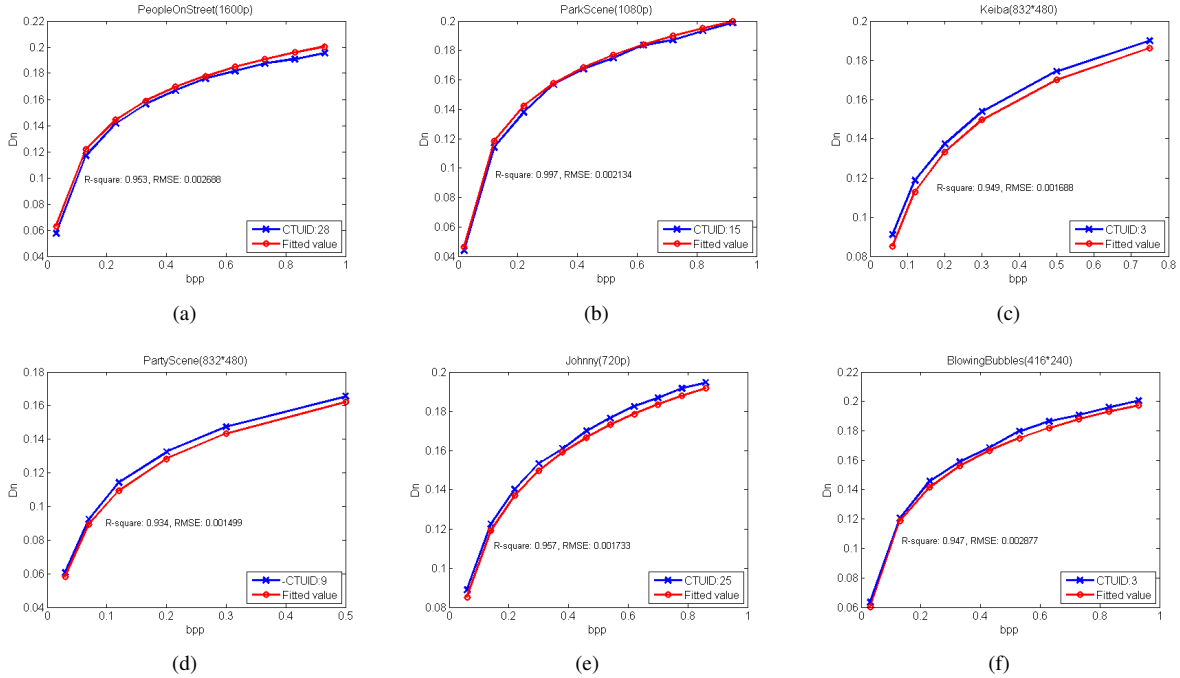


Fig. 2. Illustration of the actual and fitted R-D relationships.

Algorithm 1 The whole process of the CTU level RC.

Input: Frame-level bit rate
Output: QP of each CTU
Begin
Step1: Set $j = 0$;
Step2: If the current CTU is intra encoded, then HM intra CTU level rate control is performed;
Otherwise, compute the CTU level coding bits $R_j^* = \frac{k_j}{\sum_{i=1}^{N} k_i} R_c$;
Step3: Adjust the CTU level bit rate based on Eqn.(16);
Step4: Compute the CTU level QP;
Step5: Encode the jth CTU;
For each CU
Further adjust the CU-level QP based on divisive normalization [27][39];
End
Update the CTU level model parameters: k, using Eqn.(15);
Compute D_{real};
Step6: $j = j + 1$;
If $j > N - 1$, exit the loop;
Otherwise, jump to Step2.
End

$$\frac{\sum_{i=1}^{N} -k_i}{R_c} = \frac{-k_j}{R_j}. \tag{11}$$

$$R_j^* = \frac{k_j}{\sum_{i=1}^{N} k_i} R_c. \tag{12}$$

Accordingly, the target bits of jth CTU is determined by *Proof* end.

Here, the local optimal solution is obtained, and subsequently we prove that the local optimal solution is equivalent to the global optimization solution.

Theorem 2: U is a problem of convex quadratic optimization, and the local optimal solution U^* is also the global optimal solution. Appendix II proves this theorem.

C. Model Parameter Update

As it is difficult to obtain the parameter k relying on the CTU content before the encoding process, we estimate the optimal parameter k with an updating strategy. Here, assuming that the current to-be-encoded frame is i, and we aim to estimate the parameter based on the coding statistics of the $i-1$ frame. In particular, we assume that the coding distortion and rate of the co-located CTU are D_{real} and R_{real}, and we aim to minimize the difference between the true and estimated distortion D_{comp}, which can be expressed as squared error between the D_{real} and D_{comp},

$$e^2 = (D_{real} - D_{comp})^2. \tag{13}$$

As such, we can compute the derivative of e^2 to k,

$$\frac{\partial e^2}{\partial k} = \frac{\partial e^2}{\partial D_{comp}} \frac{\partial D_{comp}}{\partial k} = -2(D_{real} - D_{comp}) lnR. \tag{14}$$

Based on the Taylor's expansion, we have the updating strategy as follows,

$$\begin{aligned} k_{new} &= k_{old} - \delta(-2(D_{real} - D_{comp}))lnR \\ &= k_{old} + \delta_k(D_{real} - D_{comp})lnR. \end{aligned} \tag{15}$$

Basically, we can use Eqn.(1) to estimate the true distortion D_{real}. This study focuses on the modeling of videos with consistent quality. The distortion between two adjacent frames is of great importance to control the consistent quality. The distortion of the current CTU is similar to the co-located position of previous frame. Therefore, to reduce the complexity, distortion of co-located CTU is used to obtain D_{comp}.

It should be noted that δ_k in Eqn.(15) can be adaptive to the video content, and the model parameters between two consecutive frames are of great importance to achieve quality control in video coding. Regarding the rate control that produces videos with consistent quality, the model parameters are better to be consistent with the co-located position in the previous frame. Therefore, k_j of co-located CTU is used. As to the initial values of *k*, it is set to 2.3. It is worth mentioning that the initial values of k_j are not critical for CTU-level rate control, as the value will keep updating in the actual coding process. Fig. 3 shows the accuracy of *k*. From the experimental results, it can be seen that the proposed method can effectively estimates parameter.

After obtaining the parameter k_j and R_j^*, the CTU level target bit budget is finally adjusted as,

$$\begin{aligned} R_j^* &= R_j^* \times \omega_a, \\ \omega_a &= \left(1 - \frac{\sum_{p=1}^{j-1}(R_{act,p} - R_p)}{R_c}\right). \end{aligned} \tag{16}$$

Here, ω_a is an adjustment term to regularize the CTU level bit such that the frame-level budget can be met. R_{actp} and R_p are the real bits and the target bits after bit allocation, respectively. The corresponding QP can be obtained for each CTU through the R-Q model [12].

D. Proposed Rate Control Algorithm

This section summarizes the proposed CTU level rate control in algorithm 1. The proposed RC algorithm includes six steps as follows. It is worth noting that some procedures (such as bit allocations at frame level and GOP level, etc) are the same as those of $R-\lambda$ model. To improve the efficiency, divisive normalization based CU level adjustment for CU-level [27], [39] rate control was also adopted to further adjust the CU level QP as well as the corresponding Lagrange multiplier.

V. Experimental Results

In this section, we conduct experiments to compare our algorithm with the existing RC algorithms [16]–[18]. The experimental settings follow the HM LDB configuration [40]. Both Non-hierarchical (NoH) and hierarchical (H) encoding was involved in the experiment. The rate-distortion performance, rate control accuracy, subjective quality, smoothness of quality and complexity are analysed and compared.

A. R-D Performance Comparison

Here, we adopt the SSIM to compare the rate-distortion performance. The target bitrates are obtained based on compressing a sequence at fixed QP values. The values of QPs are set to be 37, 32, 27, and 22. The results are shown in Table II&III. Compared to the HM16.8 platform, the proposed scheme can achieve 16.3% (NoH) and 6.5% (H) bit rate savings on average. Moreover, it is observed that our algorithm performs better under both NoH and H configurations, and the NoH configuration can gain better performance as the H configuration leaves less room for us to improve. In Fig. 4, the overall rate-SSIM performance of our algorithm is compared with HM16.8, Li *et al.* [16], Wang *et al.* [17] and Gao *et al.* [18], and it is shown that the proposed method performs better compared to these methods in a wide range of bit rate.

B. Subjective Quality Comparison

In Fig. 5&Fig. 6, we compare the subjective quality of the RC methods under hierarchical configuration. Compared with HM16.8 and other state-of-the-art RC methods, the proposed method can produce better visual quality at similar bit rate. Experimental results also show that, compared to our scheme, other state-of-the-art RC methods are more likely to suffer from structural deformation, blocking effects as well as color artifacts, leading to lower visual quality. As a result, the visual quality is obviously degraded. Moreover, Fig. 5 shows that our scheme has better quality in the texture areas.

C. Quality Smoothness Comparison

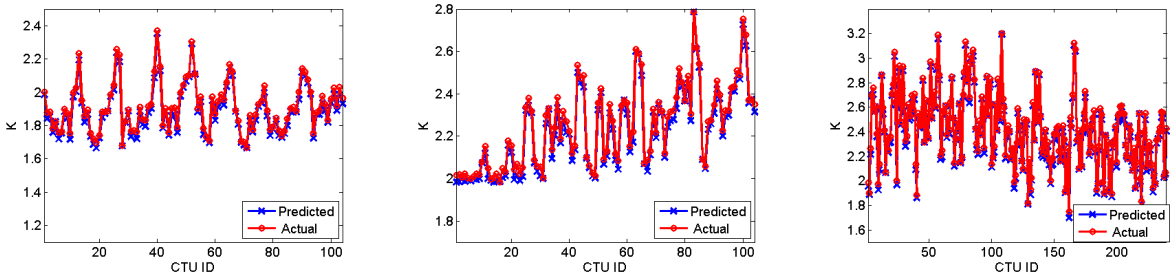
Quality smoothness is another factor influencing the visual quality of experience. In this subsection, the standard variance of SSIM is used to compute the quality smoothness, which

TABLE II
RD PERFORMANCE COMPARISON WITH DIFFERENT RATE CONTROL METHODS (NO-HIERARCHICAL).

Sequence	Ours vs. HM 16.8		Ours vs. Li *et al.* [16]		Ours vs. Wang *et al.* [17]		Ours vs. Gao *et al.* [18]	
	BD-Rate (%)	BD-SSIM	BD-Rate (%)	BD-SSIM	BD-Rate (%)	BD-SSIM	BD-Rate (%)	BD-SSIM
Class A	-12.7	0.006999	-12.3	0.006747	-9.2	0.004932	-8.4	0.004681
Class B	-14.7	0.007845	-10.3	0.006601	-7.3	0.004377	-6.3	0.003612
Class C	-17.1	0.011762	-11.7	0.006722	-8.7	0.004688	-7.6	0.004385
Class D	-26.9	0.018937	-18.2	0.012788	-13.7	0.007577	-12.3	0.006940
Class E	-10.4	0.006608	-10.1	0.005998	-7.1	0.004176	-6.1	0.003472
Average	-16.3	0.010370	-12.5	0.007771	-9.2	0.005150	-8.1	0.004618

TABLE III
RD PERFORMANCE COMPARISON WITH DIFFERENT RATE CONTROL METHODS (HIERARCHICAL).

Sequence	Ours vs. HM 16.8		Ours vs. Li *et al.* [16]		Ours vs. Wang *et al.* [17]		Ours vs. Gao *et al.* [18]	
	BD-Rate (%)	BD-SSIM	BD-Rate (%)	BD-SSIM	BD-Rate (%)	BD-SSIM	BD-Rate (%)	BD-SSIM
Class A	-5.9	0.003077	-5.8	0.002933	-3.6	0.002033	-3.3	0.001996
Class B	-4.3	0.002550	-4.5	0.002588	-2.9	0.001922	-2.7	0.001877
Class C	-4.8	0.002706	-5.5	0.002789	-3.7	0.002077	-3.5	0.001911
Class D	-12.1	0.006937	-8.0	0.004477	-6.2	0.003487	-5.9	0.003015
Class E	-5.2	0.002799	-5.1	0.002786	-3.1	0.001978	-2.9	0.001926
Average	-6.5	0.003614	-5.8	0.003115	-3.9	0.002299	-3.7	0.002145



(a) (b) (c)

Fig. 3. Comparisons between the actual and estimated model parameters (k). (a) The $60-th$ frame in BQMall(832x480) sequence; (b) The $96-th$ frame in Keiba(832x480) sequence; (c) The $200-th$ frame in Johnny(720p) sequence.

TABLE IV
COMPARISON OF S_SSIM UNDER DIFFERENT METHODS.

Sequence	HM 16.8		Li *et al.* [16]		Wang *et al.* [17]		Gao *et al.* [18]		Ours	
	NoH	H	NoH	H	NoH	H	NoH	H	NoH	H
Class A	0.0105	0.0116	0.0122	0.0133	0.0092	0.0103	0.0081	0.0096	0.0064	0.0074
Class B	0.0043	0.0052	0.0086	0.0094	0.0041	0.0047	0.0035	0.0039	0.0019	0.0021
Class C	0.0143	0.0164	0.0177	0.0188	0.0126	0.0132	0.0087	0.0099	0.0058	0.0065
Class D	0.0145	0.0166	0.0178	0.0189	0.0127	0.0129	0.0089	0.0108	0.0061	0.0093
Class E	0.0033	0.0043	0.0067	0.0099	0.0032	0.0040	0.0016	0.0020	0.0009	0.0011
Average	0.0094	0.0108	0.0126	0.0140	0.0084	0.0090	0.0062	0.0072	0.0042	0.0053

is also denoted as S_SSIM. Table IV shows that the quality smoothness of these methods. From the experimental results, it is observed that Li *et al.* [16] has the maximum fluctuation. Since Wang *et al.* [17] and Gao *et al.* [18] take the frame-coherence into account, they have achieved smoother quality when comparing to Li *et al.* [16] and HM 16.8. As the accuracy of the frame-coherence is also fully considered in our method, minimal fluctuations in terms of S_SSIM, which is 0.0042(NoH) and 0.0053(H), can be achieved. As such, our method also ensures quality smoothness compared with other state-of the-art algorithms based on the parameter inheritance scheme.

D. Buffer Occupancy Evaluation

Occupancy of buffer is an important factor in rate control, as the overflow and underflow should be avoided. Therefore, stable buffer occupancy is of great importance in evaluating RC performance. The buffer size is as shown as follow.

$$B_{uf} = D_{elay} \times T_{ar} \tag{17}$$

where D_{elay} is the delay time. T_{ar} is the bandwidth. The buffer occupancy is mainly determined by the target bits and actual bits [14]. Fig. 7 is two typical buffer occupancy under hierarchical configuration. As shown in Fig.7, the other RC methods, including Li *et al.* [16], Wang *et al.* [17], Gao *et al.* [18] and HM16.8, have higher buffer occupancy than our method. Generally speaking, our scheme can maintain lower

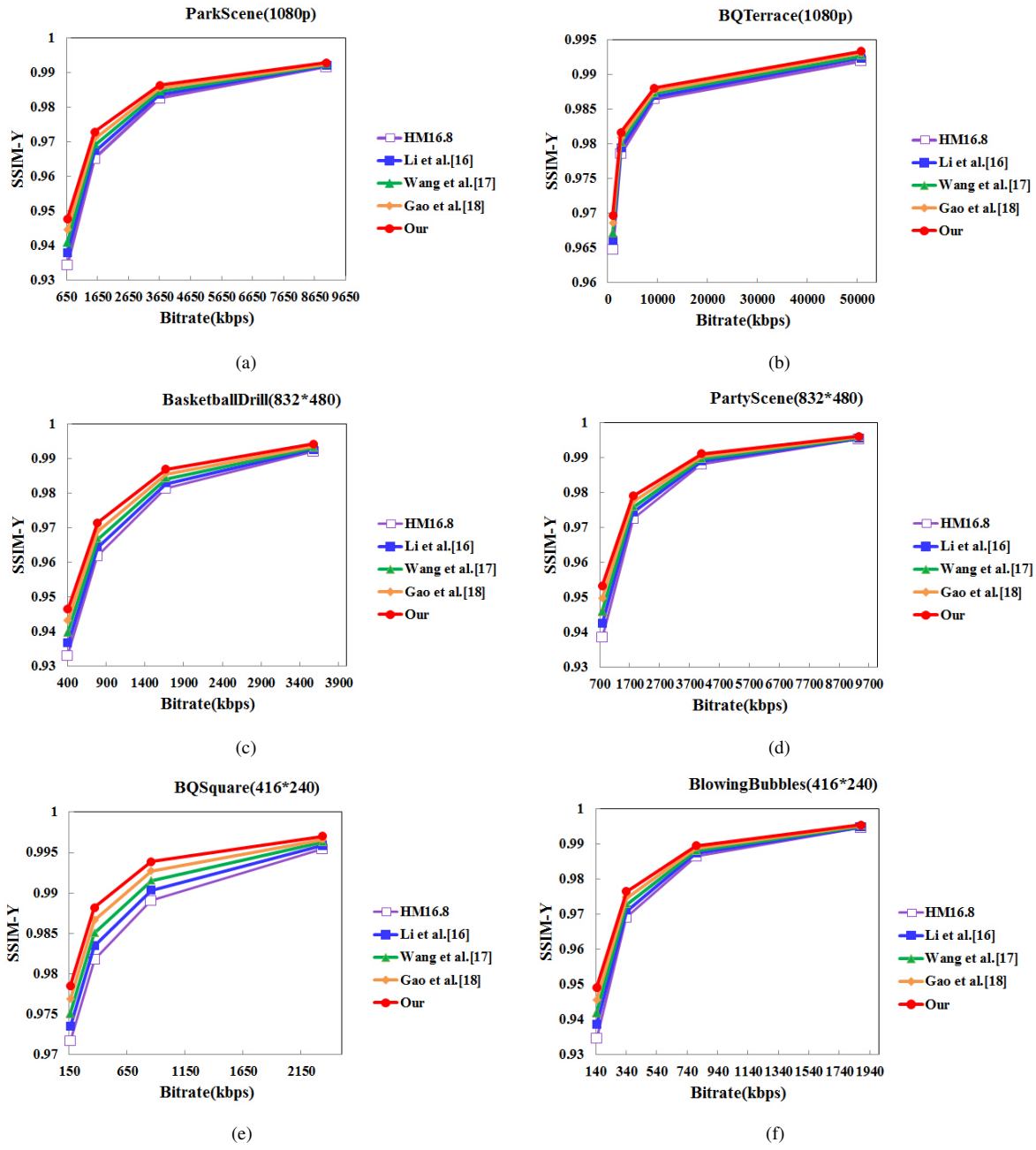


Fig. 4. R-D performance comparison between state-of-the-art RC schemes and our method under hierarchical configuration.

TABLE V
FRAME LEVEL RATE CONTROL ACCURACY COMPARISONS.

Sequence	HM 16.8		Li *et al.* [16]		Wang *et al.* [17]		Gao *et al.* [18]		Ours	
	NoH	H	NoH	H	NoH	H	NoH	H	NoH	H
Class A	0.21	1.16	0.18	1.03	1.98	9.77	0.41	2.15	0.38	2.09
Class B	0.53	2.53	0.49	1.98	1.77	5.01	0.72	3.07	0.56	2.98
Class C	0.31	1.52	0.26	1.32	1.86	9.63	0.09	1.18	0.30	1.33
Class D	0.54	2.56	0.47	2.44	3.22	10.11	0.69	3.17	0.58	2.99
Class E	1.47	6.83	1.43	5.37	0.98	3.67	1.78	5.09	1.68	4.94
Average	0.61	2.92	0.57	2.42	1.96	7.64	0.74	2.93	0.70	2.87

buffer, such that the stalling effects can be prevented for better quality of experience.

E. Bit Rate Accuracy and Complexity Comparisons

The accuracy of the bit rate at the frame level is investigated for mismatch error, which is calculated as follows,

(a) (b) (c) (d) (e) (f)

Fig. 5. Visual quality comparisons for the $120-th$ frame in BQSquare sequence (target bit rate: 160 kbps). (a) Original; (b) HM16.8 (actual bit rate: 161.11 kbps); (c) Li *et al.* [16] (actual bit rate: 160.77 kbps); (d) Wang *et al.* [17] (actual bit rate: 163.77 kbps); (e) Gao *et al.* [18] (actual bit rate: 161.72 kbps); (f) Our method (actual bit rate: 160.83 kbps).

$$Er = \frac{|R_{tar} - R_{act}|}{R_{tar}} \times 100\%. \tag{18}$$

where R_{act} and R_{tar} are the actual bit and the target bit at the frame level. TABLE V shows the bit rate errors, from which it can be seen the actual bit rates of our method is very close to the target bit rates. When compared with Wang *et al.*'s [17] and Gao *et al.*'s[18] algorithm, our scheme has higher accuracy. Moreover, although the mismatch error of our method is slightly higher than Li *et al.*'s method [17] and HM16.8, the difference is marginal.

Moreover, in Fig. 8, our method and the other four algorithms in HEVC are compared in terms of computational complexity, which is calculated as follows,

$$\Delta T = \frac{T_{pro} - T_{org}}{T_{org}} \times 100\%. \tag{19}$$

where T_{pro} and T_{org} are the encoding time of the proposed scheme and HM16.8 anchor. The results indicate that our algorithm is more complex than HM16.8, and the additional complexity of our method is close to Wang *et al.*'s algorithm [17].

VI. Conclusion

This paper proposes a coding tree unit level rate control scheme to improve the perceptual rate-distortion performance. The novelty of this paper lies in that perceptual rate-distortion model is established, and the optimization problem in CTU level rate control is solved with convex optimization to achieve the optimal performance. Extensive experimental results show that the perceptual rate-distortion performance is significantly improved based on the HEVC standard in terms of both objective and subjective evaluations.

(a) (b) (c) (d) (e) (f)

Fig. 6. Comparisons of subjective difference for the $180-th$ frame in PartyScene sequence (target bit rate: 780 kbps). (a) Original; (b) HM16.8 (actual bit rate: 780.33 kbps); (c) Li *et al.* [16] (actual bit rate: 780.11 kbps); (d) Wang *et al.* [17] (actual bit rate: 781.34 kbps); (e) Gao *et al.* [18] (actual bit rate: 780.95 kbps); (f) Our method (actual bit rate: 780.17 kbps).

APPENDIX A

Proof: The feasible utility set U for each element is a convex set and every element U_x in U ($U = (U_1, U_2, ..., U_m)$). According to the concept of convexity, to confirm that U is a convex set, it is necessary to prove that

$$\theta U_x + (1-\theta)U_y \in U; where\ 0 \leq \theta \leq 1. \tag{20}$$

Based on the rate-distortion model in Eq.(3), the distortion of the y_{th} is represented as

$$D(U_y) = ln(c \times U_y^{-k}) = lnc - klnU_y. \tag{21}$$

The utility function is defined on the basis of the distortion model. In accordance with the convex function definition, we have

$$f(U_x, U_y, \theta) = lnc - k \times ln(\theta U_x + (1-\theta)U_y). \tag{22}$$

To prove (20), $g(U_x, U_y, \theta)$ is denoted as

$$g(U_x, U_y, \theta) = \theta f(U_x) + (1-\theta)f(U_y) - f(\theta U_x + (1-\theta)U_y). \tag{23}$$

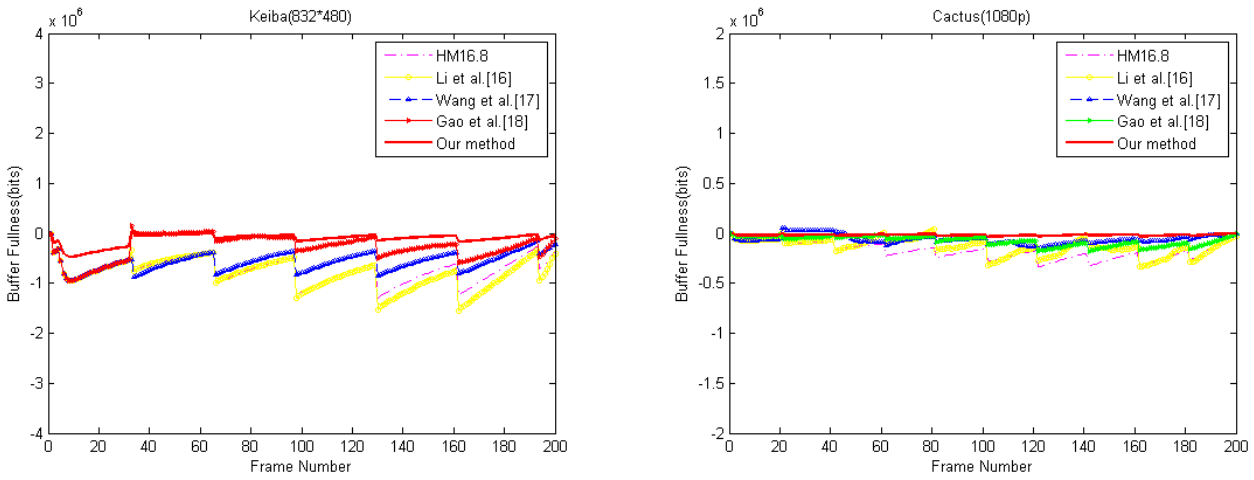
Then if $g > 0$ is proved for each element, (20) can be proved.

Eq. (23) can be further represented as

$$\begin{aligned} g(U_x, U_y, \theta) &= lnU_x^{-k\theta} + lnU_y^{-k\theta} - ln(\theta U_x + (1-\theta)U_y)^{-k} \\ &= -kln((\frac{U_x^{-\theta}U_y^{-\theta}}{\theta U_x + (1-\theta)U_y}). \end{aligned} \tag{24}$$

We know that $k > 0, 0 \leq \theta \leq 1, U_x > 0, U_y > 0$, therefore, $\frac{U_x^{-\theta}U_y^{-\theta}}{\theta U_x + (1-\theta)U_y} < 0$, $g(U_x, U_y, \theta) < 0$, and U is a convex set.

Proof end.



(a) (b)

Fig. 7. Comparisons of the buffer status for different rate control methods.

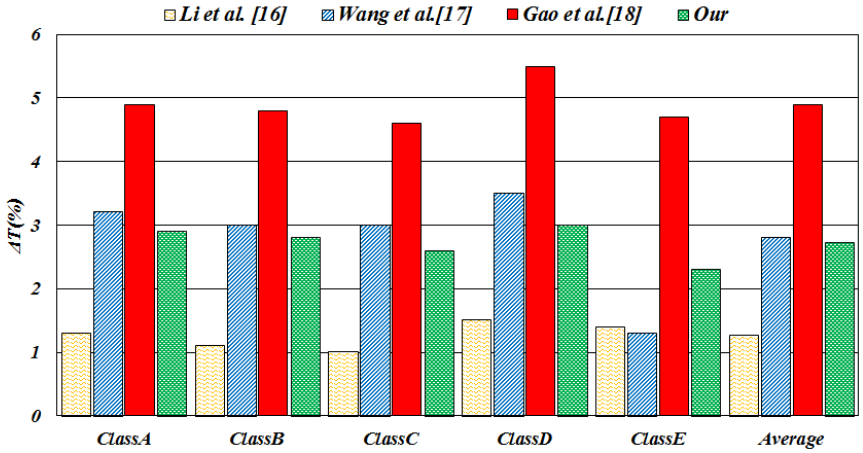


Fig. 8. Comparisons of the additional complexity.

APPENDIX B

All possible solutions U is a convex function which is defined on the convex set U. $U^* \in U$ is a local minimum off over U. U^* is a global minimum off over U.

Proof: U^*, local minimum off over U, and follows $r > 0$ such that $f(x) \geq f(U^*)$ for any $x \in U$ satisfying $x \in B[U^*, r]$.

Now let $\bar{U} \in U$ satisfy $\bar{U} \neq U^*$. Our objective is to show that $f(\bar{U}) \geq f(U^*)$. Let $\lambda \in (0, 1]$ such that $U^* + \lambda(\bar{U} - U^*) \in B[U^*, r]$. Such λ can be $r/\|\bar{U} - U^*\|$, since $U^* + \lambda(\bar{U} - U^*) \in B[U^*, r] \bigcap U$, it follows that $f(U^*) \leq f(U^* + \lambda(\bar{U} - U^*))$, and hence by inequality of Jensen

$$f(U^*) \leq f(U^* + \lambda(\bar{U} - U^*)) \leq (1-\lambda)f(U^*) + \lambda f(\bar{U}). \quad (25)$$

Thus $\lambda f(U^*) \leq \lambda f(\bar{U})$, which is followed by the desired inequality $f(U^*) \leq f(\bar{U})$. The above result is modified, which demonstrates that local minimum of strictly convex function over convex set is the strict global minimum.

Proof end.

REFERENCES

[1] B. Bross, "High efficiency video coding (hevc) text specification draft 9 (sodis)," in *11th JCT-VC meeting, Oct. 2012*, 2012.

[2] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

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

[4] B. Yan and M. Wang, "Adaptive distortion-based intra-rate estimation for h. 264/avc rate control," *IEEE Signal processing letters*, vol. 16, no. 3, pp. 145–148, 2009.

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

[6] N. Wang and Y. He, "A new bit rate control strategy for h. 264," in *Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, vol. 3. IEEE, 2003, pp. 1370–1374.

[7] S. Ma, W. Gao, F. Wu, and Y. Lu, "Rate control for jvt video coding scheme with hrd considerations," in *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, vol. 3. IEEE, 2003, pp. III–793.

[8] W. Lin, M.-T. Sun, R. Poovendran, and Z. Zhang, "Activity recognition using a combination of category components and local models for video

surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1128–1139, 2008.

[9] Z. He and D. O. Wu, "Linear rate control and optimum statistical multiplexing for h. 264 video broadcast," *IEEE Transactions on Multimedia*, vol. 10, no. 7, pp. 1237–1249, 2008.

[10] Y. Liu, Z. G. Li, and Y. C. Soh, "Adaptive mad prediction and refined rq model for h. 264/avc rate control," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 2. IEEE, 2006, pp. II–II.

[11] Z. Li, W. Gao, F. Pan, S. Ma, K. P. Lim, G. Feng, X. Lin, S. Rahardja, H. Lu, and Y. Lu, "Adaptive rate control for h. 264," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 376–406, 2006.

[12] H. Choi, J. Nam, J. Yoo, D. Sim, and I. Bajic, "Rate control based on unified rq model for hevc," *ITU-T SG16 Contribution, JCTVC-H0213*, pp. 1–13, 2012.

[13] X. Liang, Q. Wang, Y. Zhou, B. Luo, and A. Men, "A novel rq model based rate control scheme in hevc," in *Visual Communications and Image Processing (VCIP), 2013*. IEEE, 2013, pp. 1–6.

[14] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by r-lambda model for hevc," in *JCTVC-K0103, JCTVC of ISO/IEC and ITU-T, 11th meeting Shanghai, China*, 2012.

[15] J. Dong and N. Ling, "A context-adaptive prediction scheme for parameter estimation in h. 264/avc macroblock layer rate control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 8, pp. 1108–1117, 2009.

[16] S. Li, M. Xu, Z. Wang, and X. Sun, "Optimal bit allocation for ctu level rate control in hevc," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 11, pp. 2409–2424, 2017.

[17] M. Wang, K. N. Ngan, and H. Li, "Low-delay rate control for consistent quality using distortion-based lagrange multiplier," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 2943–2955, 2016.

[18] W. Gao, S. Kwong, and Y. Jia, "Joint machine learning and game theory for rate control in high efficiency video coding," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 6074–6089, 2017.

[19] D. Zhao, Y. Zhou, D. Wang, and J. Mao, "Effective macroblock layer rate control algorithm for h. 264/avc," *Computers & Electrical Engineering*, vol. 37, no. 4, pp. 550–558, 2011.

[20] M. Zhou, Y. Zhang, B. Li, and X. Lin, "Complexity correlation-based ctu-level rate control with direction selection for hevc," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 13, no. 4, p. 53, 2017.

[21] W. Gao, S. Kwong, Y. Zhou, and H. Yuan, "Ssim-based game theory approach for rate-distortion optimized intra frame ctu-level bit allocation," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 988–999, 2016.

[22] 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 Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 3, pp. 465–479, 2014.

[23] M. Karczewicz and X. Wang, "Intra frame rate control based on satd," in *JCT-VC M0257, 13th Meeting*, 2013.

[24] B. Li, H. Li, L. Li *et al.*, "Adaptive bit allocation for r-lambda model rate control in hm," in *JCTVC M0036, 13th Meeting of Joint Collaborative Team on Video Coding of ITU-T SG1 6WP3 and ISO/IEC JTC1/SC*, vol. 29, 2013.

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

[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[27] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Perceptual video coding based on ssim-inspired divisive normalization," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1418–1429, 2013.

[28] S. Wang, A. Rehman, K. Zeng, J. Wang, and Z. Wang, "Ssim-motivated two-pass vbr coding for hevc," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2189–2203, 2017.

[29] S. Wang, A. Rehman, Z. Wang, S. Ma, and W. Gao, "Ssim-motivated rate-distortion optimization for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 4, pp. 516–529, 2012.

[30] B. H. K. Aswathappa and K. Rao, "Rate-distortion optimization using structural information in h. 264 strictly intra-frame encoder," in *System Theory (SSST), 2010 42nd Southeastern Symposium on*. IEEE, 2010, pp. 367–370.

[31] C.-L. Yang, R.-K. Leung, L.-M. Po, and Z.-Y. Mai, "An ssim-optimal h. 264/avc inter frame encoder," in *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*, vol. 4. IEEE, 2009, pp. 291–295.

[32] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 6, pp. 742–752, 2005.

[33] Z. Luo, L. Song, S. Zheng, and N. Ling, "H. 264/advanced video control perceptual optimization coding based on jnd-directed coefficient suppression," *IEEE transactions on circuits and systems for video technology*, vol. 23, no. 6, pp. 935–948, 2013.

[34] B. Li, H. Li, L. Li, and J. Zhang, "$lambda$ domain rate control algorithm for high efficiency video coding," *IEEE transactions on Image Processing*, vol. 23, no. 9, pp. 3841–3854, 2014.

[35] M. Zhou, H.-M. Hu, and Y. Zhang, "Region-based intra-frame rate-control scheme for high efficiency video coding," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. IEEE, 2014, pp. 1–4.

[36] J. Si, S. Ma, X. Zhang, and W. Gao, "Adaptive rate control for high efficiency video coding," in *Visual Communications and Image Processing (VCIP), 2012 IEEE*. IEEE, 2012, pp. 1–6.

[37] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-gop based rate control for high efficiency video coding," *IEEE Journal of selected topics in signal processing*, vol. 7, no. 6, pp. 1101–1111, 2013.

[38] M. Zhou, B. Li, and Y. Zhang, "Content-adaptive parameters estimation for multi-dimensional rate control," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 204–218, 2016.

[39] S. Wang, S. Ma, D. Zhao, and W. Gao, "Lagrange multiplier based perceptual optimization for high efficiency video coding," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. IEEE, 2014, pp. 1–4.

[40] "Hm reference software 16.8. available:http://hevc.hhi. fraunhofer.de/svn/svn hevcsoftware."