Rate Control for Consistent Video Quality with Inter-Dependent Distortion Model for HEVC

Yuan Li, Huizhu Jia, Xiaodong Xie and Tiejun Huang*

National Engineering Laboratory for Video Technology
Peking University, Beijing, China
{yuanli, hzjia, donxie, tjhuang}@pku.edu.cn

Abstract—**Consistent video quality is important for video coding applications, which is also a popular optimization target for rate control. In this paper, a rate control scheme is proposed to reduce the fluctuation of video quality. First, we set up the optimization formulation of distortion and derive the distortion model by analyzing quadtree-based coding unit (CU) structure in high efficiency video coding (HEVC). Then the frame bit allocation algorithm is proposed by considering the inter-dependency. After the frame basic quantization parameter (QP) is obtained, the solution of optimization formulation finally regulates QP to maintain the consistent video quality. Experimental results show that the proposed rate control scheme can reduce the fluctuation of video quality up to 92.9% than benchmark, where the average reduction is 74.1%.**

Index Terms—**Rate control, HEVC, Consistent video quality, Inter-dependent distortion model, QP decision**

I. Introduction

Rate control is developing with the improvement of video codecs though it is not specified by the video coding standards. In various video coding applications, rate control is employed to regulate the bit stream to adapt the limited transmission channel. Except achieving the target bit rate accurately, some other optimizations also need to be considered in the rate control process, such as minimizing the total distortion, reducing the system latency, maintain the consistent video quality etc.

For the purpose of video quality optimization, previous rate control work can be classified into two categories. The first category of rate control scheme aims to achieve the minimum total distortion subject to a total bit rate constraint, which is also the classical problem for rate control. Various technologies have been proposed for this purpose. In [1] and [2], the relationship between QP and rate-distortion optimization (RDO) is studied to improve the R-D performance. Li et al. [1] proposed a linear model to predict the coding complexity of the current basic unit, which was used to obtain the QP before the RDO process. While in [2], Ma et al. proposed a partial two-pass scheme for RDO process at macroblock (MB) level to make the mode selection more accurate. Besides, R-D optimized bit allocation (BA) algorithms are also implemented widely [3-4]. Both [3] and [4] explored the BA of hierarchical structure in HEVC, where the basic concept was to ensure the high video quality of the key frames. In addition to minimizing total distortion, maintaining the consistent video quality to please human's visual experience, which is the second category of optimizing target for rate control, is also important to the implementation of video codecs. In Xie's work [5], a frame-level BA framework with a rate-complexity model was proposed to track the nonstationary characteristics, which achieved the smooth video quality. In [6], rate control scheme on the concept of "window" was proposed to control the fluctuation of distortion. In this paper, we also target to the second category of optimizing purpose.

To achieve the bit budget accurately, R-D models are usually adopted to decide the QP in rate control scheme. With the using of quadtree-based CU structure in HEVC, R-D models need to be further developed to match the relationship between bit rate and QP more accurately. In Choi's work [7], a quadratic pixel-based unified rate-quantization (URQ) model was proposed. This pixel-level model could be used for any size of rate control level. Later, in [8], Li et al. proposed an R-λ model and the corresponding rate control scheme, which achieved smaller bit rate errors and was adopted as the recommended rate control algorithm in HM software. In [9], a texture rate model was proposed to track the different characteristics of transform coefficients for different depth of CUs. The above-mentioned R-D models for HEVC are established mainly based on the assumption that the coding units between inter frames are encoded independently. However, since HEVC still adopts prediction coding structure in inter frames, the R-D models should take the inter frame coding dependency into consideration for more general situation. Besides, the effect of the quadtree-based CU structure to R-D modeling should also be taken into account.

In this paper, we first analyze the problem of maintaining the consistent video quality and establish the formulation. Then the inter-dependent distortion model is derived by considering the depth of quadtree-based CU structure. Based on the distortion model, a rate control scheme is proposed to reduce the fluctuation of the video quality. Experimental results show that the proposed rate control scheme achieves more consistent video quality than benchmark with negligible BD-Rate increased.

The rest of this paper is organized as follows. Section II describes the optimization problem formulation and distortion

* the corresponding author, Huizhu Jia is with Peking University, also with Cooperative Medianet Innovation Center and Beida(Binhai) Information Research.

model derivation. In Section III, the rate control scheme is proposed aim to reduce the fluctuation of the video quality. Section IV shows the experimental results and the conclusion is given in Section V.

II. Problem Formulation and Distortion Modeling

To minimize the fluctuation of the video quality, variation of distortions generated by the encoding process should be small. In the literature, the global variation of distortions for all frames in a sequence is usually used as the optimization goal for consistent video quality. This optimization goal is established based on the assumption that the content of video sequence is continuous without abrupt scene change. However, in some applications such as movies or sport news, scene change occurs frequently. When scene change happens, the attention of audience will change to the new scene which means that the effect of previous video quality will become weak. In this situation, the global variation of distortions cannot follow the audience's experience well. Furthermore, optimizing the global variation of distortions need to obtain the characteristics of the whole video sequence to allocating bits, which is not suitable for the real-time video coding applications. Hence we propose a local distortion variance (LDV) model to measure the fluctuation of video quality as follows:

$$LDV(i) = \mathrm{var}(D_i^*), \quad \text{where } D_i^* = \{D_{i-N}, D_{i-N+1}, \ldots D_{i+N}\} \tag{1}$$

where D_i^* is a set of distortions including the distortion of i-th frame, the distortions of N frames before i-th frame and the distortions of N frames after i-th frame. N is a constant. LDV model describes the fluctuation of the video quality for any adjacent $2N$+1 frames with the concept of sliding "window". The fluctuation of distortions in this $2N$+1 sized window is more suitable to track the actual visual experience than the global variance of distortions. In LDV model, D can be set as any measurement of distortion such as sum of absolute difference (SAD), mean squared error (MSE), structural similarity (SSIM) index etc. In this paper, we use MSE for the convenient comparison with other rate control works.

As described in Section I, the quantization step size (denote as Q in below) is the control factor of both distortion and bit rate in video coding process. The relationship between distortion and Q, which is called D-Q model, need to be analyzed. Like H.264/AVC, quantization process is adopted, in HEVC, to transform coefficients which are converted from the prediction residuals after the transform process. The distribution of transform coefficients in H.264/AVC is usually modeled with Laplacian, Cauchy, and Generalized Gaussian in the literature. However, the quadtree-based CU structure adopted in HEVC changes the characteristics of transform coefficients. [10] proposed a Mixture Model based on Laplacian (LMM) probability density function (PDF) to reflect the transform coefficients according to different depth of CU as

$$f_{LMM}(x) = \sum_{j=0}^{N_{dep}-1} \theta_j f_{CU_j}(x) \tag{2}$$

where N_{dep} is the total number of CU depth, θ_j is a weighting factor, CU_j denotes the j-th depth of CU, $f_{CU_j}(x)$ is the Laplacian PDF of CU_j as

$$f_{CU_j}(x) = \frac{\lambda_j}{2} e^{-\lambda_j |x|}. \tag{3}$$

In (3), λ_j denotes the model parameter of CU_j and can be derived as $\lambda_j = \sqrt{2}/\sigma_j$, where σ_j is the standard deviation of the transform coefficients for CU_j. Inspired by LMM, where PDF of transform coefficients is highly correlated with CU depth, the inter-dependent relationship between distortion and predicted residual [11] is proposed as

$$MAD\_R_{j,i} = MAD\_O_{j,i} + k_j \sqrt{D_{j,i-1}} + t_j \tag{4}$$

where $MAD\_R_{j,i}$ denotes the mean SAD of j-th depth CU in i-th frame between original frame and predicted frame, $MAD\_O_{j,i}$ denotes the mean SAD between two original frames, $D_{j,i-1}$ is the distortion of j-th depth CU in i-1-th frame, k_j and t_j are model parameters associated with depth j. (4) represents the relationship between the predicted residual and the distortion of reference frame. Some video sequences are tested with various QPs to verify (4) and the results are shown in Fig. 1. We can see that the R^2 is larger than 0.99, which means that the linear relationship is matched well.

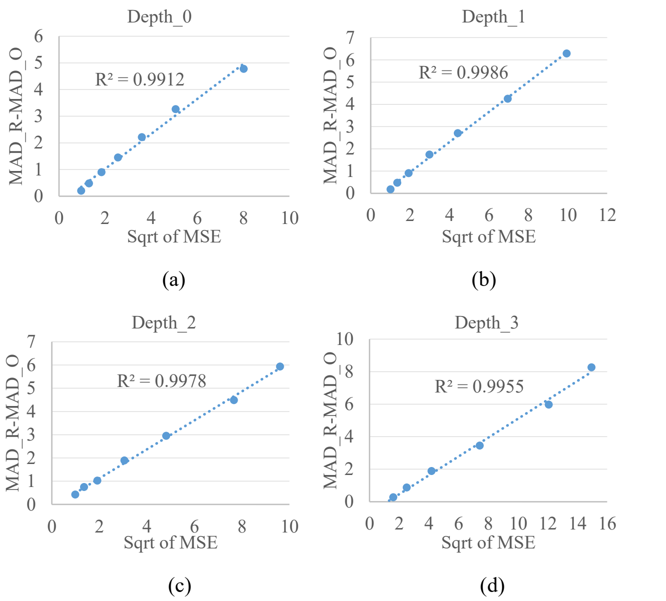


Fig. 1. Verification of (4). "*BasketballPass*" sequence, POC 2, QP from 12 to 42.

From [11], we can know that the distortion of current frame has a quadratic proportion with the MAD of current frame based on the Laplacian distribution. Combining with (2) and $\sigma=\sqrt{2}$MAD [11], we can get the distortion of each depth as

$$D_{j,i} = a_j (MAD\_R_{j,i}^2 + Q_i) + b_j \tag{5}$$

where $D_{j,i}$ and $MAD\_R_{j,i}$ has the same meaning with in (4), Q_i is based quantization step size of i-th frame, a_j and b_j are model parameters depending on the depth j. Substituting (4) into (5), the inter-dependent D-Q model can be approximated as

$$D_{j,i} = \alpha_j (Q_i + MAD\_O_{j,i}^2 + k_j^2 D_{j,i-1}) + \beta_j \tag{6}$$

where α_j and β_j are model parameters. Based on (6), the total distortion of i-th frame is

$$D_i = \tau_j \sum_{j=0}^{N_{dep}-1} D_{j,i} \tag{7}$$

where τ_j is weighting factor. (7) is the inter-dependent D-Q model with the quadtree-based CU structure considered and can be used to derived Q for consistent video quality, which will be described in Section III.

III. Proposed Rate Control Scheme

In this section, we propose a rate control scheme for video quality optimization. The target bit in a GOP, denoted as T_{GOP}, is obtained using the method in [8], which is also the recommended method in recent HM software. Our work mainly focuses on the frame level BA and the quality optimization according to the LDV model.

A. Frame Bit Allocation

Frame BA determines bit budget for each frame of GOP after the target bit in a GOP allocated. The target bit of current frame T_{Cur} is usually calculated as follows,

$$T_{Cur} = \frac{T_{GOP} - T_{coded}}{\sum_{NotCoded} \omega_i} \cdot \omega_{Cur} \tag{8}$$

where T_{coded} denotes the bits of encoded frame in current GOP, ω_i is BA weighting factor of each frame in current GOP. Traditionally, ω_i is set up equally or at fixed ratio, which will cause the fluctuation of video quality for ignoring the video content and the coding dependency. In this work, by considering the coding complexity and inter dependency, ω_i is obtained as follows,

$$\omega_i = \frac{\sum_{j}^{depth} MAD\_O_{j,i} + k_j\sqrt{D_{i-1}} + t_j}{MAD\_R_{avg} + \sum_{j}^{depth}(MAD\_O_{j,i} + k_j\sqrt{D_{avg}} + t_j)} \tag{9}$$

where MAD_R_{avg} and D_{avg} denotes the average MAD_R and distortion of previous encoded M frames (M is set as 5), k_j and t_j are model parameters like in (4). MAD_O_j for different depth of current frame are obtained from the pre-analysis process. Firstly, CU size of original frame is determined using algorithm in [12] effectively, which takes the content-dependency of depth level into consideration to pre-decide the CU depth. Then, motion search is performed to get MAD_O_j according to the CU depth. The computational complexity of pre-analysis is much lower than the complete encoding process. When current frame BA done, we use R-λ model [8] to derive the basic QP of current frame.

B. Optimization for Consistent Video Quality

LDV model described in Section II is used to optimize the fluctuation of video quality. For low delay purpose, we only use the distortion of current frame and previous frames to set up the optimization formulation with a slight modification of LDV model as

$$\min \operatorname{var}(D_i) = \frac{1}{2N+1}\sum_{j=i-2N}^{i}(D_j - D_{avg})^2 \tag{10}$$

where D_{avg} is the average distortion of the current frame and the previous $2N$ frames. From (10), we can get that

$$\operatorname{var}(D_i) = \frac{1}{2N+1}\left(\sum_{j=i-2N}^{i-1}(D_j - D_{avg})^2 + (D_i - D_{avg})^2\right) \tag{11}$$

Since the distortion of previous $2N$ frames are known, the solution of D_i can be obtained by differentiation. Then the distortion of current frame D_{Cur} can be derived as

$$D_{Cur} = \frac{1}{2N}\sum_{j=i-2N}^{i-1} D_j \tag{12}$$

which means that to get the optimal solution of LDV model, the distortion of current frame should be equal to the average distortion of previous $2N$ frames. Using D-Q model (7), QP for consistent video quality, denoted as QP_D, can be obtained. QP_D is used to regulate the frame basic QP from Section III.A by setting the QP range as [QP_D-2, QP_D +2].

C. Proposed Rate Control Scheme

The proposed rate control scheme for consistent video quality is represented completely as follows,

Step 1. Do GOP level BA using method in [8] to get T_{GOP};

Step 2. Do frame level BA for current frame using (9) and (8) to get T_{Cur};

Step 3. Calculate frame basic QP (QPR) from T_{Cur} using R-λ model [8];

Step 4. Calculate D_{Cur} by (12) and QP_D by (7), then regulate QP_R with range of [QP_D-2, QP_D +2] to get final QP of current frame QP_{Cur};

Step 5. Encode current frame with QP_{Cur}, then update model parameters with the data obtained after encoding by linear regression method;

Step 6. If current frame is the last frame in a GOP, then move to the next GOP and go to Step 1; else, move to the next frame and go to Step 2.

IV. Experimental Results

In order to evaluate the rate control scheme in Section III.C, we implement the scheme in HM 16.0. The testing configuration is set as low delay (LD) with 2 reference frames. The target of rate control is CBR and the target bit rate is generated with fixed QP sets {22, 27, 32, 37}. The rate control in HM 16.0 [8] is used as benchmark. The experimental results of bit rate error and BD-Rate are shown in Table I, where the bit rate error is measured as

$$R_{error} = \frac{\left|R_{target} - R_{actual}\right|}{R_{target}} \times 100\% \tag{13}$$

TABLE I

Bit Rate Error and Coding Performance Comparison

Sequences	R_{error}		BD-Rate
	R-λ [8]	Proposed	
ClassA	0.09%	0.11%	0.9%
ClassB	0.13%	0.02%	-1.0%
ClassC	0.05%	0.19%	2.5%
ClassD	0.13%	0.18%	1.9%
ClassE	0.11%	0.15%	1.3%
Average	0.10%	0.13%	1.1%

From Table I, we can see that both proposed rate control and [8] can achieve target bit rate accurately. For coding performance, the proposed rate control has a small BD-Rate increasing than [8] about 1.1% in average. That is because the proposed algorithm is aimed to maintain the consistent video quality which is not optimized for the coding efficiency. While [8] uses the hierarchical bit allocation, with the concept that the key frames should be allocated more bits, to obtain the coding gain. However, this will cause the fluctuation of video quality since few bits are allocated to the frames in higher layers.

The fluctuation of video quality is measured by the LDV model. N is set as 5. Table II shows the test results of average LDV and the LDV_{gain} is as follows

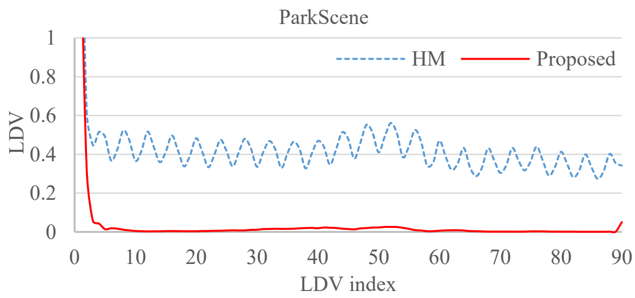
$$LDV_{gain} = \frac{\left|LDV\_avg_{proposed} - LDV\_avg_{HM}\right|}{LDV\_avg_{HM}} \times 100\% \quad (14)$$

TABLE II

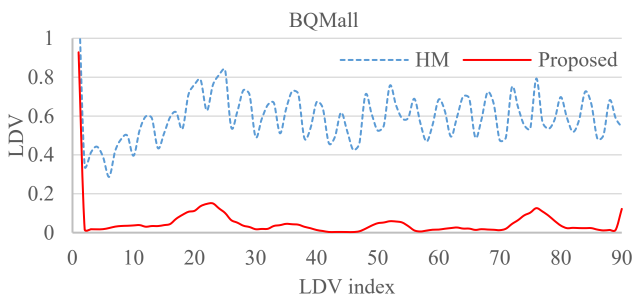
COMPARISON OF AVERAGE LDV FOR DIFFERENT SEQUENCES

Sequences	Bit rate (kbps)	LDV_avg		LDV_{gain}
		R-λ [8]	Proposed	
BasketballDrive	28353	0.54	0.10	81.5%
ParkScene	2004	0.42	0.03	92.9%
BQMall	3764	0.59	0.05	91.5%
PartyScene	1279	0.56	0.28	50.0%
BasketballPass	1549	0.57	0.31	45.6%
BlowingBubbles	496	0.59	0.10	83.1%
Average	-	-	-	74.1%

From Table II, we can see that the proposed rate control scheme can reduce the fluctuation of video quality significantly, especially in "*PartyScene*" sequence, the reduction ratio is up to 92.9%. The proposed frame level BA method with inter-dependency and the distortion optimization result in this gain.



(a)



(b)

Fig. 2. LDV comparison between HM and proposed rate control scheme. (a) "*PartyScene*" sequence, (b) "*BQMall*" sequence.

The detailed data of "*PartyScene*" sequence and "*BQMall*" sequence for each LDV index are shown in Fig. 2. Since [8] allocates more bits to the key frames and less bits to other frames, the LDV is fluctuated drastic, while the proposed is much smooth due to the D-Q model and inter-dependent BA.

V. CONCLUSIONS

In this paper, a rate control scheme with inter-dependent distortion model is proposed to reduce the fluctuation of video quality. First, we set up LDV model to measure the variance of distortion. Then the inter-dependent distortion model is derived and the frame BA algorithm is proposed. After the frame basic QP is obtained, the solution of optimization formulation finally regulates QP to maintain the consistent video quality. Experimental results show that the proposed rate control scheme can reduce the fluctuation of video quality up to 92.9% than benchmark, where the average reduction is 74.1%.

ACKNOWLEDGEMENTS

This work is partially supported by the National Science Foundation of China under contract No. 61502013, National High Technology Research and Development Program of China (863 Program) under contract No. 2015AA015903, National Key Technology Research and Development Program of the Ministry of Science and Technology of China under contract No. 2014BAK10B00, the Major National Scientific Instrument and Equipment Development Project of China under contract No. 2013YQ030967.

REFERENCES

[1] Z. G. LI, F. Pan, K. P. Lim, and G. N. Feng, "Adaptive basic unit layer rate control for JVT," in *Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-G012, 7th Meeting*, Thailand, 2003, pp. 7–14, JVT-G012.

[2] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE TCSVT.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[3] 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 Process.*, vol. 7, no. 6, pp. 1101-1111, Dec. 2013.

[4] S. Sanz-Rodriguez, and T. Schierl, "A rate control algorithm for HEVC with hierarchical GOP structures," in *Proc. IEEE ICASSP*, Vancouver, May 2013, pp. 1719-1723.

[5] B. Xie, "A sequence-based rate control framework for consistent quality real-time video," *IEEE TCSVT*, vol. 16, no. 1, pp. 56–71, Jan. 2006.

[6] Y. Li, H. Jia, C. Zhu, M. Li, X. Xie, and W. Gao, "Low-delay window-based rate control scheme for video quality optimization in video encoder," in *Proc. IEEE ICASSP*, Florence, May 2014, pp. 7333-7337.

[7] H. Choi, J. Nam, J. Yoo, D. Sim, and I. V. Bajić, "Rate control based on unified RQ model for HEVC," in *Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG118th Meeting*, San José, CA, USA, Feb. 1–10, 2012.

[8] B. Li, H. Li, L. Li, and J. Zhang, "Rate control by R-lambda model for HEVC," in *Proc. 11th Meeting Joint Collaborative Team Video Coding (JCT-VC)of ITU-T1/SC 29/WG 11*, Shanghai, China, Oct. 10–19, 2012.

[9] 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 TCSVT*, vol. 24, no. 3, pp. 465-479, Mar. 2014.

[10] 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, pp. 571–574, Oct. 2011.

[11] Y. Li, H. Jia, C. Zhu, M. Yang, X. Xie, and W. Gao, "Window-based rate control for video quality optimization with a novel INTER-dependent rate-distortion model," *Journal of Signal Process.: Image Comm.*, vol. 29, issue 10, pp. 1046-1062, Nov. 2014.

[12] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE TMM*, vol. 15, no. 2, pp. 465-470, Feb. 2013.