CONTENT ADAPTIVE QUANTIZATION PARAMETER CASCADING FOR RANDOM-ACCESS STRUCTURE IN HEVC

Kaifang Yang[1,2], Shuai Wan[2], Yanchao Gong[3,2], Yan Feng[2]

[1]School of Computer Science, Shaanxi Normal University, Xi'an, China
[2]School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China
[3]School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, China

ABSTRACT

In high-efficiency coding (HEVC), the random-access structure (RAS) is employed due to its high coding efficiency and "random-access" performance. Pictures in RAS are assigned to different temporal layers. And how to select the QP for each temporal layer in quantization parameter cascading (QPC) technique is critical for improving the coding efficiency of RAS. In order to further improve the coding efficiency of RAS, a QPC technique considering the video content characteristics (denoted as VC-QPC for short) is proposed. In VC-QPC, motion and texture complexities of a video are used for predicting the optimal quantization parameter values of temporal layers. Compared with the method in the test model of HEVC, i.e., HM14.0, the BD-rate of proposed VC-QPC is -5.60% for RAS.

***Index Terms*—**Random access structure, quantization parameter cascading, HEVC, video coding

1. INTRODUCTION

Compared with H.264/AVC [1], the latest video coding standard, i.e., high-efficiency video coding (HEVC), can achieve approximately 50% bitrate reduction with an equivalent subjective video quality [2], [3]. In the test model of HEVC, i.e., HM, three coding structures are recommended [4], [5], where the random-access structure (RAS) has the highest rate-distortion (R-D) performance due to the used effective inter-frame prediction. In the RAS, intra-coded pictures are inserted periodically in order to achieve "random-access" performance [4], [5], which makes it suitable for applications of video on demand as well as similar fields.

As shown in Fig.1, pictures within a group of picture (GOP) are organized into temporal layers in RAS, where pictures in different temporal layers are of different importance in terms of prediction. The coding efficiency of the RAS is closely related to the quantization parameter (QP) selected for coding pictures in different temporal layers (i.e., the QP cascading technique, referred to as QPC in this paper). Pictures in a lower temporal layer are of more importance in prediction since they will be directly or indirectly referenced by pictures in higher temporal layers. So pictures in a lower temporal layer has greater impact on the overall R-D performance. And smaller QP is always preferred for pictures in a lower temporal layer. Therefore, the general idea of QPC techniques [4]-[9] for the RAS is that the QP is assigned in an increasing order with the increase of temporal layers as shown in (1).

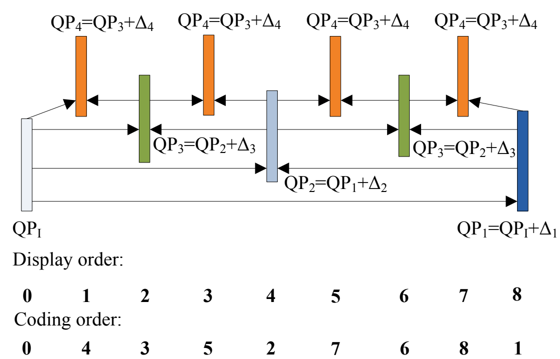

Fig. 1. An example of the RAS with GOP size being 8.

$$QP_l = \begin{cases} QP_I + \Delta_l, l = 1 \\ QP_{l-1} + \Delta_l, l > 1 \end{cases} \quad (1)$$

where QP_I is the QP of the intra-coded pictures, QP_l stands for the QP of the l^{th} temporal layer and Δ_l stands for the QP offset between neighbouring temporal layers. Generally, QP_I is set in the coding configuration file [4], [5]. Therefore, finding the optimal Δ_l for each temporal layer, denoted by Δ_l^*, is the key of different QPC techniques [4]-[9].

Currently, the QPC technique in [4], [5] with $\Delta_l^*, l \geq 1$ set as 1 is adopted in HM. In [6], QP is determined using a QP-lambda model. And an adaptive QPC technique is also proposed in [7]-[8]. However, those QPC techniques in [4]-[8] are all not optimized in terms of R-D performance.

In our previous work [9], following the spirit of Lagrangian optimization and based on the experimental observation, two QPC techniques, i.e., RDO-QPC and SRDO-QPC were proposed. Compared with the QPC techniques in [4]-[8], RDO-QPC in [9] has the highest R-D performance, however, 6 times of pre-encoding was needed for obtaining the model parameters, which limits its application in practice. SRDO-QPC involves no pre-encoding while makes a compromise between R-D performance and complexity. However, video content characteristics were ignored in SRDO-QPC, which therefore still has space for performance improvement.

In this paper, a content adaptive QPC technique (denoted as VC-QPC for short) is proposed by taking motion and texture characteristics of video sequences into consideration. VC-QPC is simple and efficient. Section 2 introduces the relationship between video content complexity and Δ_1^*. Section 3 quantitatively evaluates the video content complexity degree. The VC-QPC is proposed in Section 4 with its performance evaluation provided in Section 5. Conclusions are drawn in Section 6.

2. RELATIONSHIP BETWEEN VIDEO CONTENT COMPLEXITY AND Δ_1^*

In our previous work [9], it was observed that the $\Delta_l^*, l \geq 2$ all set to 1 has the highest or near highest R-D performance. And, Δ_1^* is the key to find the optimal QP for pictures in different temporal layers. Therefore, in this paper, in order to predict Δ_1^*, the relationship between motion complexity, texture complexity and Δ_1^* is firstly analyzed based on the experimental observation.

Table 1. Videos used in the experimental test.

Sequences (Resolution)	Total frame number	Motion complexity	Texture complexity
RaceHorse (416×240)	300	High	High
BasketballDrill (832×480)	500	Medium	Medium
Vidyo3 (1280×720)	600	Low	Low

Three video sequences, i.e., RaceHorse (WQVGA), BasketballDrill, and Vidyo3, with different motion and texture complexity were used in the experiment as shown in Table 1. And the main encoding parameters are listed in Table 2. HM14.0 [4], [5] was used for encoding. RAS was the coding structure with intra-coded pictures encoded at approximately one second interval [5], [10]. GOP size and search range were 8 and 64, respectively. QP_I were set to 22, 27, 32 and 37, respectively. Δ_1 were set to 1, 2,…, 10 and $\Delta_l, l \geq 2$ were all set to 1. Other parameters were set according to the default configuration of *encoder_randomaccess_main* [4], [5].

Table 2. Main encoding parameters.

Codec	HM14.0 [4], [5]
Encoding structure	RAS
Profile	Main
Intra period	Approximately one second [5], [10]
GOP size	8
Search range	64
QP_I	22, 27, 32, 37
Δ_1	1, 2,…, 10
$\Delta_l, l \geq 2$	1
SAO	ON

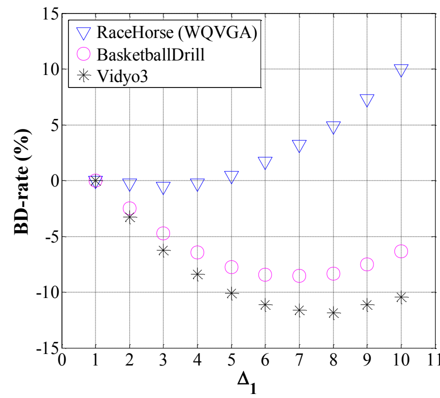

Fig. 2. Relationship between Δ_1 and BD-rate.

Encoding the video sequences using different combination of QP_I and Δ_1 under the RAS, and selecting the Δ_1 with the minimum BD-rate [11] as the Δ_1^*. For calculating BD-rate, $PSNR_{YUV}$ [12] where the luminance and chroma signals are all considered was used for representing distortion. Given the bit-rate and $PSNR_{YUV}$ values, BD-rate has been calculated using the piecewise cubic interpolation [10] and the QPC in the HM [4], [5] was used as the benchmark.

As shown in Fig. 2, the value of Δ_1^* were 3, 7 and 8 for RaceHorse (WQVGA), BasketballDrill and Vidyo3, respectively. From the experimental results, with the decrease of the video motion complexity, the value of Δ_1^* also decreases and with the increase of the video texture complexity, the value of Δ_1^* also increases. Therefore, Δ_1^* is closely related to the video content complexity. Therefore, the video motion and texture complexity degree will be used for predicting Δ_1^*.

1	1	1	1	1
1	2	2	2	1
1	2	0	2	1
1	2	2	2	1
1	1	1	1	1

Fig. 3. $h_{LP}[m_1, m_2]$ where m_1 varies from 1 (top) to 5 (down) and m_2 varies from 1 (left) to 5 (right).

3. MOTION AND TEXTURE COMPLEXITY DEGREE

Given an image $\mathbf{x}$ with width N_1 and height N_2 for $0 \le n_1 \le N_1 - 1$ and $0 \le n_2 \le N_2 - 1$. Let $\mathbf{x}(\mathbf{n}, i)$ be the luminance value at the sample location $\mathbf{n} = (n_1, n_2)$ of the i^{th} frame in a video sequence. The average inter-frame luminance difference D_v shown in (2) was used for the video motion evaluation [13].

$$D_v = \frac{1}{N_{I_v}} \sum_{i \in I_v} \left(\frac{1}{N_{B_i}} \sum_{k \in B_i} \left(\frac{1}{N_{P_{i,k}}} \sum_{\mathbf{n} \in P_{i,k}} \left(\begin{pmatrix} \mathbf{x}(\mathbf{n},k,i) \\ -\mathbf{x}(\mathbf{n},k,i-1) \\ +\mathbf{x}_{BG}(\mathbf{n},k,i) \\ -\mathbf{x}_{BG}(\mathbf{n},k,i-1) \end{pmatrix} / 2 \right) \right) \right) \quad (2)$$

$$\mathbf{x}_{BG}(\mathbf{n},k,i) = \frac{1}{32} \sum_{m_1=1}^{5} \sum_{m_2=1}^{5} \begin{pmatrix} \mathbf{x}_{BG}(n_1 - 3 + m_1, n_2 - 3 + m_2, k, i) \\ \cdot h_{LP}(m_1, m_2) \end{pmatrix} \quad (3)$$

where $I_v = \{I_1, I_2, \cdots, I_{N_{I_v}}\}$ indicates pictures in a video sequence, $N_{I_v} = \text{card}(I_v)$ is the total number of the pictures, $B_i = \{B_{i,1}, B_{i,2}, \cdots, B_{i,N_{B_i}}\}$ stands for the blocks within the i^{th} picture, and $N_{B_i} = \text{card}(B_i)$ is the total number of the blocks. $P_{i,k} = \{P_{i,k,1}, P_{i,k,2}, \cdots, P_{i,k,N_{P_{i,k}}}\}$ stands for the samples within the k^{th} block and $N_{P_{i,k}} = \text{card}(P_{i,k})$ is the total samples of $P_{i,k}$. The block width and height are all equal to 16 in this paper. $\mathbf{x}_{BG}(\mathbf{n},k,i)$ is the average background luminance calculated by a weighted low-pass filter mask with its center collocated at $\mathbf{n}$ as shown in Fig.3.

The average luminance standard deviation σ_v as illustrated in (4) was used for measure of the texture complexity degree. Please note that the block size in (4) is 4 ×4.

$$\sigma_v = \frac{1}{N_{I_v}} \sum_{i \in I_v} \left(\frac{1}{N_{B_i}} \sum_{k \in B_i} \left(\sqrt{ \frac{1}{N_{P_{i,k}}} \sum_{\mathbf{n} \in P_{i,k}} \left(\mathbf{x}(\mathbf{n},k,i) - \frac{1}{N_{P_{i,k}}} \sum_{\mathbf{n} \in P_{i,k}} \mathbf{x}(\mathbf{n},k,i)\right)^2 } \right) \right) \quad (4)$$

For RaceHorse (WQVGA), BasketballDrill and Vidyo3, D_v are 32.47, 12.60 and 5.32, and σ_v are 11.97, 7.10 and 5.27, respectively. It means D_v and σ_v reveal the motion and texture complexity degree effectively.

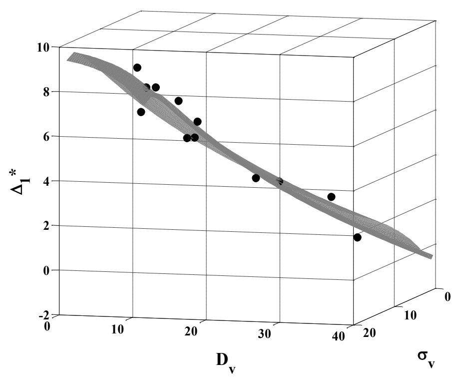

Fig. 4. Relationship between D_v, σ_v and Δ_1^*.

4. PROPOSED VC-QPC TECHNIQUE

Twelve video sequences, including Traffic, Kimono, Cactus, BasketballDrive, BQSquare, BlowingBubbles, RaceHorses (WQVGA), KristenAndSara, Vidyo1, ChinaSpeed, SlideEditing, and SlideShow, were used for establishing the prediction model of Δ_1^*. From the experimental results, D_v, σ_v and Δ_1^* can be well fitted using (5) as shown in Fig. 4. The model parameters are determined using the least-squares method. And the model parameters of p_1 to p_5 are 5.87, 1.12, -0.78, 0.03, and 0.38, respectively. The fitting accurate is 0.97 in R-square.

$$\Delta_1^* = p_1 + p_2 \ln(D_v) + p_3 (\ln(D_v))^2 + p_4 \ln(\sigma_v) + p_5 (\ln(\sigma_v))^2 \quad (5)$$

Then the QP for higher temporal layers can be achieved using (6).

$$QP_l = \begin{cases} QP_I + \text{clip3}\left(1, 10, \text{round}\left(\Delta_1^*\right)\right), & l = 1; \\ QP_{l-1} + 1, & l > 1. \end{cases} \quad (6)$$

where round(·) means a rounding to integer process and clip3(a, b, c) means limiting the value of c into a certain range [a, b].

The proposed VC-QPC is described as:

A. For the frames at the temporal base layer in a video sequence, the QP_I given in the configure file is used for encoding. Frames at other temporal layers use steps B and C for QP calculation.

B. For the second frame, calculating the video content complexity degree using (2), (3) and (4). Obtain the $QP_l, l \ge 1$ using (5) and (6).

C. For the subsequent frames in a video sequence, selecting the QP for coding according to their temporal layers and $QP_l, l \ge 1$.

The proposed VC-QPC is simple and the experiments results verify that the BD-rate improvement of the proposed VC-QPC is apparent. In other words, VC-QPC can significantly improve the R-D performance of video coding. Detailed experimental results are shown in the next section.

5. SIMULATIONS AND RESULTS

For performance evaluating purpose, QPC techniques in [4], [5], [6], and [9] (including the RDO-QPC and SRDO-QPC) were all realized. The QPC technique in the HM [4], [5] was selected as the benchmark. And the proposed VC-QPC was compared with QPC techniques in [6] and [9]. Twenty-five video sequences (including the above twelve video sequences which are used for fitting in Section 4) were encoded. All experiments were conducted using the HM14.0 under the common test conditions [10]. The coding structure was RAS with GOP size being 8. QP_I were set to 22, 27, 32, and 37, respectively.

It can be seen in Table 3, the average BD-rate of the proposed VC-QPC is -5.60% for all of the tested video sequences, and the average BD-rate is -6.42% for video sequences when the twelve fitting sequences in section 4 being removed. Compared with the methods in [6] and [9], the proposed VC-QPC has the highest R-D performance. Furthermore, the proposed technique involves no pre-encoding process, which makes it practical.

For the RA structure, temporal dependency plays an important role for the overall RD performance. And the key spirit of the proposed VC-QPC and our previous work in [9] are to utilize the temporal dependency and then derive the optimal QP offset for temporal layers. Therefore, it can be seen from Table 3 that the sequences contain slow motion, e.g. Fourpeople, have higher performance gain while fast motion sequences have lower performance gain.

Taking the video content complexity into consideration, the proposed VC-QPC is content adaptive. Without pre-encoding, it is well adapted to the scenario with scene changes. For these applications, the detection of scene changes is the key technique (while out of the scope of this paper). Once the scene change is detected, the first two frames in the new scene can be used to obtain the content complexity which is needed for calculation of QP of the subsequent frames in the same scene.

6. CONCLUSION

This paper proposed a QPC technique for the RAS, i.e., VC-QPC. Through taking the video content characteristics into consideration, a QP prediction model was established. Experimental results show that VC-QPC can significantly improve the R-D performance of video coding.

Table 3. BD-rate (%) results of different QPC techniques.

Videos	[6]	RDO-QPC	SRDO-QPC	VC-QPC
PeopleOnStreet	-1.98	-0.93	-0.48	-1.00
ParkScene	-2.14	-2.57	-2.04	-2.04
BQTerrace	-2.16	-4.17	-5.58	-5.04
BasketballDrill	-3.30	-8.53	-7.75	-8.34
BQMall	-2.13	-4.04	-4.04	-4.04
PartyScene	-3.42	-8.71	-7.65	-9.02
RaceHorses (WVGA)	-1.25	0.10	1.98	0.10
BasketballPass	-2.11	-2.05	-2.05	-1.63
FourPeople	-3.74	-11.82	-10.67	-12.58
Johnny	-3.64	-8.73	-8.73	-9.09
Vidyo3	-2.71	-10.69	-9.72	-11.09
Vidyo4	-3.73	-11.23	-10.44	-11.54
BasketballDrillText	-3.27	-8.26	-7.46	-8.15
Traffic	-2.68	-3.96	-3.96	-3.96
Kimono	-1.43	-0.89	-0.06	-1.03
Cactus	-2.81	-2.53	-2.61	-2.62
BasketballDrive	-1.33	-0.12	1.96	-0.12
BQSquare	-2.33	-7.53	-7.53	-8.49
BlowingBubbles	-2.95	-6.25	-5.74	-6.24
RaceHorses (WQVGA)	-1.66	-0.22	0.60	-0.47
KristenAndSara	-3.90	-10.43	-9.77	-10.62
Vidyo1	-2.83	-9.61	-8.87	-9.92
ChinaSpeed	-4.29	-5.82	-6.77	-7.09
SlideEditing	-0.08	0.02	0.51	0.00
SlideShow	-1.88	-4.40	-5.40	-5.96
Ave. of sequences without the twelve fitting sequences	-2.74	-6.28	-5.74	-6.42
Ave. of all sequences	-2.55	-5.33	-4.89	-5.60
Pre-encoding	None	6 times	None	None

7. ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation Research Program of China (grant 61371089), the Fundamental Research Funds for the Central Universities (grant 3102016zy019) and the PhD Start-up Fund of Xi'an University of Posts and Telecommunications (No. 101-205020012).

8. REFERENCES

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, 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, Aug. 2003.

[2] 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.

[3] ITU-T and ISO/IEC, “High efficiency video coding/information technology – high efficiency coding and media delivery in heterogeneous environments – Part 2: High efficiency video coding,” Rec. H265 and ISO/IEC 23008-2:2013, Apr./Nov. 2013.

[4] Joint Collaborative Team on Video Coding (JCT-VC), “HM Software Manual,” CVS sever at http://hevc.kw.bbc.co.uk/svn/jctvc-hm/, accessed Aug. 2016.

[5] K. McCann, B. Bross, W.-J. Han, I. K. Kim, K. Sugimoto, and G. J. Sullivan, “High Efficiency Video Coding (HEVC) Test Model 14 (HM14) Improved Encoder Description,” Joint Collaborative Team on Video Coding, JCTVC-P1002, San Jose, USA, Jan. 2014.

[6] B. Li, D. Zhang, H. Q. Li, and J. Z. Xu, “QP Determination by Lambda Value,” Joint Collaborative Team on Video Coding, JCTVC-I0426, Geneva, CH. May. 2012.

[7] T. Zhao, Z. Wang, and C. W. Chen, “Adaptive Quantization Parameter Cascading in HEVC Hierarchical Coding,” *IEEE Trans. Image Process.*, vol.25, no.7, pp. 2997–3009, Apr. 2016.

[8] M. A. Papadopoulos, F. Zhang, D Agrafiotis, and D Bull, “An adaptive QP offset determination method for HEVC,” in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 4220–4224.

[9] Y. C. Gong, S. Wan, K. F. Yang, Y. Yang, and B. Li, “Rate-distortion-optimization-based Quantization Parameter Cascading Technique for Random-access Configuration in H.265/HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, 2016, doi: 10.1109/TCSVT.2016.2539718.

[10] F. Bossen, “Common HM Test Conditions and Software Reference Configurations,” JCTVC-L1100, Feb. 2013.

[11] G. Bjøntegaard, “Calculation of Average PSNR Differences between RD Curves,” document VCEG-M33, 13th VCEG Meeting, Apr. 2001.

[12] J. –R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, “Comparison of the Coding Efficiency of Video Coding Standards-including High Efficiency Video Coding (HEVC),” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669-1684, Dec. 2012.

[13] Y. C. Gong, S. Wan, K. F. Yang, H. R. Wu, and B. Li, “A visual-masking-based estimation algorithm for temporal pumping artifact region prediction,” *Circuits Systems and Signal Processing*, vol. 36, no. 3, pp. 1264-1287, Mar. 2017.