A ROBUST HIERARCHICAL QP SETTING FOR SCREEN CONTENT CODING

Tong Tang[1], Ling Li[2], Donglin Wang[3]

Institute of Information Science and Technology, University of Science and Technology of China[1]
Institute of Software, Chinese Academy of Science[2]
Institute of Automation, Chinese Academy of Science[3]

ABSTRACT

Unlike conventional camera-captured videos (CCV), the screen content videos (SCV) are generated by computer, like text, animation, or graphics. The video contents are discontinuous, as one abrupt frame is often followed by many static frames. Therefore, the traditional hierarchical quantization parameter (QP) setting for CCV may not be suitable for SCV. In this paper, a robust hierarchical QP setting method is proposed. Concretely, a region recognition method is designed to identify the influential regions, whose quality has tremendous influence on subsequent frames. These influential regions employ a smaller QP to improve the RD (Rate Distortion) performance of encoded video. Experimental results show that the proposed method averagely reduces 1.7% bitrate, and improves RD performance for all the SCV, which outperforms other encoding methods.

Index Terms — Hierarchical QP, Screen content, High Efficiency Video Coding (HEVC).

1. INTRODUCTION

High Efficiency Video Coding (HEVC) [1] is the latest video coding standard developed by JCT-VC (Joint Collaborative Team of Video Coding), and it can save about 50% bit-rate at similar quality compared with H.264/AVC [2]. Meanwhile, with the fast development of computer technologies, screen contents become more and more important for applications like video conference, wireless display and cloud computing, where high quality SCV need to be transmitted to the client-side. SCV have some particular characteristics, such as text characters, extremely sharp edges and mixed contents, which are very different from CCV. Therefore, the screen content coding (SCC) extension has been established by JCT-VC to meet the demands of SCC [17].

Recently, there have been many new techniques for SCC. The two powerful coding techniques are intra block copy [3] and palette coding [4]. Based on these two techniques, some improved methods [5][6] have been proposed to provide better block matching. Moreover, to reduce unnecessary encoding complexity and to save bits for unnecessary motion vector precision, an adaptive motion vector resolution method was proposed [7]. And there are also some transform techniques in color space [8][9]. The above techniques all make obvious contributions in BD-rate reduction for SCC, and these techniques aim at motion estimation or residual transform. Another important issue is QP setting for video coding, which is important for motion estimation and residual transform.

SCV have distinct differences with CCV in picture characteristic distribution, which causes that traditional QP setting of CCV may be not suitable for CCV. CCV contents are uniform, the picture characteristics are continuous and stable, while it is opposite for SCV contents. Moreover, usually SCV has plenty of static frames, which are simple copies of the previous abrupt frame. Thus, the quality of that abrupt frame has great influence on subsequent frames. To illustrate the difference between CCV and SCV, *BasketballDrill* (CCV) and *Slide Editing* (SCV) are encoded with fixed QP, and the MAD (Mean Absolute Difference) values of each frame are extracted, as shown in Fig.1. We can see that CCV has relatively continuous MAD, and the MAD values fluctuate around 2. Meanwhile, the MAD values in SCV are discontinuous, and the MAD values of many frames are close to 0, which means they are static frames.

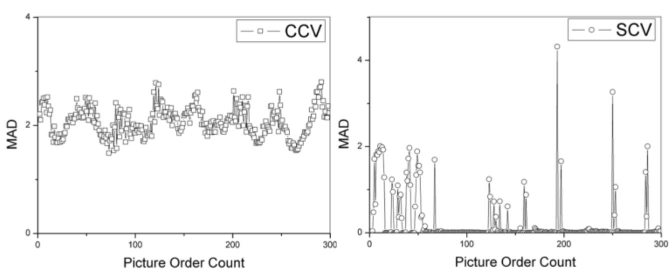


Figure 1 Frame MAD distribution comparison between CCV and SCV.

For fixed QP coding mode, the traditional QP setting is hierarchical, and different QP values are set according to the importance of reference picture in a GOP (Group of Picture), to achieve optimal RD performance as illustrated in Fig.2. The default hierarchical QP setting works well for CCV, but is not suitable for SCV. Therefore, Li et al. [10] proposed an adaptive hierarchical QP setting scheme for SCC, and smaller QP value and default QP value are allocated for the moving regions and the static regions respectively. This method can ensure that all the abrupt frames have relatively high quality, thereby improves the video quality. However, for some sequences, the moving regions have limited influence on subsequent frames, but the smaller QP settings for these regions will obviously increase bit cost, and decrease the RD performance of overall sequence.

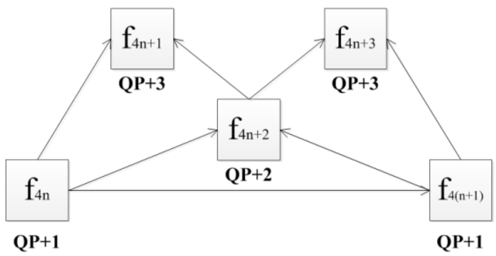


Figure 2 Hierarchical QP setting of default IBBB coding structure.

Therefore, in this paper, we propose a robust hierarchical QP setting method for SCC. Compared with method in [10], smaller QP values are only allocated for influential moving regions, which will influence plenty of subsequent frames. And the whole video quality will benefit from smaller QP setting for these moving regions. Thus the proposed method is robust for all the SCV. One of the main contributions of our proposed method is the influential region recognition. Firstly, the picture characteristic distribution of SCV is simplified, then the RD optimized problem is formulated, and finally the Lagrangian method is utilized to solve the optimization problem. Based on the optimum solution, the influential region can be recognized. Experimental results show that our proposed method outperforms other QP setting methods in terms of RD performance.

The rest of this paper is organized as follows. Section 2 introduces the proposed robust hierarchical QP setting method in detail. Section 3 presents the experimental results. And conclusion is drawn in Section 4.

2. PROPOSED APPROACH

In this section, the key frame extraction method is proposed firstly, then the recognition method of influential key frame is derived, and the moving region of influential key frame is taken as the influential regions, and finally the flowchart of proposed approach is presented.

For better mathematical modeling, the picture distribution of SCV is simplified. As shown in Fig.1, many frames of SCV are nearly static, they rarely cost bits for encoding, and their quality is determined by the previous moving frame. Thus in this paper, we make this simplification: if the frame has lot of motion information, it is thought as a key frame, whose encoding process contains complex intra prediction or inter prediction. Otherwise, the frame is thought as a static frame with skip mode. In other words, it is absolutely a copy of the previous key frame. We use fast block-matching motion estimation to measure the motion information in a frame, and a frame is classified as a key frame, if the condition in (1) is satisfied.

$$MAD^{ARPS} > \tau_1 \tag{1}$$

where MAD^{ARPS} is calculated based on the adaptive rood pattern search (ARPS) for fast block-matching motion estimation [11], and the constant threshold τ_1 is set to 1 in our experiment. The simplified example is illustrated in Fig.3. The key frames are denoted as KF, and N_i is the frame number between i-th key frame and (i+1)-th key frame.

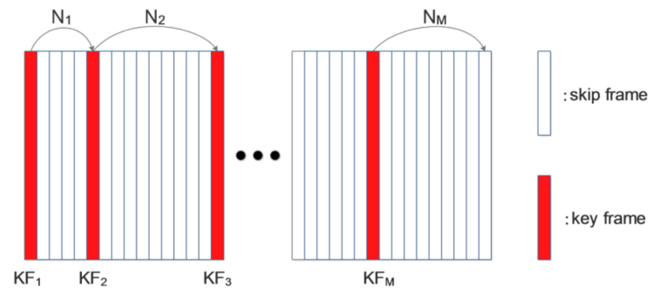


Figure 3 Simplified picture distribution of SCV.

The key frames extracted by (1) have similar RD characteristics, as shown in Fig.4. *Web Browsing* and *Conference* are both SCV. We can see that the key frames have linear RD relationship as in (2), where BPP means bits per pixels, Q is the quantization step, k is a constant model parameter.

$$BPP = k \cdot \frac{MAD}{Q} \tag{2}$$

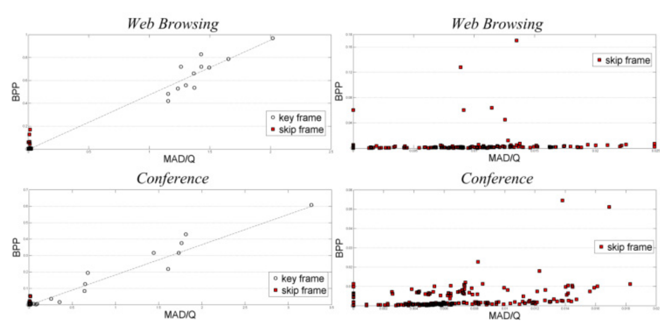


Figure 4 *R-Q* relationship investigation of key frame and skip frame for *Web Browsing* and *Conference*.

$$D = \gamma \cdot Q \tag{3}$$

Based on the R-Q model in (2) and conventional linear D-Q model in (3) [12], where γ is constant model parameter, the RD optimization problem can be formulated as in (4).

Where ΔD_{seq} is the total distortion change of the sequence as for the new-assign of QP, ΔR_{seq} is the total bitrate cost change of the sequence as for the new-assign of QP, and M is the total number of key frame in the video, D_i^0 and R_i^0 are distortion and bitrate cost respectively of the i-th key frame coded with constant quantization Q_0, D_i and R_i mean the distortion and bitrate cost respectively of i-th key frame coded with the revised hierarchical quantization Q_i.

$$\begin{cases} \min\left(\Delta D_{seq}\right) \qquad st.\ \Delta R_{seq} \le 0 \\ \Delta D_{seq} = \sum_{i=1}^{M} N_i \cdot \left(D_i - D_i^0\right) \\ \Delta R_{seq} = \sum_{i=1}^{M} \left(R_i - R_i^0\right) \end{cases} \tag{4}$$

With Lagrangian method, we can get the equation as in (5):

$$J = \Delta D_{seq} + \lambda \cdot \Delta R_{seq} \tag{5}$$

where λ is the Lagrange multiplier.

Then we have equation (6)

$$\begin{cases} \dfrac{\partial J}{\partial \lambda} = \Delta R_{seq} = 0 \\ \dfrac{\partial J}{\partial \vec{Q}} = \dfrac{\partial\left(\Delta D_{seq}\right)}{\partial \vec{Q}} + \lambda \cdot \dfrac{\partial\left(\Delta R_{seq}\right)}{\partial \vec{Q}} = 0 \end{cases} \tag{6}$$

where $\vec{Q}$ is the Q vector (Q_1, $Q_2 \ldots Q_M$).

Considering (2), (3), (4), (6), we have the derivation (7)

$$\sqrt{\lambda} = \frac{Q_0}{\sum_{i=1}^{M} MAD_i} \cdot \sum_{i=1}^{M} \sqrt{\frac{MAD_i \cdot \gamma \cdot N_i}{k}} \tag{7}$$

And we finally get the solution as in (8). Due to the space limitations, the detailed derivation is not listed here.

$$Q_i = \frac{\sqrt{MAD_i} \cdot \sum_{j=1}^{M}\left(\sqrt{MAD_j \cdot N_j}\right)}{\left(\sum_{j=1}^{M} MAD_j\right) \cdot \sqrt{N_i}} \cdot Q_0 = \Omega_i \cdot Q_0 \tag{8}$$

where Ω_i is the identifier to recognize whether the frame or region is influential. If Ω is lower than the constant threshold τ_2 (set to 1.1 in our experiment), the current frame or region is influential.

Fig.5 shows the Ω values of key frames for *Web Browsing* and *Conference*. We can see that for *Web Browsing*, all the key frames have small Ω values. This is because all the key frames are followed by many skip frames (with large *N*). That means all the key frames have great influence on subsequent frames. Thus, the method in Li [10] can perform well for sequences like *Web Browsing*. However, for *Conference*, some key frames do not have great influence on subsequent frames, as marked with black dot in Fig.5. But these frames would still be allocated smaller QP in Li [10], which degrades the RD performance. While in our scheme, this can be avoided. Therefore, the identifier Ω makes our scheme robust for various screen content videos.

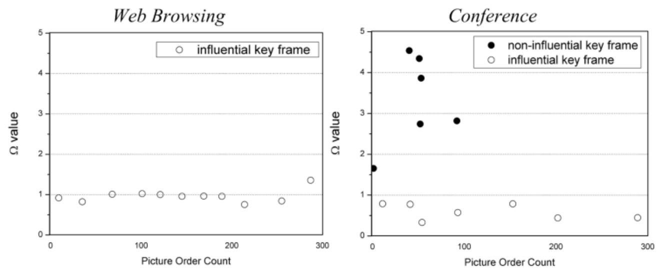


Figure 5 Ω values of key frames for *Web Browsing* and *Conference*.

In general, identifier Ω can measure influence of key frames under the premise of considering frame complexity. And the moving regions in the influential key frames are influential regions.

The flowchart of our proposed QP setting method is shown in Fig.6. $QP_{(4n)}$ is the QP assigned to the frame f_{4n} in Fig. 2. In brief, only the influential regions will be new-assigned a smaller QP, other regions will use the default hierarchical QP setting. And whether the region is influential or not is determined by the identifier Ω in (8). It should be noted that if the QP is new-assigned a smaller QP, the corresponding Lagrange Multiplier will be changed synchronously.

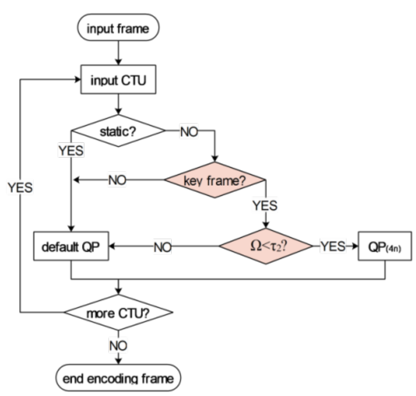


Figure 6 Flowchart of proposed QP setting method.

3. EXPERIMENTAL RESULT

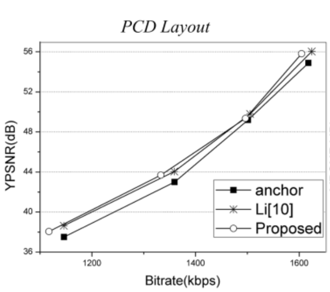
To verify the performance of the proposed approach, it is implemented into HEVC reference software for SCC HM-16.10+SCM-8.0 [13]. And the testing sequences contain the screen content videos in common test condition (CTC) [14] and other screen sequences which are once used by HEVC Range Extension [15] (including *Conference*, 720p, YUV 444; *Map*, 720p, YUV 420; *PPT*, 1080p, YUV 444; *Slide Editing*, 720p, YUV 420, *Web Browsing*, 720p, YUV 444;

PCD Layout, 1080p, YUV 444). All the sequences are encoded with IBBB structure, the configuration parameters are default except that the RDOQ and RDOQTS are off and the encoding frame rate is 30fps. The coding efficiency is measured in terms of the Bjøntegaard delta (BDBR) [16], which is used to represent the average and bit rate differences. The anchor of BDBR is the default hierarchical QP encoding scheme in HM, and the encoding QP set is (22, 27, 32, 37).

Table 1 BDBR COMPARISON

Sequence	BDBR(%)					
	Li [10]			proposed		
	Y	U	V	Y	U	V
Conference	+2.5	+3.7	+3.7	**-2.1**	**-1.7**	**-1.6**
Map	+3.5	+1.0	+3.9	**-1.2**	**-1.5**	**-0.5**
PPT	-0.8	-1.3	-1.1	**-0.8**	**-1.3**	**-1.1**
Slide Editing	-1.0	+0.1	-0.6	**-1.0**	**-1.2**	**-2.0**
PCD Layout	-1.9	-2.1	-1.6	**-2.3**	**-2.5**	**-2.3**
Web Browsing	-3.3	-1.9	-4.2	**-3.0**	**-1.8**	**-3.9**
Average	-0.2	-0.1	0.0	**-1.7**	**-1.7**	**-1.9**

Table 1 shows the BDBR results of the proposed approach and the reference scheme of Li [10]. We can see that our proposed approach can averagely achieve 1.7% BDBR saving for Y component, which obviously outperforms Li [10]. Besides, the proposed approach can achieve BDBR saving for all the testing video sequences. Due to page limits, the results of six sequences are listed. Meanwhile, Li [10] can only save BDBR for part of videos, and is not robust enough. Li [10] works well for sequences like *PPT* or *Web Browsing*. In these sequences, most key frames are followed by plenty of skip frames, which means the N_j in equation (8) is large for key frame j. The large N_j ensures the identifier Ω_i is small as long as the frame has motion, thus the method in Li [10] can save BDBR for these sequences. However, for other sequences like *Map* or *Conference*, the video content distribution is more complex, and not all the key frames are influential. Hence, not all the moving regions need a smaller QP. Therefore, the method in [10] may cause RD performance degradation.



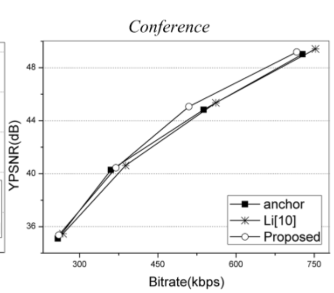


Figure 7 *RD* curves of three encoding methods for *PCD layout* and *Conference*.

Fig.7 compares the RD curves for *PCD Layout* and *Conference*. For *PCD Layout*, both Li [10] and our proposed method have obviously better RD performance than the anchor, since most of the key frames in *PCD Layout* have plenty of subsequent skip frames. And for *Conference*, our proposed method still achieves better RD performance than the anchor, while the RD performance of Li [10] is worse than the anchor. For each QP, the YPSNR of Li [10] improves compared with the anchor, but the bitrate cost increases too, which causes BDBR increase.

Except for RD performance, the encoding complexity is also important. The encoding complexity is measured with the encoding time increase (*TI*) as defined in (9).

$$TI = \frac{T_{comp} - T_{anchor}}{T_{anchor}} \times 100\% \qquad (9)$$

where T_{comp} represents the encoding time of the compared method, and T_{anchor} is the encoding time of the anchor method.

The average *TI* results are shown in Table 2. The average complexity increase of Li [10] and our proposed method are both negligible. Although the proposed method applies fast motion estimation method (ARPS is used in this paper) to obtain frame MAD, the additional time consumption is small.

Table 2 COMPLEXITY INCREASE COMPARISON

Sequence	Li [10]	Proposed
Conference	-2.1%	-2.7 %
Map	+3.4%	+1.3%
PPT	-0.1%	+1.0%
Slide Editing	+0.2%	-0.8%
PCD Layout	+0.7%	+0.8%
Web Browsing	+0.8%	+3.0%
Average	+0.5%	+0.4%

4. CONLCUSION

In this paper, a robust hierarchical QP setting method for screen content coding is proposed. By analyzing the picture characteristics of screen content video, the content characteristic distribution is simplified, the RD optimization problem is mathematically modeled, and the identifier of influential region recognition is derived, which is used to recognize whether the region is worthy of a smaller QP. Based on the identifier, a robust hierarchical QP setting scheme is presented. Experimental results show that the proposed approach is robust enough, can achieve BDBR saving for all testing screen content videos, and outperforms other encoding methods.

5. ACKNOWLEDGEMENT

This work is partially supported by the NSF of China (under Grant 61672491, 61732020).

REFERENCES

[1] 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.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, Jul. 2003.

[3] C. Pang, Y.-K. Wang, V. Seregin, K. Rapaka, M. Karczewicz, X. Xu, S.Liu, S. Lei, B. Li, and J. Xu “Non-CE2: Intra block copy and inter signaling unification”, JCTVC-T0227, 20th Meeting: Geneva, Feb2015.

[4] P. Onno, X. Xiu, Y.-W. Huang, and R. Joshi, “Suggested combined software and text for run-based palette mode”, JCTVC-R0348, 18th Meeting: Sapporo, July 2014.

[5] W. Zhu, K. Zhang, J. An, “Inter-Palette Coding in Screen Content Coding”, IEEE Transactions on Broadcasting, pp. 1-7, 2017.

[6] L. Zhao, K. Zhou, J. Guo, S.Wang, T. Lin, “A Universal String Matching Approach to Screen Content Coding”, IEEE Transactions on Multimedia, pp. 1-14, 2017.

[7] B. Li, J. Xu, G. Sullivan, Y. Zhou, and B. Lin “Adaptive motion vector resolution for screen content”, JCTVC- S0085, 19th Meeting: Strasbourg, FR, Oct 2014.

[8] L. Zhang, J. Chen, J. Sole, M. Karczewicz, X. Xiu, Y. He, and Y. Ye,“SCCE5 Test 3.2.1: In-loop color-space transform”, JCTVC-R0147, 18th Meeting: Sapporo, July 2014.

[9] L. Zhang, X. Xiu, J. Chen, M. Karczewicz, Y. He, Y. Ye, J. Xu, J. Sole, W. Kim, “Adaptive Color-Space Transform in HEVC Screen Content Coding”, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 6, pp. 446-459, 2016.

[10] J. Li, B. Li, J. Xu, and R. Xiong, “An Adaptive Hierarchical QP Setting for Screen Content Coding”, Int’l Conf. on Visual Communications and Image Processing (VCIP), Singapore, Singapore, 2015.

[11] Y. Nie, K.K. Ma, “Adaptive rood pattern recognition search for fast block matching motion estimation,” IEEE Transactions on Image Processing, Vol.11, No.12, pp.1442-1448, 2002.

[12] S. Hu, H. Wang, S. Kwong, T. Zhao, and C.-C. J. Kuo, “Rate control optimization for temporal-layer scalable video coding,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 8, pp. 1152–1162, 2011.

[13] https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.10+SCM-8.0/ .

[14] F. Bossen, “Common HM test conditions and software reference configurations”, ITU-T SG16 WP3 and ISO/IECJTC1/SC29/WG11 Document JCTVC-K1100, Shanghai, China, 10–19 Oct. 2012.

[15] ftp://hevc@ftp.tnt.uni-hannover.de/testsequences/FrExt-candidate-sequences/screen_content.

[16] G. Bjøntegaard, “Calculation of Average PSNR Differences between R-D Curves,” ITU-T SG16 Q.6 VCEG 13rd Meeting, Austin, TX, USA, Doc. VCEG-M33, Apr. 2001.

[17] W. Peng; F. Walls; R. Cohen; J. Xu; J. Ostermann; A. Macinnis; T. Lin, “Overview of Screen Content Video Coding: Technologies, Standards, and Beyond”, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 6, pp. 393-408, 2016.