SPATIOTEMPORAL PERCEPTION AWARE QUANTIZATION ALGORITHM FOR VIDEO CODING

Yunyao Yan[1,2,4], Guoqing Xiang[2,#], Yuan Li[2], Xiaodong Xie[2], Wei Yan[3,], Yungang Bao[4]*

[1] School of Electronic and Computer Engineering, Peking University, Shenzhen, China
[2] National Engineering Laboratory for Video Technology, Peking University, Beijing, China
[3] School of Software and Microelectronics, Peking University, Beijing, China
[4] Peng Cheng Laboratory, Shenzhen, China
Email: {yunyaoyan, gqxiang, yuanli, donxie}@pku.edu.cn, yanwei@ss.pku.edu.cn, baoyg@ict.ac.cn

ABSTRACT

Adaptive quantization (AQ) proves to be an effective tool to improve coding performance. In this paper, we propose an adaptively spatiotemporal perception aware quantization algorithm to increase subjective coding performance. First, the perceptual complexity models are conducted with spatial and temporal characteristics to measure the spatiotemporally perceptual redundancies, respectively. With the help of the models, the adaptively spatial and temporal quantization parameter (QP) offsets are then calculated for each coding tree unit (CTU), respectively. Finally, the perceptually optimal Lagrange multiplier of each CTU is determined with the spatial-temporal QP offset. Experimental results show that the proposed algorithm reduces 8.6% BD-Rate with SSIM (Structural Similarity Index Metric) in average over the AVS2 (the second generation of Audio Video Coding Standard) reference software RD17.0 in Low-Delay P (LDP) configurations. The subjective assessment proves the proposed algorithm can significantly reduce the bit rates with the same subjective quality.

***Index Terms*—**Adaptive quantization, Perceptual video coding (PVC), Just noticeable distortion (JND), Spatially perceptual complexity, Temporally perceptual complexity, Subjective quality

1. INTRODUCTION

AQ is an important tool for video coding to improve coding performance. It adjusts the quantization parameter (QP) for different coding units (CU) with the corresponding spatial, temporal or perceptual characteristics. According to the different characteristics, AQ schemes can be divided into two categories. One type of scheme is designed with objective characteristics. For example, in the HM16.0, the spatial AQ method, named TM5 Model [1], measures quantization step by the spatial activity of the CU related to the frame-level average spatial activity. However, compared to AQ-2 Method, which is a spatial AQ method in x265 [2], the performance is limited because the TM5 Model utilizes the minimum variance of sub CUs to denote the spatial activity of the whole CU, which means the spatial activity of the large size CU is estimated insufficient. In addition to the spatial activity, the temporal inter-frame dependency is also a significant characteristic of video coding. In [3], the temporal dependent RDO (Rate-Distortion Optimization) is proposed and the motion prediction SSD (Sum of Squared Difference) is utilized as the temporal complexity to calculate the final Lagrange multiplier for each CU. However, the QP is only adjusted according to the Lagrange multiplier of each CU empirically within small intervals. This leads to the incomplete removal of temporal redundancy.

Although the above mentioned AQ methods can improve the coding performance, they have not considered the characteristics of the human visual system (HVS). Thus, the perceptual redundancy is removed insufficiently. To improve the subjective quality of videos, the other AQ scheme is designed with the perceptual characteristics into consideration.

SSIM-based AQ is proposed in [4], which takes the SSIM as the perceptual distortion metric to guide the Lagrange multiplier and QP offset for CUs in all depths. However, this method only considers spatial variance without the effect of content on HVS. In [5], perceptually temporal characteristics are considered in the AQ method by incorporating JND (Just Noticeable Distortion) into SATD (Sum of Absolute Transformed Difference), which can reflect perceptually interframe dependency. The QP offset is generated more visual-friendly with this method. However, the spatial characteristic is inadequate consideration.

[#] The co-first author, [*] the corresponding author. This work is partially supported by the National Key Research and Development Program of China under contract No. 2016YFB0401904 and National Science Foundation of China under contract No. 61971047.

978-1-7281-1331-9/20/$31.00 © 2020 IEEE

In conclusion, the AQ methods based on objective characteristics cannot reach subjective coding performance well and the consideration of spatial or temporal perception redundancy in existed methods is inadequate. In this paper, we propose a novel CTU-level adaptive quantization algorithm to reduce not only spatially but also temporally perceptual redundancies. First, we establish the spatial perception aware adaptive quantization model to generate the spatial perception QP offset. Then, the temporal perception aware adaptive quantization model is constructed and the temporal perception QP offset is produced by perceptually temporal characteristic. Finally, the perceptually optimal QP and the corresponding Lagrange multiplier are determined adaptively for each CTU based on the obtained QP offsets. Experimental results show that compared with RD17.0 [6] of AVS2 [7], the proposed algorithm reduces by 8.6% BD-rate with SSIM for LDP configurations on average. In addition, the subjective test shows that the proposed algorithm can reduce the remarkable bit rates in the case of similar subjective quality.

The rest of the paper is organized as follows. Section 2 reviews two AQ methods with perceptual characteristics in detail. In Section 3, our proposed spatiotemporal perception aware quantization algorithm is given. Experimental results are presented and analyzed in Section 4, and finally, conclusions are drawn in Section 5.

2. RELATED WORKS

2.1. SSIM-based AQ

The AQ methods based on the perceptual characteristics can help to achieve better subjective quality. Perceptual characteristics include spatial and temporal perception characteristics. We take the SSIM-based AQ in [4] as an example of spatial perception methods to review here.

As is demonstrated in [8], SSIM is more correlated with human perceptual quality than PSNR (Peak Signal to Noise Ratio) as an image quality metric. To achieve a better subjective performance, the distortion based on SSIM (*dSSIM*) is used as the spatial perception complexity to replace the sum of mean square error (SSE) for each CU. By applying the condition that the overall rate of coding the frame is kept the same after using *dSSIM* as the distortion metric in RDO, the Lagrange multiplier can be computed as,

$$\lambda_{d,k}=\frac{2\sigma_{d,k}^2+c_2}{exp\left(\frac{1}{N_d}\sum_{j=1}^{N_d}log\left(2\sigma_{d,j}^2+c_2\right)\right)}\lambda_{SSE}, \quad (1)$$

where $\sigma_{d,k}^2$ is the variance of k-th CU at depth d and c_2 is constant. λ_{SSE} is the default Lagrange multiplier and N_d is the number of CUs at depth d. The QP offset can be determined by $\lambda_{d,k}$, since there is the direct relationship between the Lagrange multiplier and QP, which can be denoted as,

$$\lambda=\beta\cdot 2^{(QP\text{-}12)/3}, \quad (2)$$

where β is the constant related to configurations in High-Efficiency Video Coding (HEVC) [9]. Finally, the QP offset is computed as,

$$\Delta QP_{d,k}=3\cdot\left(s_{d,k}\text{-}\frac{1}{N_d}\sum_{j=1}^{N_d}s_{d,j}\right), \quad (3)$$

where $s_{d,k}=log\left(2\sigma_{d,k}^2+c_2\right)$.

However, the ΔQP_k adjusted by equation (3) can only reflect the spatial characteristics. The AQ method in [4] does not consider the perceptual temporal characteristics sufficiently. The performance may be restricted when this method is used on the sequences with complex spatial variance distribution and slow motion between frames, which is analyzed in [5].

2.2. Perceptually Temporal AQ

The temporal perception characteristic is another key factor in perceptual video coding. The perceptually temporal AQ in [5] is taken as an example here. The sum of absolute perceptual transform difference (SAPTD), as the perceptually temporal characteristic, reflects the perceptually interframe dependency, which is defined as incorporating JND into SATD. Similar to the logarithm operation in x265 [2], the original QP offset of a CU is calculated as,

$$\Delta QP_{d,k}^*=log(SAPTD_{d,k}^*), \quad (4)$$

where $SAPTD_{d,k}^*$ is the minimum SAPTD for k-th CU at depth d. The frame-level QP offset for final adjustment is denoted as,

$$\Delta\overline{QP}_d=\frac{1}{N_d}\sum_{k=0}^{N_d\text{-}1}\Delta QP_{d,k}^*, \quad (5)$$

Therefore, visual-friendly QP offset for each CU is decided as,

$$\Delta QP_{d,k}=s_T\times(\Delta QP_{d,k}^*\text{-}\Delta\overline{QP}_d), \quad (6)$$

where s_T is the empirical parameter. The results in [5] show that it can obtain better subjective performance for video coding. But there are still some problems should be concerned. For example, when used for the sequence of scene changing, which means the temporal dependency is very weak, the performance of this method may be limited. Therefore, to achieve a better subjective performance, both spatial and temporal perception characteristics need to be considered.

3. PROPOSED METHOD

3.1. Spatial Perception Aware AQ Model

It is well known that the visual resolution of the HVS is limited, which means some distortion below certain thresholds cannot be perceived by human eyes. The JND technique estimates such threshold through the visual characteristics, such as the luminance adaption, contrast masking, contrast sensitivity, etc. As the research in [10], the HVS is more sensitive to the regions with obvious statistical regularities in comparison with others and the JND threshold of these regions is lower. Therefore, we adopt the JND model by considering both the visual regularity and the contrast masking in [10] to estimate the perceptual spatial complexity. The JND model is denoted as,

$$JND(x,y)=L_A(x,y)+V_M(x,y)\text{-}0.3\cdot min\{L_A(x,y),V_M(x,y)\}, \quad (7)$$

where x and y are the pixel positions. $L_A(x,y)$ is the luminance adaptation, which is expressed as,

$$L_A(x,y)=\begin{cases}17\times\left(1\text{-}\sqrt{\frac{B(x,y)}{127}}\right) & if\ B(x,y)<127\\ \frac{3}{128}\times(B(x,y)\text{-}127)+3 & else\end{cases}, \quad (8)$$

where $B(x,y)$ is the mean luminance in the local region. Another factor $V_M(x,y)$ in the JND model is visual masking and it is calculated by considering both visual regularity and luminance contrast as follows,

$$V_M(x,y)=f\big(L_c(x,y),\mathcal{N}(x,y)\big)=\frac{1.84\cdot L_c^{2.4}}{L_c^2+26^2}\cdot\frac{0.3\cdot\mathcal{N}^{2.7}}{\mathcal{N}^2+1}, \quad (9)$$

where $L_c(x,y)$ denotes the luminance contrast and $\mathcal{N}$ is defined as the number of quantified orientation differences to estimate the visual regularity.

As described before, the larger the JND variance (σ_{JND}^2), the higher the spatial complexities of the CTUs, which means these CTUs can be quantized more. While the smaller the σ_{JND}^2 of the CTU, the more sensitive the human eyes are to it, which means the CTU should have the smaller QP. Therefore, we define the temporary spatial perception QP offset as,

$$\Delta QP_i^{S*}=log\big(\sigma_{iJND}^2\big), \quad (10)$$

where σ_{iJND}^2 is the σ_{JND}^2 of i-th CTU. Then, as analyzed in [11], the frame-level QP needs to be taken into consideration for the final quantization, which is expressed as,

$$\overline{QP^S}=\frac{1}{N}\sum_{i=1}^{N}\Delta QP_i^{S*}, \quad (11)$$

Thus, the final spatial perception QP offset is defined as,

$$\Delta QP_i^S=s_1\times(\Delta QP_i^{S*}\text{-}\overline{QP^S}), \quad (12)$$

where s_1 is an empirical coefficient to fully exploit the performance of perceptual spatial characteristics. In this paper, the s_1 is set as 2.6 which can reach the acceptable average performance within various testing sequences.

3.2. Temporal Perception Aware AQ Model

The natural video sequence contains critical temporal perception characteristics, which are related to the inter-frame dependencies. The temporal perception characteristic for each CTU can be obtained between the reference frame and the current frame by motion estimation. Usually the smaller the difference between them, the larger the interframe dependency. The motion compensation predicted (MCP) error (D^{MCP}) is the difference between the reconstructed frame and the original frame, which can be stricter to compute the interframe dependency. Thus, MCP error is taken as the temporal perception complexity in this paper. However, we cannot truly obtain the reconstructions for the future frames, since these CTU are not really coded at the point. To estimate these differences, we adopt the source distortion temporal propagation model in [3] and the MCP error of current CTU between the current frame t and the future frame t+1 can be expressed as,

$$\begin{aligned}D_{i,t+1}^{MCP}&=\|F_{i,t+1}\text{-}\widehat{F}_{i,t}\|^2=\|F_{i,t+1}\text{-}F_{i,t}+F_{i,t}\text{-}\widehat{F}_{i,t}\|^2\\&=\alpha\cdot\left(\|F_{i,t+1}\text{-}F_{i,t}\|^2\text{-}\|F_{i,t}\text{-}\widehat{F}_{i,t}\|^2\right)=\alpha\cdot\left(D_{i,t+1}^{OMCP}+D_{i,t}\right), \quad (13)\end{aligned}$$

where $F_{i,t+1}$ is the pixel value of the i-th CTU in frame t+1 and $\widehat{F}_{i,t}$ is the pixel value of the i-th CTU in current reconstructed frame t. $\|\cdot\|$ indicates the L_2-norm operator and α is a constant. $D_{i,t+1}^{OMCP}$ denotes the MCP error between the original frame t and t+1 and $D_{i,t}$ is the distortion of the i-th CTU in frame t. However, the $D_{i,t}$ cannot be truly obtained before the compression of the current frame t. As analyzed in [3], $D_{i,t}$ can be estimated as,

$$D_{i,t}=D_{i,t}^{MCP}\cdot\mathcal{F}(\theta)=D_{i,t}^{MCP}\cdot\mathcal{F}\left(\frac{\sqrt{2Q}}{D_{i,t}^{MCP}}\right), \quad (14)$$

where $D_{i,t}^{MCP}$ is the MCP error between the reconstructed frame t-1 and the original frame t, which can be obtained by motion estimation with pre-analysis. $\mathcal{F}(\theta)$ is the statistical curve, where θ is $\sqrt{2Q}/D_{i,t}^{MCP}$ and Q is the quantization step.

In summary, the MCP error of current CTU for the future frame can be expressed as follows,

$$D_{i,t+1}^{MCP}=\alpha\cdot\left(D_{i,t+1}^{OMCP}+D_{i,t}^{MCP}\cdot\mathcal{F}\left(\frac{\sqrt{2Q}}{D_{i,t}^{MCP}}\right)\right). \quad (15)$$

As above mentioned, the smaller the MCP error, the larger the interframe dependency between the two CTUs, which means the CTUs should have smaller QPs. Similar to the expression of the spatial perception QP offset, the CTU-level temporal perception QP offset ΔQP_i^{T*} is denoted as,

$$\Delta QP_i^{T*}=log\left(\sum_{j=t}^{t+m}D_{i,j}^{MCP}\right), \quad (16)$$

where $D_{i,j}^{MCP}$ is the MCP error of i-th CTU in the j-th frame and m is the reference frames chain length, which is set as 1 in this paper. The frame level QP offset is calculated as,

$$\overline{QP^T}=\frac{1}{N}\sum_{i=1}^{N}\Delta QP_i^{T*}, \quad (17)$$

After that, we have the final temporal perception aware QP offset for current CTU as,

$$\Delta QP_i^T=s_2\times(\Delta QP_i^{T*}\text{-}\overline{QP^T}), \quad (18)$$

where s_2 is an empirical coefficient and its effect is similar to that of s_1 and it is set as 1.5 in this paper.

3.3. Spatiotemporal Perception Aware AQ Model

Human eyes usually are affected by both spatial and temporal perception redundancies. Therefore, we construct an adaptive spatiotemporal perception aware model to eliminate both of the spatial and temporal perception redundancy. Considering that there can be an overlap between those two kinds of redundancies, the spatio-temporally optimal QP offset is expressed as,

$$\Delta QP_i^{ST}=\begin{cases}\Delta QP_i^S+\Delta QP_i^T-s\cdot min\{\Delta QP_i^S,\Delta QP_i^T\} & if\ Sgn(i)==1\\ \Delta QP_i^S+\Delta QP_i^T & else\end{cases}, \quad (19)$$

where s is an empirical coefficient, which is set as 0.6 in this paper. $Sgn(\mathrm{i}) == 1$ means that $sign(\Delta QP_i^S)$ is equal to $sign(\Delta QP_i^T)$ of i-th CTU, where $sign(\gamma)$ is -1 when negative γ and 1 for positive γ. It can be seen that when $sign(\Delta QP_i^S)$ is equal to $sign(\Delta QP_i^T)$, the spatial and temporal perception redundancies have some overlap effects on the HVS. In this case, we consider reducing the impact of those overlap effects. In another case, the spatial and temporal perception redundancies have the opposite effects for human eyes, which means there is no overlap effect. Therefore, the final ΔQP_i^{ST} is obtained by directly adding ΔQP_i^S and ΔQP_i^T. As a consequence, the perceptually optimal QP for the current CTU is denoted as,

$$QP_i=QP_{frame}+\Delta QP_i^{ST}, \quad (20)$$

where QP_{frame} is the frame-level QP. Finally, the Lagrange multiplier is adjusted with the QP_i according in RD17.0 as,

$$\lambda_i=w\cdot 2^{(QP_i-11)/4}, \quad (21)$$

where w is the constant related to the frame type and interframe structure. In brief, the flow of the proposed adaptively spatiotemporal perception aware quantization algorithm is shown as Fig. 1.

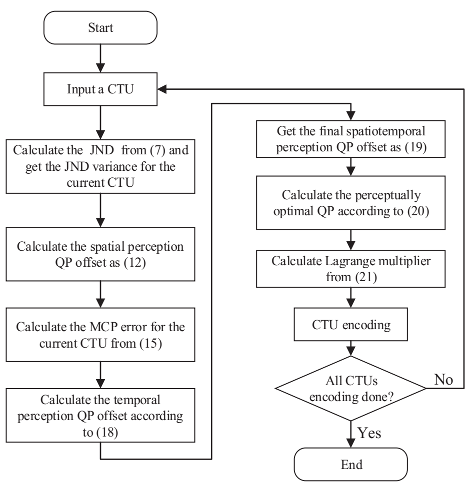


Fig. 1. Algorithm flow of the proposed adaptively spatiotemporal perception quantization.

4. EXPERIMENTAL RESULTS

4.1. Experiment Setup

Experiments are performed on the platform of the AVS2 reference software RD17.0 to evaluate the proposed algorithm. The simulation environment is set as suggested by the common test conditions (CTC) [12] specified by AVS-P2 under LDP configurations. Simulations were conducted on 17 sequences with resolution ranging from 1920×1080 to 416×240. The QP values are set to 20, 26, 32 and 38, respectively. Two other related approaches are also tested for the coding performance comparisons, which are the SSIM-based AQ in [4] and the perceptually temporal AQ in [5], respectively.

4.2. Coding Performance Comparisons and Discussions

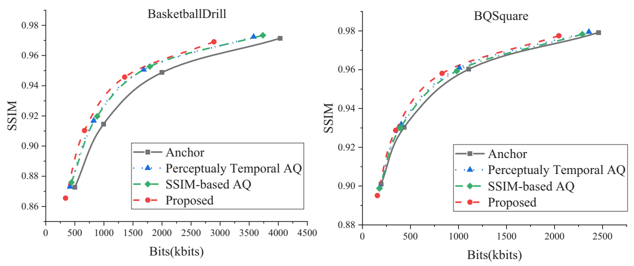
The rate-distortion coding performance of the proposed spatiotemporal AQ method is measured in terms of BD-rate based SSIM over RD17.0 as shown in Table 1. The AQ methods are performed on the CTU with the size of 64×64 for all test methods. Considering the computational complexity of temporal perception aware AQ model, we utilize the unrestricted center-Biased diamond search method in [13] with the search range 64×64 to estimate the perceptually temporal complexity. The QP offset is allowed at the CTU level within the range [-9, +9]. It can be seen that the proposed method achieves -8.6% BD-rate with LDP on average over RD17.0. Fig. 2 shows examples of coding performance comparison for two sequences. Comparing the results of the SSIM-based AQ in [4] and the perceptually temporal AQ in [5], the proposed algorithm can achieve better performance since the proposed scheme considers both spatial and temporal perception redundancies.

For an intuitive illustration of the proposed method's behavior, Fig. 3 shows the varying QP offsets over different CTU locations in the 3rd frame of the sequence "Sunflower". The warmer the CTU corresponds to a larger QP offset while the cooler the CTU associates with a smaller QP offset. It can be seen that for the spatial perception aware AQ (SPAQ), the smaller QP offsets are applied to the CTUs in regular regions which are much more sensitive to HVS and the larger QP offsets are used for the CTUs with complex texture. In Fig. 3(b), the perceptually spatial complexities of CTUs in tubular flower regions are larger than those in other regions, which means these CTUs can be quantized more.

For the temporal perception aware AQ (TPAQ), the smaller QP offsets are applied to the CTUs in the regions with smaller perceptually temporal complexities to protect the interframe dependencies and the larger ones are used for the objects with less interframe dependencies. In Fig. 3(a), the foreground with the bee has motion while the background keeps almost static. Thus, the interframe dependencies of the CTUs with the bee are less than the CTUs with sunflower, which can be quantized more, while the CTUs with the sunflower should be protected more as Fig. 3(c) shown.

Table.1. Coding performance comparison in terms of BD-rate using SSIM

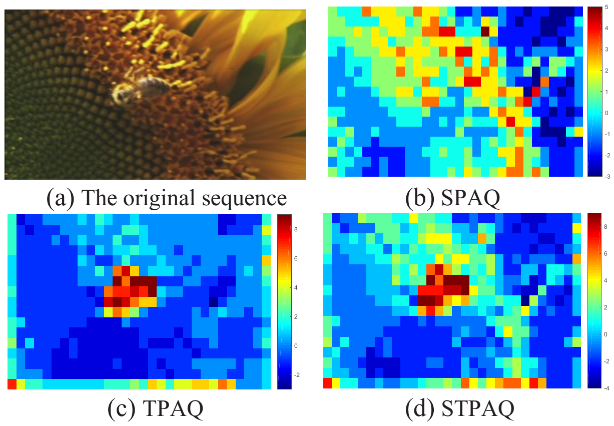
Resolution	Sequence	Pro.	[5]	[4]
1080p	Sunflower	-8.0%	-8.1%	-6.1%
	Cactus	-5.2%	-4.3%	-4.1%
	BasketballDrive	-9.0%	-6.8%	-7.5%
	Pedestrian_area	-4.7%	-2.8%	-2.1%
WVGA	BQMall	-5.0%	-4.8%	0.4%
	PartyScene	-8.5%	-6.4%	-2.7%
	BasketballDrill	-26.6%	-19.3%	-18.3%
	RaceHorses	0.0%	-0.8%	0.5%
WQVGA	BasketballPass	-21.6%	-15.8%	-12.7%
	BlowingBubbles	-1.9%	-4.6%	1.1%
	BQSquare	-15.9%	-9.7%	-7.3%
	RaceHorses	-5.1%	-3.5%	-4.3%
720p	Harbour	-0.6%	-0.6%	1.5%
	Crew	-4.2%	-2.0%	-5.0%
	City	-7.8%	-4.8%	-6.2%
	Vidyo1	-10.6%	-7.8%	-7.1%
	Vidyo4	-11.6%	-6.6%	-5.9%
Average		-8.6%	-6.4%	-5.1%



(a) BasketballDrill (b)BQSquare

Fig. 2. BD-rate based on SSIM comparison for different sequences.

For the spatiotemporal perception aware AQ (STPAQ), the QP offset includes the consideration of both spatial and temporal perception characteristics. In other words, the smaller QP offsets are applied for the CTUs with both more interframe dependencies in temporal domains and regular in spatial domains. While the larger QP offsets are utilized for the CTUs with high complexity in both perceptually spatial and temporal domains. A smaller QP offset tends to lead the encoder to select the coding option with smaller distortion and higher consumed bits, which in turn provides better perceptual coding quality while the larger QP offsets decrease the consumed bits in the CTUs which are not perceptually important. As Fig. 3(d) shown, the CTUs in the regions with the bee and the tubular flower can be assigned the larger QP offsets to save bits without too much perceptual distortion while the CTUs in the areas with sunflower seeds and ligulate flowers are distributed the smaller QP offsets due to the low perceptually spatial and temporal complexities. The method of SPAQ, TPAQ and STPAQ can achieve -1.8%, -6.7% and -8.0% BD-Rate with SSIM, respectively. It can be seen that the STPAQ can achieve better coding performance than those methods which only consider the spatial or the temporal perceptual characteristics alone.



(a) The original sequence (b) SPAQ (c) TPAQ (d) STPAQ

Fig. 3. A visualized representation of CTU-based QP offsets for sequence "Sunflower".

4.3. Subjective Quality Comparison

Finally, the subject quality assessment experiments are performed with the RD17.0 and the proposed method for all testing sequences. The double stimulus continuous quality scale (DSCQS) [14] method is used as our subjective quality evaluation. 12 observers (8 of them are relative field and others native) are employed and all of them have been trained before the test. The displaying order of videos is random and unknown to the observes. And then observers are asked to give the mean opinion score (MOS) from the consecutive numbers ranging from 1 to 5. Fig. 4 shows the average MOS comparisons of all test sequences for frame QP = 20, 26, 32 and 38 with LDP configurations. It can be seen that the MOS values of the proposed algorithm are close to those of the anchor RD17.0, which means the subject qualities for the proposed algorithm and the anchor RD17.0 are similar, while the proposed algorithm can reduce generous bitrates. To show the subjective quality directly, the visual comparison of the sequence "BasketballDrive" is shown as an example in Fig. 5. It can be seen that our method can save 34% bitrates with almost the same subjective quality.

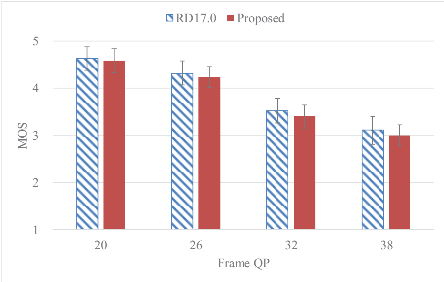


Fig. 4. MOS comparisons between the RD17.0 and the proposed algorithm.

(a) (b)

Fig. 5. Subjective comparisons for sequence "BasketballDrive" (a) RD17.0: 29640bits, SSIM: 0.8992, MOS:2.75, (b) Proposed: 19312bits, SSIM:0.8928, MOS:2.67.

5. CONCLUSIONS

In this paper, we propose an adaptive perception aware quantization model to achieve better subjective coding performance for video coding. The spatial and temporal perception aware model is firstly established respectively to measure the perceptual complexity. With the help of the models, the adaptively spatial and temporal QP offsets are then calculated for each CTU. Finally, the perceptually optimal Lagrange multiplier of each CTU is determined with the spatial-temporal QP offset. Experiment results have demonstrated that the proposed algorithm can improve the performance of the RD17.0 codec by -8.6% in terms of BD-rate with SSIM for the LDP configurations. Subjective quality evaluation proves that the proposed algorithm can save generous bitrates with a similar visual quality comparing to RD17.0. It should be noted that the proposed algorithm is not limited to the use on RD17.0, but also can be applied on other platforms. In the future, we will study better AQ methods to improve more subjective coding performance.

6. REFERENCES

[1] ISO/IEC/JTC1/SC29/WG11, MPEG-2 Test Model 5, Rate Control and Quantization Control, Mar. 1993 (Chapter 10)

[2] http://hg.videolan.org/x265.

[3] T. Yang, C. Zhu, X. Fan and Q. Peng, "Source Distortion Temporal Propagation Model for Motion Compensated Video Coding Optimization," 2012 IEEE International Conference on Multimedia and Expo, Melbourne, VIC, 2012, pp. 85-90.

[4] C. Yeo, H. L. Tan and Y. H. Tan, "SSIM-based adaptive quantization in HEVC," 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, 2013, pp. 1690-1694.

[5] G. Xiang, H. Jia, M. Yang, X. Zhang, X. Huang, J. Liu and X. Xie, "A perceptually temporal adaptive quantization algorithm for HEVC." Journal of Visual Communication and Image Representation vol.50, pp. 280-289, Jan. 2018.

[6] AVS2.0 Reference Software [Online]. Available at: http://www.avs.org.cn/AVS2/en/download.asp

[7] Z. He, L. Yu, X. Zheng, S. Ma and Y. He, "Framework of AVS2-video coding," 2013 IEEE International Conference on Image Processing, Melbourne, VIC, 2013, pp. 1515-1519.

[8] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," in IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[9] G. J. Sullivan, J. Ohm, W. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.

[10] J. Wu, G. Shi, W. Lin and C. C. J. Kuo, "Enhanced just noticeable difference model with visual regularity consideration," 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, 2016, pp. 1581-1585.

[11] G. Xiang, H. Jia, M. Yang, Y. Li and X. Xie, "A novel adaptive quantization method for video coding. " Multimedia Tools and Applications, vol. 77, no. 12, pp. 14817-14840, Jun. 2018.

[12] X. Zheng, "Common test conditions of AVS-P2", document AVS-N2020, Audio Video Coding Standard Workgroup of China, Shenzhen, China ,2014.

[13] J. Y. Tham, S. Ranganath, M. Ranganath, A. A. Kassim, "A Novel Unrestricted Center-Biased Diamond Search Algorithm for Block Motion Estimation", IEEE Transactions on Circuits & Systems for Video Technology, vol. 8, no. 4, pp. 369-377, Aug. 1998.

[14] Recommendation I. 500–11. Methodology for the Subjective Assessment of the Quality of Television Pictures. Recommendation ITU-R BT. 500–11. ITU Telecom. Standardization Sector of ITU, 2002