



# Adaptive QP offset selection algorithm for virtual reality 360-degree video based on CTU complexity

Zhi Liu<sup>1</sup> · Ke Yang<sup>1</sup> · Xianya Fu<sup>2</sup> · Mengmeng Zhang<sup>1</sup> · Zhao Wang<sup>3</sup> · Fuqi Mao<sup>1</sup>

Received: 20 April 2020 / Revised: 10 August 2020 / Accepted: 16 September 2020 /

Published online: 25 September 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

Virtual reality 360-degree video requires ultra-high resolution to provide realistic feeling and dynamic perspective. Huge data volume brings new challenges to coding and transmission. Quantization parameter (QP) is one of the key parameters to control output bitrate and reconstruction quality during coding process. Many QP offset selection algorithms designed for this kind of video are based on latitude or Equirectangular Projection (ERP) weight maps, which cannot adapt to the situation of the flat block in tropical area or the complex block in polar area. In this paper, a new metric to measure complexity of Coding Tree Unit (CTU) is designed, and an adaptive QP offset selection algorithm is proposed based on CTU complexity to improve the quantization process. Each CTU is classified into one of the five categories according to its complexity, and then different QP offset value is determined for each category. By improving the quality of the visually sensitive area and reducing the bitrate of the flat one, the efficiency of the encoder is improved. The experimental results show that, compared with the HM16.20, the WS-PSNR increases by 0.40 dB, the BD-rate reduces by 1.99%, and the quality of visually sensitive areas has improved significantly.

**Keywords** Virtual reality · 360-degree video · QP offset selection · CTU complexity

---

✉ Zhi Liu  
lzliu@ncut.edu.cn

✉ Mengmeng Zhang  
obtx7rw@163.com

<sup>1</sup> School of Information Science and Technology, North China University of Technology, Beijing 100041, China

<sup>2</sup> Yi Xinda (Beijing) Technology Co., Ltd, Beijing 100080, China

<sup>3</sup> Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China

## 1 Introduction

Virtual reality 360-degree video (hereinafter referred to as 360-degree video) requires ultra-high resolution to provide realistic feeling and dynamic perspective [11]. The huge amount of data presents challenges to coding and transmission. In the Equirectangular Projection (ERP) format, the 360-degree video image is stretched to varying degrees, generating a lot of redundant data, which significantly affects encoding performance [2, 9, 21]. Therefore, researches on coding optimization of 360-degree video have received widespread attention. In order to improve coding performance, an adaptive quantization parameter (QP) offset selection algorithm based on Coding Tree Unit (CTU) complexity is proposed to optimize the quantization process in High Efficiency Video Coding (HEVC). The main research contents and innovations are as follows:

The characteristics of video images in the ERP projection format are researched, a new metric for measuring CTU complexity is designed, and an adaptive QP offset selection algorithm based on CTU complexity is proposed. In this algorithm, CTUs are classified into five categories according to complexity, and different QP offset values are determined for each category. In this way, the flat blocks in the equatorial region and the complex blocks in the polar region can be quantified effectively. The coding performance is enhanced by improving the quality of the complex region and reducing the bit rate of the flat region. Experimental results demonstrate that, compared with the latest reference software HM 16.20, the proposed algorithm can achieve 0.40 dB WS-PSNR improving and 1.99% BD-rate saving. The proposed algorithm can significantly improve 360-degree video coding performance.

## 2 Related work on QP offset selection

In HEVC, QP is one of the key parameters that control the output bit rate and reconstruction quality in the encoding process [4, 14]. Different from traditional video, the complexity of each area of 360-degree video varies greatly, and the large-scale QP distribution cannot achieve the desired effect. Reducing the bit rate or improving the reconstruction quality of some areas requires the determination of QP according to the characteristics of 360-degree video [23]. In [26], a new entropy equalization optimization (EEO) method is proposed to enhance the encoding performance of virtual reality 360 videos. This method proposes a spherical bit rate equalization strategy to obtain the Lagrange multiplier of the coding block, which is used in the rate-distortion optimization process in video coding. Then the QP value of each block is determined according to dynamic adaptation. Based on the EEO method, two algorithms, EEOA-ERP and EEOA-CMP, are proposed to improve the compression efficiency of ERP and CMP images respectively. Experimental results show that compared with the latest algorithm WSU-ERP, EEOA-ERP achieves a 0.37% reduction in BD-Rate in the low-latency (LD) configuration. Compared with HM16.17 under normal test conditions, the objective quality of EEOA-CMP in random access (RA) configuration improves by 2.6%.

At present, many QP offset selection algorithms perform QP selection in strips. In the F meeting of the Joint Video Experts Team (JVET), an adaptive QP scheme with number F0038 [15] based on ERP characteristics was proposed. The idea of the proposal is to adjust the QP according to the weight of WS-PSNR which has a good correspondence with the distortion distribution of ERP projection video. After compensation, the QP in the polar region can be increased while the QP in the equatorial region remains unchanged.

The F0038 scheme provides ideas for many QP offset selection algorithms. In [10], a coding optimization method based on spherical average weight is proposed to improve the

coding efficiency of 360-degree video. Reduce coding errors by assigning finer quantization parameters to areas with larger weights. The CTU is adaptively quantized according to the WS-PSNR weight. Since the decoder can derive the QP offset based on the projection type and the spatial position of the CTU, there is no need to transmit other parameter information. Experimental results show that compared with fixed QP coding, the proposed adaptive quantization algorithm can greatly improve the coding performance in ERP, rotating ERP and CMP formats.

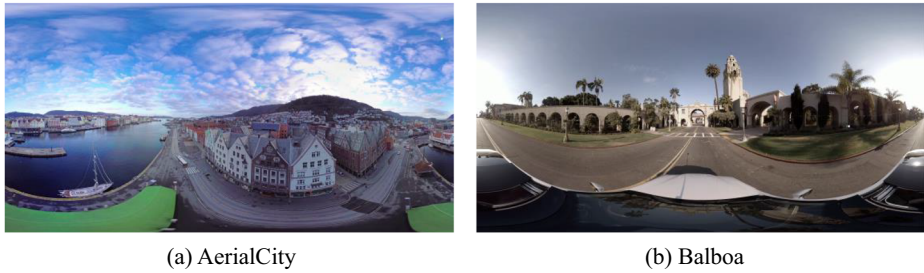
In [11], a rate-distortion optimization scheme based on spherical domain is proposed. QP is adaptively selected according to the WS-PSNR weight. The weight adopts the center point position of each CTU and defines new WSAD and WSATD. Based on HEVC, the optimal rate-distortion relationship in the spherical domain is derived, which saves the code rate.

In [24], a new adaptive algorithm is presented based on the weight of WS-PSNR and using the normalization method. This algorithm defines a new weight of WS-PSNR to adjust QP. Compared with the two methods, the WS-PSNR weight is analyzed in detail, and the idea that QP values will decrease in tropical area and increase in polar area is proposed.

An adaptive QP offset selection algorithm based on CTU complexity is proposed in this paper, which is quite different from the above-mentioned existing QP selection algorithm. The literatures [13, 15, 18, 24] use the correspondence between the weight of WS-PSNR and the distortion distribution of ERP projection video to select the QP offset. The closer to the polar area, the greater the QP value is, and the overall bit rate can be saved by reducing the bit rate in the polar area. The literatures [13, 18, 24] carried out in-depth exploration on the correspondence between WS-PSNR and QP offset to obtain a more efficient QP offset selection scheme. The premise of these algorithms is that in the ERP projection format, the closer to the polar area, the more serious the stretching is, and the video content is relatively fixed. There is a large flat area near the polar area, and the important areas are concentrated on the tropical area. However, as the types of 360-degree videos continue to increase, scenes become more and more complex [7]. These algorithms cannot adapt well to the situation where there are the flat block in tropical area or the complex block in polar area. The video images are classified at the CTU complexity in this paper, different QP offset values are determined for each category. By improving the quality of the visually sensitive area and reducing the bitrate of the flat one, the efficiency of the encoder is improved.

### 3 Adaptive QP offset selection algorithm based on CTU complexity

360-degree video shooting is roughly divided into two categories: fixed-position cameras and movable cameras [20]. As shown in Fig. 1, the video sequences are mostly shot outdoors with the sky at the top of the video and the ground at the bottom. In the AerialCity sequence (fixed-position cameras), the viewer focuses on the houses in tropical area and the vehicles moving on the road; in the Balboa sequence (movable cameras), the landscape of the buildings on both sides is more attractive as the vehicles move forward [12]. The viewer is more sensitive to moving and complex objects, rather than the background and the stationary object. In addition, many pixels are similar in the area of sky or ground, and the fluctuation amplitude in the CTU is very small. It is unnecessary to quantify in detail. Therefore, if more code rate can be allocated to the human visually sensitive area during the encoding process and non-visually sensitive areas or flat areas are



**Fig. 1** 360-degree video image (a) AerialCity (b) Balboa

roughly coded, the coding performance can be improved under the condition that the increase of bitrate is within the acceptable range, so as to obtain the image with better subjective quality.

For the huge data volume of 360-degree video and the limitation of the current QP offset selection problem, a new metric to measure the complexity of CTU is designed, and an adaptive QP offset selection algorithm based on CTU complexity is proposed to improve the quantization process in this paper. The algorithm consists of three parts including calculating the texture complexity of each CTU by using gradient, classifying each CTU into one of the five categories according to its complexity, and determining different QP offset value for each category. By improving the quality of the visually sensitive area and reducing the bitrate of the flat one, the encoder can achieve higher efficiency.

### 3.1 Complexity calculation based on CTU gradient

There are various connections among the pixels in the video frame, so the texture complexity can be calculated using the difference between the current pixel and the adjacent pixels (upper left pixel, upper pixel, left pixel). Among the methods of image processing, the gradient reflects the connection characteristics between pixels [10]. The video image is stored as a two-dimensional array, and the image gradient is the derivative of the pixel  $p(i, j)$ :

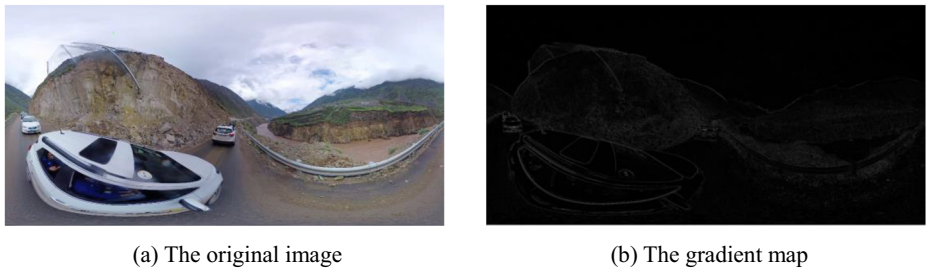
$$\text{Grad}(x, y) = |\text{dx}(i, j)| + |\text{dy}(i, j)| \quad (1)$$

$$\text{dx}(i, j) = p(i, j) - p(i, j-1) \quad (2)$$

$$\text{dy}(i, j) = p(i, j) - p(i-1, j) \quad (3)$$

Image gradient is commonly used for edge detection [8]. The direction of the gradient points to the direction in which the image changes fastest. At the edge, the pixel value jumps and the calculated gradient value is large, so the gradient can obviously reflect the existence of edge in the image. The effect of image gradient processing of a frame in the 360-degree video sequence DrivingInCountry is shown in Fig. 2.

In video coding, the texture complexity of a block is often considered. The gradient can reflect the direction and speed of the pixel change, which can reflect the pixel fluctuation range within the block and highlight the pixel jump. Therefore, the texture complexity can be expressed by



**Fig. 2** DrivingInCountry (a) The original image (b) The gradient map

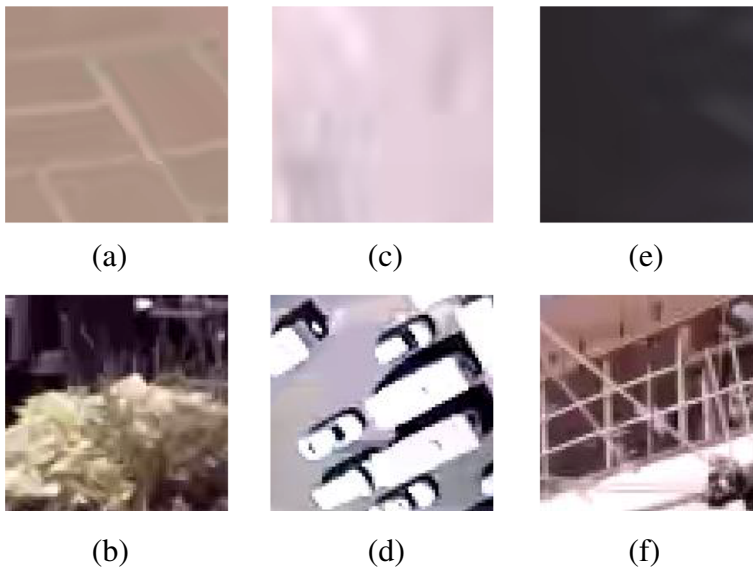
calculating the gradient of the CTU. This paper proposes the following two equations which can be used to calculate the CTU complexity of polar area  $T_p$  and tropical area  $T_e$ :

$$T_p = \sum_{j=2}^H \sum_{i=2}^W \left( \eta |p_{i,j} - p_{i,j-1}| + (1-\eta) |p_{i,j} - p_{i-1,j}| \right) \tag{4}$$

$$T_e = \sum_{j=2}^H \sum_{i=2}^W \left( |p_{i,j} - p_{i,j-1}| + |p_{i,j} - p_{i-1,j}| + |p_{i,j} - p_{i-1,j-1}| \right) \tag{5}$$

where  $W$  represents the width of the CTU,  $H$  represents the height of the CTU,  $p_{i,j}$  represents the current pixel,  $p_{i,j-1}$  represents the upper pixel,  $p_{i-1,j}$  represents the left pixel,  $p_{i-1,j-1}$  is the upper left pixel,  $\eta = h/\hat{H}$ ,  $h$  is the ordinate value of the upper left corner of each CTU, and  $\hat{H}$  is the height of the image. The value of  $h$  is defined as 0 at both ends of the image and 100 in the middle. Since the ERP projection format has the feature that the latitude increases as the image is stretched and deformed, it is necessary to make a distinction between tropical area and polar area. Equations (4) and (5) are used to calculate the CTU complexity in polar area and tropical area, respectively. In Eq. (4), the closer CTU is to polar area, the smaller  $\eta$  value is and the smaller its reference value to the upper pixel is; the larger  $1 - \eta$  value is, the more important the reference value is to the left pixel. Equation (5) is used for tropical area. In tropical area, the horizontal stretch becomes smaller, and the correlation between the current pixel and the left pixel is reduced. Increasing the upper left reference pixel helps to make the calculation result reflect the fluctuations in the block better.

The CTUs obtained from the three video sequences (GasLamp, Landing2, Harbor) is illustrated in Fig. 3. Their corresponding texture complexity is (a) 9965, (b) 112,277, (c) 916, (d) 155,668, (e) 351, (f) 119,680. From the subjective evaluation and data, Eqs. (4) and (5) can accurately reflect the texture complexity of each CTU. Flat CTU has a low texture complexity value, such as (c), (e). These areas are single in color and have no obvious ups and downs, which are mostly used as a fixed background in video images. From the pixel point of view, these areas have similar pixel values, and most of them are the same. Relatively the CTU with complex texture and with edges has higher  $T_p$  and  $T_e$  values, such as (b), (d), (f). These areas are mostly detailed areas in the video image, and the content is richer. The adjacent pixel values are quite different, and some pixels in the entire complex area are greatly different. In addition, areas with simple textures, like (a), are also well represented in terms of texture complexity values and are not treated as flat blocks. This is because there is a big difference between the texture part of the pixel and the adjacent pixel. This shows that  $T_p$  and  $T_e$  can



**Fig. 3**  $64 \times 64$  CTU example

reflect the texture of the block completely. In turn, it provides a guarantee for the accuracy of CTU categories classification.

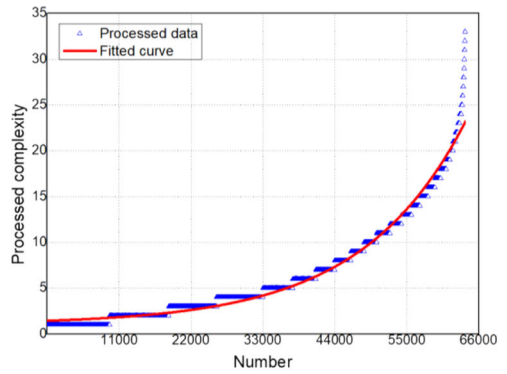
### 3.2 CTU categories classification based on CTU complexity

The larger part of the 360-degree video picture is the areas with simple texture, such as the sky and the ground. Nevertheless, the human eye pays more attention to the moving things and the more complicated texture areas [5]. Therefore, it is not enough to divide it into flat and complex parts. The entire image needs to be further classified into multiple parts. Usually, the number of gray levels that a person can distinguish with the naked eye is about 16 levels. Therefore, the CTU is first classified into 16 categories for experimental testing. On the premise that the performance of the algorithm is basically unchanged, the classification intervals are merged.

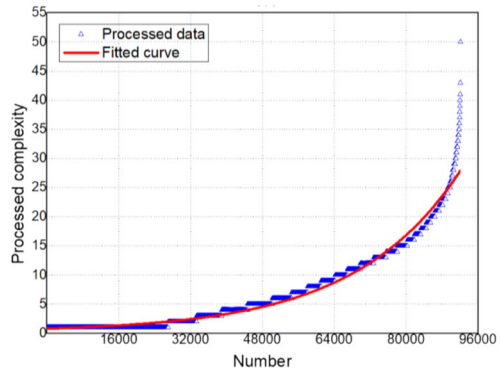
Equations (4), (5) are used to acquire the texture complexity of each CTU texture for all three frames of the test sequence. The data is screened, a small amount of non-universal data is excluded to make the overall distribution of data more regular and coherent. The total amount of data is 334,770 in this paper. The polynomial function is used to fit the processed data and get the fitting curve [6]. The data distribution statistical chart is shown in Fig. 4. After the fitting, a small number of extreme values were discarded. Data curve becomes linear and continuous, with a more intuitive trend. As shown in Fig. 4, a large number of CTUs have low complexity. As the complexity of the texture increases, the number of CTUs is gradually decreasing.

In order to classify CTU scientifically, the abscissa of the fitted curve is chosen at equal intervals, and the corresponding ordinate value is used as the threshold for the initial partitioning interval. After selecting the initial threshold, all video sequences are tested repeatedly, and adjusted the threshold interval according to the effect of each experiment. When setting equal intervals, many low complexity CTUs are classified into medium complexity intervals. The number of CTUs in the high complexity interval is extremely small. This

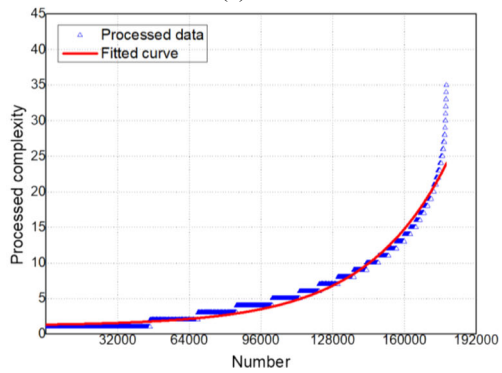
**Fig. 4** CTU complexity data distribution chart



(a) 4K



(b) 6K



(c) 8K

situation leads to unreasonable quantification. In the actual experiment process, CTU is gradually classified into appropriate intervals, and the partitioning intervals are gradually combined to reduce the complexity under the premise of not affecting the algorithm effect to a large extent. Finally, all CTUs are classified into five categories in the intra image to reflect the importance of CTU. The threshold of the best partition interval is obtained according to the texture. CTU categories classification interval and the number of CTUs in the corresponding interval is shown in Table 1. As shown in the table, the number of CTUs in



**Table 1** CTU categories classification interval and number

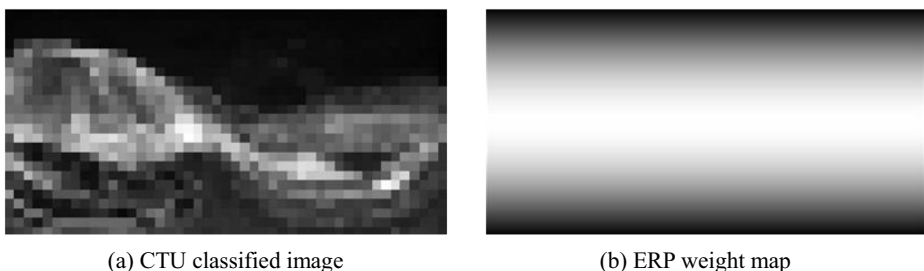
Class	CTU category	Classification interval	Number of CTUs
4 K	5	[0, 33,750)	41,170
	4	[33,750, 45,000)	10,150
	3	[45,000, 67,500)	10,130
	2	[67,500, 112,500)	10,100
	1	[112,500, 150,000)	10,290
6 K	5	[0, 45,000)	42,550
	4	[45,000, 67,500)	10,040
	3	[67,500, 112,500)	10,320
	2	[112,500, 157,500)	10,020
	1	[157,500, 200,000)	9850
8 K	5	[0, 45,000)	110,390
	4	[45,000, 90,000)	33,830
	3	[90,000, 112,500)	11,550
	2	[112,500, 202,500)	11,420
	1	[202,500, 250,000)	11,960

the final debugged range is roughly the same as the curve in the graph. This reflects the rationality of interval classification.

As seen in Fig. 5, CTU classified image and ERP weight map are compared. In order to better reflect the CTU categories grading situation, here is the rendering of the initial 16 categories. In the CTU classified image, 16 Gy scales represent the complexity of CTU. High brightness indicates high complexity. It can be clearly seen that ERP weight maps only focus on latitude changes, which is shown in the following aspects: video content is highly concerned for tropical area, and the distortion should be minimized through QP negative compensation to reduce the quantization step size; the video content is less concerned for polar area, and the QP is positively compensated to increase the quantization step size. Therefore, ERP weight maps cannot adapt to the situation of flat block in tropical area or complex block in polar area. However, the CTU classified image can reflect the CTU categories of different complexity. At the same latitude, CTUs of different complexity can be distinguished, solving the problem that many algorithms cannot adapt to the situation of flat block in tropical area or complex block in polar area.

### 3.3 QP offset selection based on CTU category

Since the adjustment of QP directly affects the quality of the reconstructed video, in order to improve the image quality of the visually sensitive area, more code streams should be allocated



**Fig. 5** Image area classification comparison (a) CTU classified image (b) ERP weight map



[25]. The visually sensitive areas have different levels of complexity and should distinguish the QP offset value. Through subjective evaluation, comparing a large number of CTUs and their corresponding complexity calculation results, it can be observed that there is no obvious texture in the CTU of category 5, basically pure color blocks, or small fluctuations in the block, while textures appear in the CTU of category 4, and the pixels produce large fluctuations. Therefore, the category 4 and above are set as the visually sensitive area. The proposed algorithm allocates the continuous QP offset value for different categories, which enables the CTU in the visually sensitive area to carry out fine quantification. The continuous QP offset value also avoids the block effect between different categories. The initial QP offset values are set for the experiment. Corresponding to  $QP_{base}= 22, 27, 32, 37$ , the category 4 is set to  $QP_{offset}=- 1, -2, -3, -4$ , the category 3 is set to  $QP_{offset}=- 2, -3, -4, -5$ , the category 2 is set to  $QP_{offset}=- 3, -4, -5, -6$ , the category 1 is set to  $QP_{offset}=- 4, -5, -6, -7$ . The QP offset values are adjusted by testing all video sequences multiple times. The final QP offset selection for visually sensitive areas is shown in Table 2.

For non-visually sensitive area (CTU of category 5), different selection methods are performed. After the spherical surface is projected into the ERP plane, since different latitudes adopt different degrees of pixel sampling, different pixel regions in different rectangular planes have different pixel redundancy. The degree of pixel redundancy at different heights in ERP can be measured by  $1/\cos \theta, \theta \in [-\pi/2, \pi/2]$ , indicating image area. The range of the upper and lower parts corresponds to  $\pm\pi/2$ . The  $1/\cos \theta$  at the low latitude has a small value, which means that the pixel redundancy is small, corresponding to the smaller quantization step size; the  $1/\cos \theta$  at the high latitude has a large value, the pixel redundancy is larger, and the large quantization step size is set. Therefore, the quantization step  $Q_{step}$  is a function of  $1/\cos \theta$  [22]. Adjusting  $Q_{step}$  requires indirect adjustment of QP. If a global QP is set through the configuration file, there is a  $QP_{offset}$  at different latitudes relative to tropical area. Setting the  $QP_{offset}$  ensures similar quantization at each latitude distortion. The relationship between the two in HEVC and its transformation is as follows:

$$Q_{step} = 2^{\frac{QP-4}{6}} \tag{6}$$

$$Q_{QP} = 4 + 6\log_2(Q_{step}) \tag{7}$$

Referring to Eq. (7), setting  $QP_{offset}$  as the logarithm function of  $1/\cos \theta$ , Eq. (8) is proposed for QP offset selection of non-visually sensitive area:

$$QP_{offset} = \begin{cases} \text{around}\{4 + 2\log_{1/2}[\cos(\theta)]\}, & 2\log_{1/2}[\cos(\theta)] < 11 \\ 11, & \text{others} \end{cases} \tag{8}$$

**Table 2** QP offset selection for the visually sensitive area

CTU category	$QP_{base}=22$	$QP_{base}=27$	$QP_{base}=32$	$QP_{base}=37$
1	-4	-5	-7	-8
2	-3	-4	-6	-7
3	-1	-2	-4	-5
4	0	-1	-3	-4

Where  $\text{around}(\cdot)$  indicates rounding. Considering that setting too large  $QP_{\text{offset}}$  will seriously affect subjective quality,  $QP_{\text{offset}}$  should be set less than 11 through repeated experiment adjustments. For the CTU of category 5,  $QP_{\text{offset}}$  increases as  $\cos\theta$  decreases. The QP offset selection for non-visually sensitive areas is complete.

### 3.4 Algorithm flow

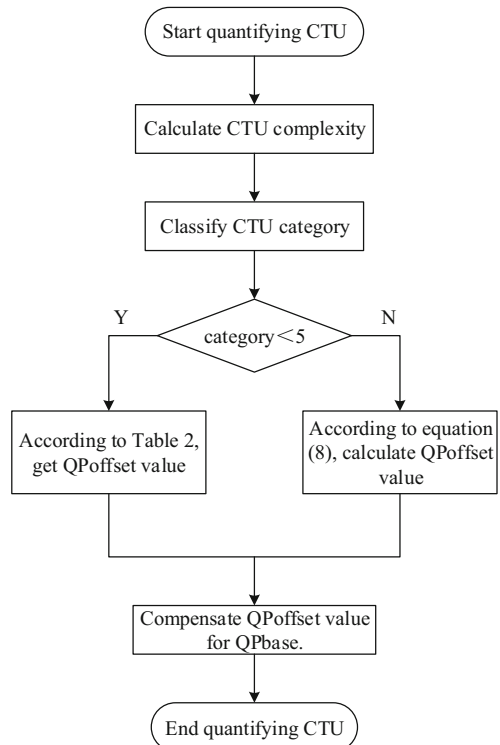
The algorithm flow is shown in Fig. 6. The main steps are as follows:

- (1) Get the current CTU pixel value, and calculate the complexity according to the Eqs. (4) and (5).
- (2) Classify CTUs according to the complexity and determine CTU category.
- (3) Determine QP selection method according to CTU category: the QP offset selection for visually sensitive areas (category 1, 2, 3, 4) is shown in Table 2. The QP offset value for non-visually sensitive area (category 5) is calculated by Eq. (8).

## 4 Experimental results and discussion

The proposed algorithm is integrated into HM16.20 and 360Lib-4.0 to test the coding performance and video image quality. The hardware platform for the experiment is set to

Fig. 6 Algorithm flowchart



Intel(R) Xeon(R) CPU E5–2640 v4, and the memory is 64.0GB. The main coding parameters are all I frame coding mode (All Intra Main10, AI-Main10), each sequence is encoded in 100 frames, and the initial QP is set to 22, 27, 32, 37 respectively. In order to evaluate the comprehensive coding performance of the proposed algorithm, BD-rate is calculated from the Excel table provided by JVET to measure the relationship between bitrate and image quality [19]. If  $\Delta$ BD-rate is negative, it means the overall coding performance is improved, and image quality is measured in *WS-PSNR* which is defined by *WMSE* as follows:

$$WS-PSNR = 10\log\left(\frac{MAX^2}{WMSE}\right) \tag{9}$$

$$WMSE = \sum_{i=0}^{width-1} \sum_{j=0}^{height-1} \left(y(i,j)-y'(i,j)\right)^2 \cdot W(i,j) \tag{10}$$

$$W(i,j) = \frac{w(i,j)}{\sum_{i=0}^{width-1} \sum_{j=0}^{height-1} w(i,j)} \tag{11}$$

Where *MAX* is the maximum value of the image pixel,  $y(i,j)$  and  $y'(i,j)$  represent the original pixel value and the reconstructed pixel value, respectively,  $W(i,j)$  is the weight scale factor of the normalized sphere. The calculation of the weight scale factor for different projection formats is different, and the weight scale factor for ERP projections is mentioned above. The calculation of  $\Delta$ *WS-PSNR* is as follows:

$$\Delta WS-PSNR = WS-PSNR_{proposed} - WS-PSNR_{HM16.20} \tag{12}$$

And the calculation of  $\Delta$ *Time* is as follows:

$$\Delta Time = \frac{T_{proposed} - T_{HM16.20}}{T_{HM16.20}} \times 100\% \tag{13}$$

This article uses sixteen standard test sequences recommended by GoPro [1], InterDigital [3], Nokia [16] and Letin VR [17]. For the accuracy of the quality assessment, the sequences are converted to low-resolution ERP videos before encoding. The encoding size is  $4096 \times 2048$  for 8 K and 6 K videos and the encoding size is  $3328 \times 1664$  for 4 K videos. The experimental data is shown in Table 3.

In Table 3, the experimental results show that compared with the standard algorithm, the *WS-PSNR* is increased by 0.40 dB, the BD-rate is reduced by 1.99% on average, and the coding time is increased by only 4.04%. The coding performance of the Gaslamp video test sequence, the SkateboardInLot video test sequence, the SkateboardTrick video test sequence, and the Harbor video test sequence are improved. This is because the video content is characterized by the existence of large texture complexes in the above five sequences. The lower-level regions account for the majority of the entire video content, while the regions with higher texture complexity are extremely complex. The CTUs of these regions belong to category 1 or 2. In this way, after a large number of CTUs with low texture complexity are compensated by QP, the code rate is greatly reduced, and some details of the image are lost, the distortion is enhanced, and the quality is degraded. However, since these regions are flat,

**Table 3** Comparison results of proposed method and HM16.20

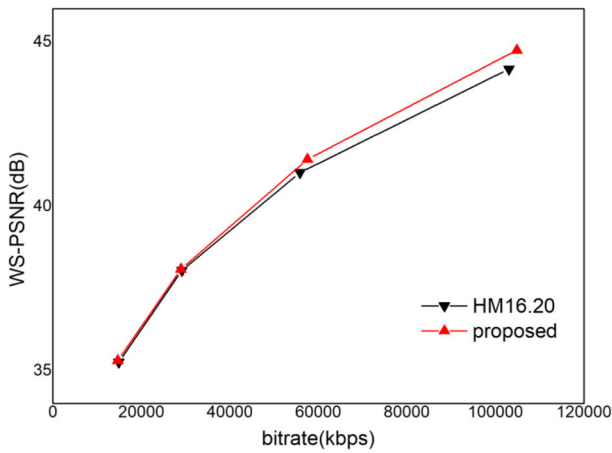
Class	Sequence	$\Delta$ WS-PSNR(dB)	$\Delta$ BD-rate(%)	$\Delta$ Time(%)
4 K	AerialCity	0.24	-2.14	2.63
	DrivingInCity	0.19	-1.54	0.32
	PoleVault_le	0.48	-2.21	2.70
	DrivingInCountry	0.47	-1.68	1.62
6 K	Balboa	0.35	-1.73	4.38
	BranCastle2	0.58	-1.90	6.73
	Broadway	0.31	-2.09	4.61
	Landing2	0.31	-1.46	4.43
8 K	ChairliftRide	0.37	-0.44	6.92
	Gaslamp	0.27	-3.82	4.10
	Harbor	0.32	-2.61	3.83
	KiteFlite	0.54	-1.55	5.57
	SkateboardInLot	0.35	-3.06	3.32
	SkateboardTrick	0.41	-2.70	3.62
	Train_le	0.58	-2.30	3.70
Average	Trolley	0.67	-0.67	6.13
		0.40	-1.99	4.04

there is less impact on the evaluation of the reconstructed video quality. However it has the effect of reducing the bitrate, while the CTU with less texture complexity is significantly compensated by the large QP negative offset selection. In addition, the quality of reconstructed video has been significantly improved. In contrast, the coding performance of the Trolley video test sequence, the Landing2 video test sequence, and the DrivingInCity video test sequence is improved slightly. One important reason is that there are a large number of areas with high texture complexity in all parts of the above three sequences, which are greatly compensated by QP. Another reason is that the code rate is too large and cannot be balanced well. Figure 7 shows the RD curve comparisons of different sequences.

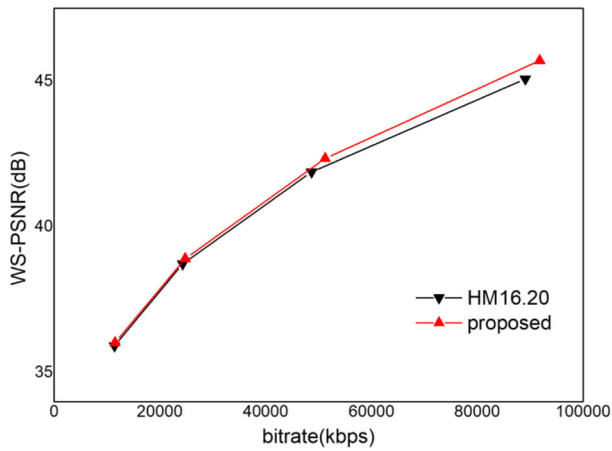
It is observed that the proposed algorithm outperforms the coding framework HM16.20 at both high bitrate and low bitrate, indicating that the reconstructed video quality can be improved by using the presented algorithm which researches the characteristics of 360-degree video. Smaller QP is set for the CTU of category 1–4 which is complex in texture to perform fine processing and larger QP is set for the CTU of category 5 which is simple in texture to perform rough processing. Compared to setting fixed QP value or strip QP to quantize, 360-degree video content can be quantified reasonably by this method.

To further evaluate the performance of the proposed algorithm, the algorithm proposed in this paper is compared with [13, 24], and the results are shown in Tables 4 and 5. It can be found that the algorithm proposed in this paper is superior to the reference algorithms in WS-PSNR and BD-rate for the same sequence, and can obtain better coding performance. This is due to that, compared with the algorithms in [13, 24], the proposed algorithm is more flexible to identify the plain block in tropical area and the complex block in polar area using CTU complexity.

As described above, the presented algorithm can signally improve the quality of the visually sensitive area in the image. In this part, several frames are intercepted to show subjective quality, and the details are enlarged to compare carefully. As seen in Figs. 8 and 9, the decoded images of the 8th frame of KiteFlite and the 4th frame of DrivingInCity are shown. Compared with the HM16.20, the presented algorithm can improve the block effect that exists in the



(a) Gaslamp



(b) SkateboardInLot

Fig. 7 RD curves for the proposed algorithm and the HM16.20 (a) Gaslamp (b) SkateboardInLot

image and result in clearer edge and detail. Although the objective quality improvement of the KiteFlite sequence is not obvious from the comprehensive data, the visual perception of the human eye is better. Since the visually sensitive areas are more refined, such as the text and graphics on billboard, the overall quality of the image has been improved from subjective

Table 4 Comparison of the proposed method and ref. [13]

Sequence	$\Delta$ BD-rate(%)
Harbor	1.0
SkateboardInLot	-0.1
SkateboardTrick	-2.3
Train_le	-1.7
Average	-0.8

**Table 5** Comparison of the proposed method and ref. [24]

Class	Sequence	$\Delta$ WS-PSNR(dB)	$\Delta$ BD-rate(%)
4 K	AerialCity	0.19	-2.34
	PoleVault_le	0.44	-0.51
6 K	Balboa	0.24	-3.03
	BranCastle2	0.57	-0.70
	Broadway	0.27	-3.59
	Landing2	0.29	-1.96
8 K	ChairliftRide	0.41	0.46
	Gaslamp	0.24	-4.72
	Harbor	0.31	-2.01
	KiteFlite	0.54	-0.35
	SkateboardInLot	0.30	-2.16
Average	Trolley	0.67	0.73
		0.37	-1.68

evaluation. In Fig. 9, since the camera and the cars are moving, the complex areas of the texture are further distorted. The text on the red bus glass is completely blurry and has a



(a) KiteFlite(HM16.20)



(b) KiteFlite(proposed)

**Fig. 8** Subjective Quality Comparison of KiteFlite (a) KiteFlite (HM16.20) (b) KiteFlite (proposed)



(a) DrivingInCity(HM16.20)



(b) DrivingInCity(proposed)

**Fig. 9** Subjective Quality Comparison of DrivingInCity (a) DrivingInCity (HM16.20) (b) DrivingInCity (proposed)

significant block effect in HM16.20. However, the text details can still be faintly recognized and the picture of block is smooth in the proposed algorithm.

In addition, the presented algorithm which is used for most test sequences outperforms the HM16.20 substantially in the higher bitrate segment, but the performance improvement is small in the lower bitrate segment. This shows that the proposed adaptive QP offset selection algorithm in this paper works better when optimizing video with higher bitrate.

## 5 Conclusions

In this paper, an adaptive QP offset selection algorithm based on CTU complexity is proposed. The proposed algorithm can achieve better quantization through fine processing of complex texture region and rough processing of simple texture region, and solves the problems of non-uniformity ERP sampling and the limitations of many QP offset selection algorithms. The algorithm mainly includes three parts: (1) Gradient is used to calculate the texture complexity of a CTU; (2) each CTU is classified into one of the five categories according to its



complexity; (3) different QP offset value is determined for each category. Smaller QP is set for the CTU of category 1–4 which is complex in texture to perform fine processing and larger QP is set for the CTU of category 5 which is simple in texture to perform rough processing. The experimental results show that compared with the HM16.20, the WS-PSNR increases by 0.40 dB, the BD-rate reduces by 1.99%, the time only increases by 4.04%, and the quality of visually sensitive areas has improved significantly.

**Acknowledgements** This work is supported by the Beijing Municipal Natural Science Foundation (No.4202018), Great Wall Scholar Project of Beijing Municipal Education Commission (CIT&TCD20180304), and the National Natural Science Foundation of China (No.61972023).

**Data availability** Some or all data, models, or code generated or used during the study are available from the corresponding author by request.

## Compliance with ethical standards

**Conflict of interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Abbas A, Adsumilli B (2016) AHG8: new GoPro test sequences for virtual reality video coding, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 4th Meeting, document JVET-D0026, Chengdu, China
2. Alshina E, Boyce J, Abbas A, Ye Y (editors) (2017) JVET common test conditions and evaluation procedures for 360° video, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 JVET-H1030, Macau
3. Asbun E, He Y, He Y, Ye Y (2016) AHG8: InterDigital test sequences for virtual reality video coding, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 4th Meeting, document JVET-D0039, Chengdu, China
4. Bai HH, Zhu C, Zhao Y (2007) Optimized multiple description lattice vector quantization for wavelet image coding. *IEEE Trans Circ Syst Video Tech* 17:912–917
5. Bai H, Lin W, Zhang M, Wang A, Zhao Y (2014) Multiple description video coding based on human visual system characteristics. *IEEE Trans Circ Syst Vid Technol* 24(8):1390–1394
6. Chevallier O, Zhou N, Cercueil JP, He J, Loffroy R, Wang YXJ (2019) Comparison of tri-exponential decay versus bi-exponential decay and full fitting versus segmented fitting for modeling liver intravoxel incoherent motion diffusion MRI. *NMR Biomed*, Article Number e4155, <https://doi.org/10.1002/nbm.4155>
7. Duanmu F, Kurdoglu E, Liu Y, Wang Y (2017) View Direction and Bandwidth Adaptive 360 Degree Video Streaming using a Two-Tier System, Proc. of IEEE International Symposium on Circuits and Systems, pp. 1–4, Baltimore, Maryland, USA
8. He YS, Ni LM (2019) A novel scheme based on the diffusion to edge detection. *IEEE Trans Image Process* 28(4):1613–1624
9. He Y, Vishwanath B (2016) AHG8: InterDigital’s projection format conversion tool, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 4th Meeting, document JVET-D0021, Chengdu, China
10. Hoang ND (2019) Automatic detection of asphalt pavement raveling using image texture based feature extraction and stochastic gradient descent logistic regression. *Automat Construct* 105
11. Jarvinen A (2016) Virtual reality as trend Contextualising an emerging consumer Technology into trend analysis, Proceedings of 2016 Future Technologies Conference (FTC), pp. 1065–1070
12. Khalil JE, Munteanu A, Lambert P (2019) Scalable wavelet-based coding of irregular meshes with interactive region-of-interest support. *IEEE Trans Circ Syst Vid Technol* 29(2):2067–2081
13. Li Y, Xu J, Chen Z (2017) Spherical domain rate distortion optimization for 360 degree video coding [C]. *IEEE International Conference on Multimedia and Expo (ICME)*, Hong Kong, 709–714

14. Papadopoulos MA, Zhang F, Agrafiotis D, Bull D (2016) AN ADAPTIVE QP OFFSET DETERMINATION METHOD FOR HEVC, IEEE International Conference on Image Processing ICIP, pp. 4220–4224
15. Racape F, Galpin F, Rath G, Francois E (2017) AHG8: adaptive QP for 360 video coding, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 JVET-F0038, Hobart
16. Schwarz S, Aminlou S (2016) Tampere pole vaulting sequence for virtual reality video coding, “Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 4th Meeting, document JVET-D0143, Chengdu, China
17. Sun W, Guo R (2016) Test sequences for virtual reality video coding from letinvr, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 4th Meeting, document JVET-D0179, Chengdu, China
18. Sun Y, Yu L (2017) Coding optimization based on weighted-to-spherically-uniform quality metric for 360 video [C]. IEEE Vis Commun Image Process (VCIP):1–4
19. Sun Y, Lu A, Yu L (2016) AHG8: WS-PSNR for 360 degree video objective quality evaluation, Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 4th Meeting, document JVET-D0040, Chengdu, China
20. Tang CZ, Wang O, Liu F, Tan P (2019) Joint Stabilization and Direction of 360 degrees Videos, ACM Trans Graphics 38(2)
21. Tran HTT, Ngoc NP, Bui CM, Pham MH, Thang TC (2017) An evaluation of quality metrics for 360 videos, 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, pp. 7–11
22. Xiang GQ, Jia HZ, Yang MY, Li Y, Xie XD (2018) A novel adaptive quantization method for video coding. *Multimed Tools Appl* 77(12):14817–14840
23. Xu YW, Yi SQ, Lin LQ, Chen WL, Zhao TS (2019) GOP structure-independent quantization parameter cascading in video coding. *IEEE Access* 7:76274–76282
24. Zhang MM, Zhang J, Liu Z, An CZ (2019) An efficient coding algorithm for 360-degree video based on improved adaptive QP Compensation and early CU partition termination. *Multimed Tools Appl* 78(1): 1081–1101
25. Zhao T, Wang Z, Chen CW (2016) Adaptive quantization parameter cascading in HEVC hierarchical coding. *IEEE Trans Image Process* 25(7):2997–3009
26. Zhou Y, Tian L, Zhu C, Jin X, Sun Y (2020) Video coding optimization for virtual reality 360-degree source [J]. *IEEE J Select Topics Signal Process* 14(1):118–129

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.