

# Perceptual sensitivity-based rate control method for high efficiency video coding

Huanqiang Zeng<sup>1</sup> · Aisheng Yang<sup>1</sup> · King Ngi Ngan<sup>2</sup> ·  
Miaohui Wang<sup>2</sup>

Received: 28 June 2015 / Revised: 20 September 2015 / Accepted: 8 October 2015 /  
Published online: 20 October 2015  
© Springer Science+Business Media New York 2015

**Abstract** The newest video coding standard—*high efficiency video coding* (HEVC) optimizes its coding efficiency in terms of *sum of square error* (SSE), which does not fully consider the perceptual characteristic of the input video. Thus, the HEVC is not effective in the sense of perceptual video coding. To address this problem, an efficient *perceptual sensitivity-based rate control* (PSRC) method for HEVC is proposed based on the *human visual system* (HVS) observation that the region with less perceptual sensitivity can tolerate more distortion. In the first stage, the proposed method develops a perceptual sensitivity measurement to evaluate the perceptual sensitivity of each *coding tree unit* (CTU) and each frame, which is then used to guide the bit allocation so that more bits will be allocated to those regions with higher perceptual sensitivity. To meet the target bits, an improved R- $\lambda$  model is utilized to determine the *quantization parameter* (QP). Experimental results have shown that the proposed method is able to improve the perceptual coding performance, compared with the original rate control in HEVC.

**Keywords** Perceptual video coding · HEVC · Human visual system · Rate control · Perceptual sensitivity

## 1 Introduction

*High efficiency video coding* (HEVC) is the newest video coding standard [21, 24], which is a successor of the H.264/AVC [10] and developed by *joint collaborative team on video coding* (JCT-VC) standard committee formed by *video coding experts group* (VCEG) and *moving picture experts group* (MPEG). Compared with H.264/AVC, HEVC is able to double the

---

✉ Huanqiang Zeng  
zeng0043@e.ntu.edu.sg

<sup>1</sup> School of Information Science and Engineering, Huaqiao University, Xiamen 361021, China

<sup>2</sup> Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, China

compression efficiency while keeping almost the same subjective quality of the reconstructed video. The improvement of coding efficiency is achieved by exploiting many new techniques, such as *coding unit* (CU) with variable block sizes varying from  $4 \times 4$  to  $64 \times 64$ , *prediction unit* (PU), *transform unit* (TU), quad-tree structure, advanced *motion vector* (MV) prediction, and so on [28]. Due to its high coding performance, the HEVC is expected to be used in various applications, such as HDTV, digital satellite broadcasting, and so on.

As we known, rate control plays an important role in video coding with the goal of maximizing video quality under the constraints of the bit rates and codec buffer in a robust and accurate manner. Similar to H.264/AVC, the HEVC also adopts Lagrangian *rate distortion optimization* (RDO) [29] to identify the optimal encoding setting for achieving the optimal *rate distortion* (RD) performance. However, the HEVC still optimizes the coding performance in terms of *sum of squared error* (SSE), which is not well correlated to the perceptual quality as discussed in [9]. Moreover, since the video quality is ultimately judged by human eyes, it is very essential to develop a *perceptual* rate control method for the HEVC with the goal of providing the optimal perceptual quality under a given bit rate.

In this paper, an efficient perceptual rate control method for the HEVC, called the *perceptual sensitivity-based rate control* (PSRC), is proposed by considering the perceptual characteristics of the input video content. In our approach, a perceptual sensitivity measurement is firstly developed to evaluate the perceptual sensitivity for each *coding tree unit* (CTU) and each frame. Then, the bit allocation is guided by the obtained perceptual sensitivity and an improved R- $\lambda$  model is exploited to determine the *quantization parameter* (QP) for meeting the allocated target bits. Experimental results have demonstrated that the proposed PSRC method is able to improve the perceptual coding performance, compared with the original rate control in HEVC.

The rest of this paper is organized as follows. The related works are briefly reviewed in Section 2. The proposed perceptual sensitivity-based rate control, PSRC, method for HEVC is presented in Section 3 in detail. Extensive simulation results are documented and discussed in Section 4. Finally, the summary is drawn in Section 5.

## 2 Related works

There have been many works on rate control for previous video coding standards (e.g., H.264/AVC) [5]. In these methods, the common way is to establish the rate distortion/quantization model based on the characteristics of the residual or the input video and then use this model to determine the suitable QP. Among them, considering the predicted residual follows the Laplacian distribution, Chiang et al. [6] presented a quadratic rate distortion model that uses the *mean absolute difference* (MAD) to estimate the complexity of the input video. Liu et al. [16] presented a switched MAD prediction scheme to reduce the MAD abruptness and a linear rate quantization model to describe the relationship between the bits and the QP. Kwon et al. [11] presented a rate control method for H.264, in which the inter dependency between RDO and rate control is addressed by QP estimation and update, and the bits for coding header information are estimated as a function of the nonzero MV elements. An et al. [1] suggested an iterative RDO method for H.264 by using the primal-dual decomposition and sub-gradient projection. Dong et al. [8] suggested a context-adaptive model parameter prediction scheme by using the spatial and temporal correlations so that the accuracy of the MAD, model parameter and bits matching could be significantly improved. Tsai et al. [23] improved the rate control

performance of intra coding by applying a Taylor series and scene change aware-based rate quantization step size model. However, these methods are not applicable to the HEVC, because HEVC exploits more complicated quad-tree coding structure with variable-sized CUs, PUs and TUs, which is different from that of previous video coding standards. Therefore, it is very necessary and more challenging to develop a novel rate control algorithm for HEVC.

For that, multiple rate control methods for HEVC have been recently proposed. Choi et al. [7] presented a quadratic pixel-based unified rate quantization model for HEVC based on the number of pixels by considering the size of PU varies from CU to CU. Based on the observation that the Lagrangian multiplier is a more important parameter to achieve the target bits than QP, Li et al. [13] suggested a rate- $\lambda$  model to perform rate control for HEVC instead of the conventional rate-quantization model. It should be pointed out that this rate- $\lambda$  model has been incorporated to the HEVC reference software. Moreover, Lee et al. [12] presented a frame-level rate control scheme for HEVC based on the texture and non-texture models. More specifically, the texture model is established by using the transform residual that is categorized into three accounts: low, medium and high-textured CUs, while the non-texture model is developed by considering the different characteristics of the non-texture bits in various depths of CUs. By taking into account the inter-frame dependency between the coding frame and its reference frame, Wang et al. [26] proposed an inter-frame dependency based rate and distortion models. Based on these models and a mixed Laplacian distribution of residual information, a new  $\rho$ -domain frame-level Rate-group of picture (Rate-GOP) based rate control is presented.

Unfortunately, the above-mentioned rate control methods for H.264/AVC and HEVC do not consider the *human visual system* (HVS) characteristic and might be inefficient in the sense of perceptual video coding. Since the video quality is ultimately judged by the human eye, how to develop a perceptual rate control method by incorporating the HVS characteristics attracts more and more attentions from both the academic community and industry. Although there are some existing perceptual rate control methods for H.264/AVC (e.g., [18, 27]), they cannot be directly applied to HEVC due to the different coding structures. For the HEVC, considering saliency represents the probability of human attention, Li et al. [14] incorporated the graph-based visual saliency into the quantization control so that the larger QP will be assigned to the CU with lower probability of attention. Li et al. [15] developed a weight-based *unified rate quantization* (URQ) scheme for perceptual video coding of conversational video in HEVC. Based on the observation that human is usually attracted by the face in conversational video, a hierarchical perceptual model of face is used to compute the weight map, which is then utilized to guide the bit allocation. In general, the perceptual rate control method for the HEVC has not been well investigated. Therefore, we focus on developing a perceptual rate control method for the HEVC.

### 3 Proposed Perceptual Sensitivity-Based Rate Control (PSRC) for HEVC

The proposed PSRC method consists of three parts as described in the following three subsections: 1) the perceptual sensitivity measurement is developed to evaluate the perceptual characteristic of the input video; 2) Bit allocation is performed based on the obtained perceptual sensitivity; and 3) An improved R- $\lambda$  model is utilized to meet the target bits, including QP determination and parameters update.

### 3.1 Perceptual sensitivity measurement

Obviously, the development of perceptual sensitivity measurement for guiding the video coding process shall consider the following requirements. First, the perceptual sensitivity measurement shall be able to well describe both the HVS perception and video coding distortions (e.g., quantization, artifacts). Second, the perceptual sensitivity measurement shall have low complexity and be easy to be incorporated into the video coding process (e.g., rate control). Therefore, although there are many existing visual quality assessment metrics that have good measurement of HVS perception [19, 20], they are not applicable to video coding process. As we know, the conventional video codec optimizes the RD performance in terms of *SSE*, which is computational efficient. However, the correlation between *SSE* and HVS perception is poor [29]. Further study finds that there is a strong linear relationship between the *mean squared error* (MSE) and HVS perception [2, 22].

Motivated by this, we propose a perceptual sensitivity measurement as a function of MSE, which can not only better indicate the HVS perception than MSE but also be easier to be incorporated into the video codec. For each CTU, the proposed *perceptual sensitivity measurement* (PSM) is defined by weighting the MSE based on its perceptual characteristic and the PSM of the whole frame can be computed by simply summing the *PSM* of each CTU in the frame:

$$\begin{aligned} PSM_i &= 1 + k_i \times MSE_i \\ PSMF &= \sum_{i=1}^N PSM_i \end{aligned} \quad (1)$$

where  $MSE_i$  is the mean squared error between the original and reconstructed  $i$ -th CTU,  $N$  is the number of CTU in the current frame  $f$ ,  $k_i$  is the perceptual weighting factor to indicate the perceptual characteristic of  $i$ -th CTU. Intuitively, the visibility of distortion in a video scene depends on its spatial texture and temporal motion. As a result, the perceptual weighting factor  $k_i$  is computed by considering both spatial texture and temporal motion. First, for the spatial texture, in general, from the spatial viewpoint, highly spatial texture region can tolerate more distortion than the low/medium spatial texture region. Hence, the *spatial texture complexity* (*STC*) is developed as the spatial perceptual feature for each CTU to evaluate its texture complexity:

$$STC_i = \frac{\sigma_m}{\sigma_f} \times \frac{1}{\sigma_i} \quad (2)$$

where  $\sigma_i$  is the variance of the  $i$ -th CTU in current frame  $f$ ,  $\sigma_f$  is the variance of the current frame, and  $\sigma_m$  is the variance of the mean value of all the CTUs in the current frame, which can be computed by:

$$\sigma_m = \frac{1}{N} \sum_{i=1}^N \left( m_i - \sum_{i=1}^N \frac{m_i}{N} \right)^2 \quad (3)$$

where  $m_i$  is the mean value of the  $i$ -th CTU in the current frame  $f$ ,  $N$  is the number of CTU in the current frame. This  $STC_i$  makes full use of two texture masking properties: global smoothness  $\frac{\sigma_m}{\sigma_f}$  and local contrast  $\frac{1}{\sigma_i}$ . One can see that the smaller the  $STC$  value is, the larger the variance is, which means that the current CTU is more likely to contain more complex texture information.

Second, for the temporal motion, from the temporal viewpoint, people are more interested in a moving region than a stationary region. Therefore, *temporal motion activity* (TMA) is presented as the temporal perceptual features for each CTU, namely, using the MV length to evaluate the motion activity. As suggested in [17], once the speed of the moving object is very fast that exceeds the spatial-temporal resolution capacity of humans, the moving regions will be smoothed and thus cause motion blurring. For these motion blurring regions, people always cannot perceive good visual quality and might ignore these regions. Therefore, the  $TMA_i$  for each CTU can be computed as:

$$TMA_i = \begin{cases} 1 & \text{if } (|x_i| + |y_i|) > L; \\ \sqrt{x_i^2 + y_i^2} + 1 & \text{otherwise.} \end{cases} \quad (4)$$

where  $MV_i = \{x_i, y_i\}$  is the MV of the  $i$ -th CTU in the current frame, and  $L$  is empirically-determined as 8 from the extensive experiments. We can see that a smaller  $TMA$  value indicates that the CTU has lower motion activity, especially  $TMA=1$  means that the CTU is likely to fall in a motionless region or a motion blurring region.

Based on the above analysis, for the current CTU, we can individually obtain two perceptual features—spatial texture complexity (i.e.,  $STC$ ) and temporal motion activity (i.e.,  $TMA$ ). In order to make the spatial  $STC$  and temporal  $TMA$  contribute equally to the evaluation of perceptual distortion, the simple strategy by using the product of these two perceptual features is exploited to compute perceptual weighting factor,  $k_i$ :

$$k_i = STC_i \times TMA_i \quad (5)$$

It should be pointed out that the proposed perceptual sensitivity measurement in (1) can well indicate the texture masking property of HVS. To be more specific, if the distortion with the same MSE occurs in both the complex texture and smooth CTUs, the perceptual quality reduction in the complex texture CTUs tends to be smaller than that in smooth one; similarly, if the distortion with the same MSE encounters in both the moving and stationary CTUs, the perceptual quality reduction will be easier to be perceived in the moving regions.

### 3.2 Bit allocation guided by perceptual sensitivity

For the perceptual rate control, the key problem is how to allocate the bits for each frame and each CTU based on their perceptual sensitivity, which can be computed by using the  $PSM$  developed in Section 3.1. And bit allocation is performed as the order of the GOP level, frame level, and CTU level as follows.

#### 1) Pre-analysis process

Before bit allocation, we perform a pre-analysis process to obtain the  $PSM$  of each CTU and frame. In this pre-analysis process, by using only one nearest previous frame as the reference frame, only the  $2N \times 2N$  partition mode for CTU is performed to get the MV for each CTU, and the  $MSE$  is computed between the original CTU and its best matching one. In this process, the spatial texture complexity  $STC_i$  and temporal motion activity  $TMA_i$  for each CTU can be computed by using (2) and (4),

respectively. Then, the PSM for each CTU  $PSM_i$  and each frame  $PSMF$  can be computed by using (1). Note that the first video frame is a special case that there is no reference frame for performing the motion estimation and thus the  $PSM_i$  are equal and directly set as 1 for all the CTUs in the first video frame.

## 2) GOP level bit allocation

Note that the GOP level bit allocation is conducted by using the commonly-used way as suggested in [13]. Suppose  $R_T$  is the target bit rate, the frame rate is  $f$ ,  $N_{GOP}$  is the GOP length,  $N_{coded}$  is the number of frame that has been coded,  $SW$  is the size of smooth window, which is used to make the bit rate change smoother,  $R_{used}$  means the bits that have been used. First, the average bits for each frame can be simply computed as:

$$R_{FraAvg} = \frac{R_T}{f} \quad (6)$$

Then, the target bits for each frame can be computed as below, consisting of two items: the average bits per frame and the buffer status.

$$T_{AvgFra} = R_{FraAvg} + \frac{R_{FraAvg} \cdot N_{coded} - R_{used}}{SW} \quad (7)$$

Finally, the target bits for the current GOP is

$$T_{GOP} = N_{GOP} \cdot T_{AvgFra} \quad (8)$$

## 3) Frame level bit allocation

Frame level bit allocation is performed after obtaining the bits for the current GOP. One can easily perceive that different frame in a GOP has different influence in the subjective quality of the whole GOP. Considering that the frame with higher perceptual sensitivity shall be assigned with more bits, the bits are allocated for each frame based on its perceptual sensitivity  $PSMF$  computed in (1). Therefore, the target bits of the  $j$ -th frame in the current GOP can be determined as:

$$T_{Fra}^j = \frac{T_{GOP} - R_{used}^{GOP}}{\sum_{NotCodedFrames} PSMF_i} PSMF_j \quad (9)$$

where  $R_{used}^{GOP}$  means the bits that have been used for coding the frames in the current GOP and  $PSMF_j$  is the weight of each frame in the current GOP. One can see that the higher perceptual sensitivity of the frame is, the more bits will be assigned.

## 4) CTU level bit allocation

Similarly, different CTU in a frame has different effect on the subjective quality of the whole frame. Hence, the target bits of current CTU  $T_{CurrCTU}$  are allocated according to its perceptual sensitivity computed in (1) as follows.

$$T_{CurrCTU} = \frac{T_{CurrFra} - Bit_{header} - Coded_{Fra}}{\sum_{NotCodedCTUs} PSM_i} \cdot PSM_{CurrCTU} \tag{10}$$

where  $Bit_{header}$  is the estimated bits of all headers, including slice header, MVs, prediction modes, etc., which is estimated according to the actual header bits of previous coded pictures belonging to the same level.

By using the above-mentioned bit allocation guided by perceptual sensitivity, one can see that the frame and the CTU with higher perceptual sensitivity value will be assigned with more bits so that the perceptual visual quality of the reconstructed video can be improved.

### 3.3 An improved R-λ model

After the CTU level bit allocation, the question becomes how to determine the  $QP$  so as to meet the target bits for each CTU. For simplicity, the R-λ model [13] is used, as it has higher accuracy than the traditional rate-quantization model:

$$\lambda = \alpha \cdot bpp^\beta \tag{11}$$

where  $\alpha$  and  $\beta$  are the model parameters that are related to the characteristic of the input video. Hence, different CTUs shall have different  $\alpha$  and  $\beta$  values. The initial values of  $\alpha$  and  $\beta$  are empirically-determined as 3.2003 and -1.367, respectively. Note that the initial values of  $\alpha$  and  $\beta$  are not critical, since they will keep updating in the encoding process. The update principle is based on the assumption that the collocated CTUs in different frames but belong to the same frame level may share the same parameters  $\alpha$  and  $\beta$ . And  $bpp$  means the bits per pixel, which can be computed as:

$$bpp = \frac{T_{CurrCTU}}{w \cdot h} \tag{12}$$

where  $w$  and  $h$  are width and height of the CTU, respectively.

One can see that the target bits  $T_{CurrCTU}$  for each CTU can be computed by (10) and  $T_{CurrCTU}$  is then used to compute  $bpp$  by (12). Finally, the  $\lambda$  value can be derived by (11) and is then exploited to determine the  $QP$  value according to the following equation (13).

$$QP = 6.3256 \ln \lambda + 21.8371 \tag{13}$$

To keep the visual quality consistency, the  $\lambda$  value and the determined  $QP$  value for the current CTU are clipped in a narrow range, as follows.

$$\lambda_{lastCTU} \cdot 2^{-\frac{1.0}{3.0}} \leq \lambda_{currCTU} \leq \lambda_{lastCTU} \cdot 2^{\frac{1.0}{3.0}} \tag{14}$$

$$QP_{lastCTU} - 3 \leq QP_{currCTU} \leq QP_{lastCTU} + 3$$

where  $\lambda_{currCTU}$  is the  $\lambda$  value of the current CTU,  $\lambda_{lastCTU}$  is the  $\lambda$  value of previous encoded CTU,  $QP_{currCTU}$  is the  $QP$  of the current CTU and  $QP_{lastCTU}$  is the  $QP$  of previous encoded CTU.

In order to adapt to the characteristic of the input video, the parameters need to be continually updated during the encoding process. In this work, there are some parameters: the perceptual sensitivity of each CTU  $PSM_i$  and each frame  $PSMF$ ,  $\alpha$  and  $\beta$ . The former two parameters will be updated in a GOP period in the pre-analysis process described in Section 3.2. The latter two parameters,  $\alpha$  and  $\beta$ , will be updated by using the real encoded  $bpp$  (i.e.,  $bpp_{real}$ ), the real used  $\lambda$  value (i.e.,  $\lambda_{real}$ ) and the perceptual sensitivity values as:

$$\begin{aligned}\lambda_{comp} &= \alpha_{old} \cdot \left( \frac{bpp_{real}}{PSM_i} \right)^{\beta_{old}} \\ \alpha_{new} &= \alpha_{old} + \delta_{\alpha} \cdot (\ln \lambda_{real} - \ln \lambda_{comp}) \cdot \alpha_{old} \\ \beta_{new} &= \beta_{old} + \delta_{\beta} \cdot (\ln \lambda_{real} - \ln \lambda_{comp})\end{aligned}\quad (15)$$

where  $\lambda_{comp}$  is the  $\lambda$  that is computed from the old model parameters by using  $bpp_{real}$  value and the PSM value,  $\delta_{\alpha}$  and  $\delta_{\beta}$  are empirically set to be 0.1 and 0.05, respectively. Note that the final values of  $\alpha$  and  $\beta$  are clipped in a pre-determined range [0.05, 20] and [-3.0, -0.1] as suggested in [13], respectively.

### 3.4 Proposed PSRC method

In summary, the proposed PSRC method can be described as below:

---

#### Function Encode-HEVC (Input video)

##### Begin

For each GOP in the input video

Allocate the target bits for the current GOP by (8)

Pre-analysis to compute the perceptual sensitivity  $PSM_i$  and  $PSMF$  by (1)

##### Begin

For each frame in the current GOP

1) Allocate the target bits for the current frame by (9)

##### Begin

For each CTU in the current frame

1.1) Allocate the target bits for the current CTU by (10)

1.2) Estimate the  $\lambda$  value by (11)

1.3) Compute the  $QP$  by (13)

1.4) Encode the current CTU

1.5) Update CTU-level parameters

##### End

2) Updating the Frame-level parameters

##### End

Updating the GOP-level parameters

##### End

---



**Table 1** The performance of proposed PSRC method under the Low delay setting and Random access setting, compared with the original rate control in HEVC [13]

Sequences	Resolutions	Low Delay (%)		Random Access (%)	
		$\Delta$ BR (%)	$\Delta$ SSIM	$\Delta$ BR (%)	$\Delta$ SSIM
Class A	2500×1600	-8.82	+0.0106	-6.65	+0.0074
Class B	1920×1080	-7.69	+0.0084	-6.03	+0.0065
Class C	832×480	-5.94	+0.0062	-5.27	+0.0052
Class D	416×240	-5.48	+0.0058	-4.91	+0.0040
Class E	1280×720	-6.77	+0.0076	-5.67	+0.0061
Average		-6.94	+0.0077	-5.71	+0.0058

## 4 Experimental results and discussions

To evaluate the performance, the proposed *perceptual sensitivity-based rate control* (PSRC) method is incorporated into the HEVC reference software (i.e., HM10.0) and tested on multiple commonly-used sequences. The test conditions are set as standard Low Delay with IPPP structure setting and Random Access setting as suggested in [4], respectively. Moreover, the QP value is set as 22, 27, 32 and 37, respectively. The first 100 frames of each test sequences are encoded. For each sequence under each QP, the target bit rate is set as the corresponding bit rate resulted from the same encoding setting without enabling the rate control scheme, and the initial QP is set as the current QP. All the simulation experiments are conducted on a computer with 3.6 GHz Intel i7 Core 4 processors, 8 GB memory, and Win 7 operating system.

In this work, we exploit the commonly-used visual quality assessment metric—SSIM [25] to measure the perceptual quality of the reconstructed video, which is obtained by simply averaging the corresponding SSIM value of each frame. The proposed PSRC method is compared with the original rate control method in HEVC [13]. And the average difference between their rate-SSIM curves is measured according to the method in [3] by using the following performance indexes:  $\Delta$ SSIM means the average SSIM changes;  $\Delta$ BR means the total bit rate changes (in percentage); “+” means increment; and “-” means decrement. In addition, *bit rate error* (BRE) [13] is used to indicate the mismatch between the target and real bit rate.

**Table 2** The bit rate error of the original rate control in HEVC [13] and proposed PSRC method for HEVC under the low delay and random access setting

Sequences	Resolutions	Low Delay (%)		Random Access (%)	
		[13]	Proposed	[13]	Proposed
Class A	2500×1600	0.58	0.61	1.22	1.37
Class B	1920×1080	0.36	0.34	0.94	1.11
Class C	832×480	0.44	0.48	0.82	0.95
Class D	416×240	0.32	0.33	1.04	1.06
Class E	1280×720	0.53	0.52	1.10	1.02
Average		0.45	0.46	1.02	1.10

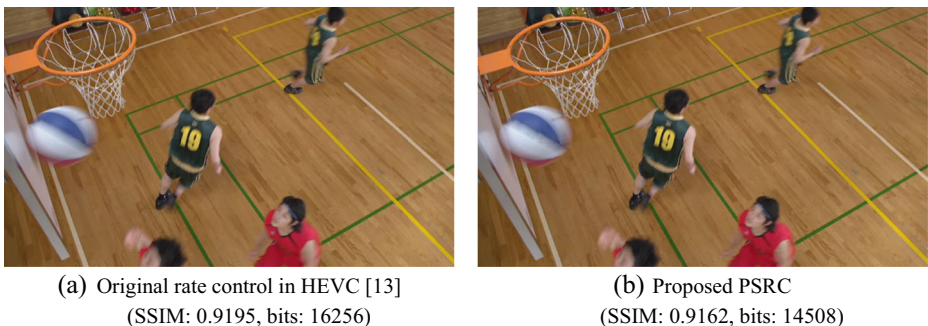


**Fig. 1** The reconstructed video frame—“*BQTerrace*”, 17th frame,  $QP=32$ , Low Delay Setting)

$$BRE = \frac{|R_{\text{target}} - R_{\text{actual}}|}{R_{\text{target}}} \times 100\% \quad (16)$$

where  $R_{\text{target}}$  and  $R_{\text{actual}}$  are the target bit rate and actual output bit rate, respectively. The lower  $BRE$  value indicates a better match between the target and real output bit rate.

Table 1 shows the performance of the proposed PSRC method for the HEVC under low delay and random access setting respectively, compared with the original rate control in HEVC [13]. One can see that the proposed PSRC method can achieve, on average, 6.94 % bit rate reduction and 0.0077 SSIM improvement for low delay setting. And for random access setting, on average, 5.71 % bit rate saving and 0.0058 SSIM increment can be obtained by the proposed PSRC method. In addition, Table 2 shows the bit rate error of the proposed PSRC method and the original rate control in HEVC [13] under low delay and random access setting, respectively. One can see that the proposed PSRC method is able to achieve similar bit rate mismatch between target bits and actual output bits to that of the original rate control in HEVC [13]. Moreover, Figs. 1 and 2 show the examples of the reconstructed frame by the original rate control in HEVC [13] and proposed PSRC method under Low Delay setting and Random Access setting, respectively. One can see that the proposed PSRC method can effectively allocate the bits according to the perceptual sensitivity of the input video so as to significantly reduce the bits while keeping the similar perceptual quality. Similar arguments can be applied to other sequences. Therefore, the proposed PSRC method is able to effectively improve the perceptual coding performance of the HEVC. In addition, another merit of the proposed PSRC method is that it has negligible computational overhead, as one can perceive that the complexity of the perceptual sensitivity measurement is fairly small, compared with the encoding process.



**Fig. 2** The reconstructed video frame—“*BasketballDrill*”, 3rd frame,  $QP=27$ , Random Access Setting)

## 5 Conclusions

In this paper, an efficient perceptual rate control method, called the *perceptual sensitivity-based rate control* (PSRC), is proposed for HEVC. The perceptual coding gain is achieved by adaptively allocating the bits for each frame and CTU in the RDO process based on their perceptual sensitivity. More specifically, the CTU with less perceptual sensitivity, which can tolerate more distortion, will be assigned with fewer bits. The above-mentioned perceptual sensitivity is evaluated by using two perceptual features that can effectively reflect the human perception, including spatial texture complexity and temporal motion activity. Experimental results have shown the efficiency of the proposed PSRC method on the improvement of the perceptual coding performance, compared with the original rate control in HEVC.

**Acknowledgments** This work was supported in part by a grant from the Research Grants Council of the Hong Kong SAR, China (Project CUHK 415712), in part by the National Natural Science Foundation of China under the Grants 61372107 and 61401167, in part by the Xiamen Key Science and Technology Project Foundation under the Grant 3502ZZ20133024, in part by the Opening Project of State Key Laboratory of Digital Publishing Technology under the grant FZDP2015-B-001, and in part by the High-Level Talent Project Foundation of Huaqiao University under the Grants 14BS201 and 14BS204.

## References

1. An C, Nguyen TQ (2008) Iterative rate-distortion optimization of H.264 with constant bit rate constraint. *IEEE Trans Image Process* 17(9):1605–1615
2. Bhat A, Kannangara S, Zhao Y, Richardson I (2012) A full reference quality metric for compressed video based on mean squared error and video content. *IEEE Trans Circ Syst Video Technol* 22(2):165–173
3. Bjontegaard G (2001) Calculation of average PSNR differences between RD-Curves. Document VCEG-M33, VCEG 13th meeting
4. Bossen F (2012) Common test conditions and software reference configurations. Document JCTVC-J1100
5. Chen Z, Ngan KN (2007) Recent advances in rate control for video coding. *Signal Process Image Commun* 22(1):19–38
6. Chiang T, Zhang Y-Q (1997) A new rate control scheme using quadratic rate distortion model. *IEEE Trans Circ Syst Video Technol* 7(1):246–250
7. Choi H, Nam J, Yoo J, Sim D, Bajić IV (2012) Rate control based on unified RQ model for HEVC. Joint Collaborative Team on Video Coding (JCT-VC), ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, 8th Meeting, pp. 1–10
8. Dong J, Ling N (2009) A context-adaptive prediction scheme for parameter estimation in H.264/AVC macroblock layer rate control. *IEEE Trans Circ Syst Video Technol* 19(8):1108–1117
9. Girod B (1993) What's wrong with mean squared error. In: *Digital Images and Human Vision*. Cambridge, Ma: MIT Press, pp. 207–220
10. ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 (2010) Advanced video coding for generic audiovisual services. JCTVC-A202, Dresden, Germany
11. Kwon D-K, Shen M-Y, Kuo C-CJ (2007) Rate control for H.264 video with enhanced rate and distortion models. *IEEE Trans Circ Syst Video Technol* 17(5):517–529
12. Lee B, Kim M, Nguyen TQ (2014) A frame-level rate control scheme based on texture and non-texture rate models for high efficiency video coding. *IEEE Trans Circ Syst Video Technol* 24(3):465–479
13. Li B, Li HQ, Li L, Zhang JL (2014) Lambda domain rate control algorithm for high efficiency video coding. *IEEE Trans Image Process* 23(9):3841–3854
14. Li Y, Liao W, Huang J, He D, Chen Z (2014) Saliency based perceptual HEVC. *IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1–5
15. Li S, Xu M, Deng X, Wang Z (2014) A novel weight-based URQ scheme for perceptual video coding of conversational video in HEVC. *IEEE International Conference on Multimedia & Expo*, pp. 1–6
16. Liu Y, Li ZG, Soh YC (2007) A novel rate control scheme for low delay video communication of H.264/AVC standard. *IEEE Trans Circ Syst Video Technol* 17(1):1152–1162

17. Oh H, Kim W (2013) Video processing for human perceptual visual quality-oriented video coding. *IEEE Trans Image Process* 4(4):1526–1535
18. Ou T, Huang Y, Chen HH (2011) SSIM-based perceptual rate control for video coding. *IEEE Trans Circ Syst Video Technol* 21(5):682–691
19. Ou YF, Ma Z, Liu T, Wang Y (2011) Perceptual quality assessment of video considering both frame rate and quantization artifacts. *IEEE Trans Circ Syst Video Technol* 21(3):286–298
20. Seshadrinathan K, Bovik AC (2010) Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Trans Image Process* 19(2):335–350
21. Sullivan J, Ohm JR, Han WJ, Wiegand T (2012) Overview of the High Efficiency Video Coding (HEVC) standard. *IEEE Trans Circ Syst Video Technol* 22(12):1649–1668
22. Tan HL, Li ZG, Tan YH, Rahardja S, Yeo C (2010) A perceptually relevant MSE-based image quality metric. *IEEE Trans Image Process* 19(2):335–350
23. Tsai W-J, Chou T-L (2010) Scene change aware intra-frame rate control for H.264/AVC. *IEEE Trans Circ Syst Video Technol* 20(12):1882–1886
24. Ugur K, Andersson K, Fuldseth A, Bjøntegaard G et al (2010) High performance, low complexity video coding and the emerging HEVC standard. *IEEE Trans Circ Syst Video Technol* 20(12):1688–1697
25. Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 13(4):600–612
26. Wang S, Ma S, Wang S, Zhao D, Gao W (2013) Rate-GOP based rate control for high efficiency video coding. *IEEE J Select Top Signal Process* 7(6):1101–1111
27. Wang S, Rehman A, Wang Z, Ma S, Gao W (2012) SSIM-motivated rate-distortion optimization for video coding. *IEEE Trans Circ Syst Video Technol* 22(4):516–529
28. Wang ZX, Zeng HQ, Chen J, Cai CH (2014) Key techniques of high efficiency video coding standard and its extension. *IEEE International Conference on Industrial Electronics and Applications (ICIEA2014)*, pp. 1169–1173
29. Wiegand T, Schwarz H, Joch A, Kossentini F, Sullivan GJ (2003) Rate-constrained coder control and comparison of video coding standards. *IEEE Trans Circ Syst Video Technol* 13(7):688–703



**Huanqiang Zeng** received the B.S. and M.S. degrees from Huaqiao University, Xiamen, China, and the Ph.D. degree from Nanyang Technological University, Singapore, all in electrical engineering. He was a Research Associate with Temasek Laboratories, Nanyang Technological University, and a Post-Doctoral Fellow with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong.

He is currently a Professor with the School of Information Science and Engineering, Huaqiao University. His research interests include visual information processing, video communication, 3-D/multiview video processing, and computer vision. Dr. Zeng has been serving as the Associate Editor for several international journals such as *International Journal of Image and Graphics*, the Technical Program Committee Member for multiple international conferences, and a reviewer for multiple international journals and conferences.



**Aisheng Yang** received the B.S. degree from Jiangxi Science & Technology Normal University, Nanchang, China. He is pursuing the master degree with the School of Information Science and Engineering, Huaqiao University. His research interests are image processing, perceptual video coding, and rate control.



**King Ngi Ngan** received his B.Sc. (Hons) and Ph.D. degrees, both in Electrical Engineering from Loughborough University, U.K., in 1978 and 1982, respectively. He joined the Department of Electronic Engineering, the Chinese University of Hong Kong as Chair Professor in 2003. He has been appointed Chair Professor at the University of Electronic Science and Technology, Chengdu, China, under the National Thousand Talents Program since 2012. Previously he was a Full Professor at the School of Computer Engineering, Nanyang Technological University, Singapore, and the School of Electrical, Electronic and Computer Engineering, University of Western Australia, Australia. He was appointed as Honorary Professor of Tunku Abdul Rahman University in Malaysia, Adjunct/Visiting Professor of Zhejiang University, Shanghai Jiaotong University, Shanghai University, Huazhong University of Science and Technology and Huaqiao University in China, Distinguished Adjunct Professor of the National Sun Yat-sen University in Taiwan, and Adjunct Professor of the University of Western Australia and RMIT University in Australia. He served as the External Examiner for the B.Eng. degree program of Malaysia's Multimedia University, and was appointed as IEEE Distinguished Lecturer of the IEEE Circuits and Systems Society in 2006–2007.

Prof. Ngan is active in many professional activities. He was an associate editor of the Journal on Visual Communications and Image Representation, U.S.A., and an area editor of EURASIP Journal of Image Communications, and served as an associate editor of IEEE Transactions on Circuits and Systems for Video Technology and Journal of Applied Signal Processing. He chaired a number of prestigious international conferences on video signal processing and communications and served on the advisory and technical

committees of numerous professional organizations. He co-chaired the IEEE International Conference on Image Processing (ICIP) in 2010 in Hong Kong. He has published extensively including 3 authored books, 6 edited volumes and over 300 refereed technical papers in the areas of visual signal processing, image/video coding and multimedia communications, which are his main research interests. He also holds 11 patents on image coding/transmission techniques.



**Miaohui Wang** is currently pursuing the Ph.D. degree in the Department of Electronic Engineering, the Chinese University of Hong Kong (CUHK). Before that he received his B.S. degree in Applied Mathematics from Xi'an University of Posts and Telecommunications (XUPT) and M.A. degree from School of Computer Science, Fudan University in 2007 and 2010 respectively. His current research interests cover a wide range of topics related with video coding, computer vision and code optimization, including rate control, transform coding, error concealment and image inpainting.

Mr. Wang received the Best Thesis Award in Shanghai city (2011) and Fudan University (2012), respectively.