A Rate Control Algorithm for HEVC Considering Visual Saliency

Henglu Wei, Wei Zhou ✉, Rui Bai and Zhemin Duan
Northwestern Polytechnical University, Xian, China
E-mail: zhouwei@nwpu.edu.cn

***Abstract*—In this paper, visual saliency is used to guide the coding tree unit (CTU) level bit allocation process in high efficiency video coding (HEVC) to improve the visual quality. At first, a saliency detection algorithm is proposed. With the detected saliency map, the distortion of each CTU is weighted by the corresponding saliency, so that the distortion of the salient areas is more critical. Then, the optimal bit allocation problem constraint by the picture level target bits and minimum quality fluctuation is built. Numerical method is used to solve the bit allocation problem. Experiment results show that quality gaining in salient areas is up to 0.8658 dB, the gaining of saliency weighted PSNR is up to 1.0318 dB.**

I. Introduction

A rate control scheme usually resolves two main problems. The first is how to allocate proper bits to each coding block according to the buffer status or bit budget, and the second is how to adjust the encoder parameters to properly encode each block with the allocated bits [1]. For the first problem, proper bits mean that the allocated bits should not only meet the constraint, but also reach the optimization goal. In [2], inter-frame dependency was considered to reflect the correlation among consecutive frames, and then efficient frame-level bits allocation algorithm was proposed. Li et al. [3] attempted to solve the coding tree unit (CTU) level optimal bit allocation problem, they developed an optimization formulation with rate distortion (R-D) estimation, and obtained a closed form solution by recursive Taylor expansion method. Li et al. [4] considered both inter-frame dependence and CTU level R-D performance. Many R-D models have been proposed to calculate λ and QP, e.g., Quadratic model [5], ρ-domain model [6] and λ-domain model [7].

All of the above rate control algorithms aimed at minimizing the average objective distortion, while human visual system (HVS) was not considered. For HVS based video coding, region of interest (ROI) coding is the hot topic. Some algorithms have been proposed to improve the quality in ROIs. Zeng et al. [8] developed perceptual sensitivity to guide bit allocation. Zhang et al. [9] built RD models for ROIs and non-ROIs respectively. Bai et al. [10] used the average saliency and mean absolute difference as weights, and allocated bits for each CTU according to these two weights. However, bit allocation for ROIs and non-ROIs in these algorithms is not optimal in RDO sense, as the bits difference between ROIs and non-ROIs is manually determined.

In this paper, we propose a saliency based CTU level rate control algorithm for Inter coded frames in HEVC. Our algorithm mainly focuses on the first problem, i.e., proper bit allocation. In the proposed rate control algorithm, saliency is used as the weight to measure the distortion in different regions, so that the overall distortion is penalized by the distortion in salient areas (SA), i.e., ROI. At first, a saliency detection algorithm is proposed to detect the saliency. Then, the optimal CTU level bit allocation problem is built constraint by frame level target bits and minimum weighted distortion. The problem is solved by the numerical method. Finally, the bits re-allocation and the RDO parameters clipping algorithms are designed to make up the mismatch between the allocated bits and the actually consumed bits.

The paper is organized as follows. The proposed saliency detection algorithm and saliency weighted distortion are presented in Section II. Section III introduces the proposed rate control algorithm. Section IV presents experimental results and Section V concludes the paper.

II. The Proposed Saliency Detection Algorithm

The proposed saliency detection algorithm is mainly composed of 3 parts: static saliency detection, dynamic saliency detection and fusion. In the proposed algorithm, information in compressed HEVC domain, besides information in pixel domain, is utilized to generate dynamic feature map to reduce complexity.

A. Static Saliency Detection

To detect static saliency, static features are extracted at first. In the proposed algorithm, four static feature maps are extracted: one luminance feature map, two chroma feature maps and one texture feature map. Luminance and chroma feature maps are the down-sampling of Y, Cb and Cr color component, represented by SF^L, SF^{Cb} and SF^{Cr} respectively. The down-sampling process can be expressed as,

$$\begin{cases} SF_k^L = \sum_{i=1}^{64} Y_{k,i}/64 \\ SF_k^{Cb} = \sum_{i=1}^{16} Cb_{k,i}/16 \\ SF_k^{Cr} = \sum_{i=1}^{16} Cr_{k,i}/16 \end{cases} \tag{1}$$

where $Y_{k,i}$ ($Cb_{k,i}$, $Cr_{k,i}$) is the i-th element in the k-th 8×8 (4×4) block of Y (Cb, Cr) color component. It should be noted that the video format in our experiment is 4:2:0 YCbCr, then the extracted luminance feature map and chroma feature maps are of the same size.

Texture feature can be represented by spatial frequency. DCT is a powerful tool to transform images from pixel domain into spatial frequency domain. Therefore, we use 8×8 DCT to extract texture features from Y component. Our texture feature map, represented by SF^T, is a set of vectors, each of which is composed of 5 lowest frequency DCT AC coefficients, as following,

$$SF_k^T = \left\{ AC_k^{(0,1)}, AC_k^{(1,0)}, AC_k^{(2,0)}, AC_k^{(1,1)}, AC_k^{(0,2)} \right\} \quad (2)$$

where $AC_k^{(i,j)}$ is the coefficient with coordinate (i, j) in the k-th DCT block.

Each static feature map can generate a saliency map by calculating the saliency of each static feature. The saliency of one static feature is measured by the distance from itself to the surroundings, as the following,

$$SM_k^F = \sum_{i \in W(k)} G(i,k) \cdot \|F_i - F_k\|_2 \quad (3)$$

where F_i is the i-th feature in the static feature map F, $F \in \left\{ SF^L, SF^{Cb}, SF^{Cr}, SF^T \right\}$, $W(k)$ is k-centric 9×9 square window, and $G(\cdot)$ is Gaussian kernel function. Finally, the static saliency map can be obtained by the following equation,

$$SM = 0.4 \cdot SM^L + 0.15 \cdot SM^{Cb} + 0.15 \cdot SM^{Cr} + 0.3 \cdot SM^T \quad (4)$$

B. Dynamic Saliency Detection

Dynamic feature indicates the movement in frame. It is time-consuming to exhaustively search the movement of the entire frame. To reduce the complexity, motion vector (MV) generated by the HEVC encoder is used instead. The MV of each 8×8 block composes the dynamic feature map, represented by DF. However, for dynamic background videos, the movement of background may cause comparable MV to that of foreground, which can introduce spurious SA. Therefore, the feature of dynamic background should be filtered out. Different from the complex and varied motion of foreground, the motion of background tends to be regular, which results in the relatively simple coding structure and smaller residual. In this paper, the mask containing coding structure and residual information is designed to filtered out the motion of dynamic background, as following,

$$mask_k = \begin{cases} 0, & if\ d_k \cdot r_k < Th \\ 1, & if\ d_k \cdot r_k > Th \end{cases} \quad (5)$$

where d_k is the splitting depth in CTU of the k-th 8×8 block, and r_k is the average residual of the k-th 8×8 block. The threshold Th is defined as,

$$Th = 2 \cdot \frac{\sum d_k \cdot r_k}{n} \quad (6)$$

where, n is number of pixels in a frame. Then, the dynamic saliency map can be obtained by normalizing and masking the dynamic feature map,

$$DM_k = \frac{\|DF_k\|_2}{\max\limits_k \|DF_k\|_2} \cdot mask_k \quad (7)$$

C. Adaptive Fusion

With the static and dynamic saliency map, the final saliency map can be obtained by fusion. In the proposed algorithm, the linear fusion scheme is used, as following,

$$FM_k = a_1 \cdot SM_k + a_2 \cdot DM_k + a_3 \cdot MM_k \quad (8)$$

where a_1, a_2 and a_3 are weighting factors, MM is the mixed saliency map, $MM_k = SM_k \cdot DM_k$. The weighting factors are defined as,

$$\begin{cases} a_1 = 1 \\ a_2 = \left(\frac{\sigma^{SM}}{\sigma^{DM}} \right)^{\frac{1}{2}} \\ a_3 = 2 \cdot \left(\frac{\sigma^{SM}}{\sigma^{FM}} \cdot \frac{\sigma^{DM}}{\sigma^{FM}} \right)^{\frac{1}{2}} \end{cases} \quad (9)$$

where σ is standard deviation of corresponding saliency map. After fusion, normalization is carried out, and the final saliency map is obtained.

III. The Proposed Rate Control Algorithm

A. Saliency Weighted Distortion

The traditional picture quality metrics, i.e., PSNR and MSE, have been widely criticized for not correlating well with perceived quality measurements [11]. The studies over the past few years have shown that the addition of video saliency maps improves the performance of most quality metrics [12]. Therefore, the visual saliency is used to weight the distortion, similar as [13], to be compliant well with the perceived quality. The saliency weighted distortion in this paper is defined as,

$$d^s = (0.01 + s)\, d = wd \quad (10)$$

where d is MSE of the CTU, s is the maximum visual saliency within the CTU and $w = 0.01 + s$ is the weighting factor.

As shown in (10), the weighting factor w is proportional to visual saliency s. For SA, the relative large weighting factor penalizes the objective distortion d, which is in accordance with the wide acknowledge that visual distortions appearing in salient regions might be more visible and, therefore, more annoying [14].

B. CTU Level Initial Bit Allocation

In [15], a linear distortion model, i.e., λ-D model, is proposed, where distortion is described by the Lagrange multiplier λ, as the following,

$$d^{curr} = \frac{d^{prev}}{\lambda^{prev}} \cdot \lambda^{curr} \quad (11)$$

where d^{curr} is MSE of the current picture, d^{prev} is MSE of the previous picture, λ^{curr} is the Lagrange multiplier of the current picture and λ^{prev} is the Lagrange multiplier of the previous picture. Another rate estimation model widely used in HEVC is λ-R model [7] which can be expressed as,

$$\lambda = \alpha \cdot bpp^{\beta} \quad (12)$$

where α and β are context related model parameters, and bpp is the abbreviation of bit per pixel.

In this paper, we use the above λ-D model and λ-R model to facilitate the proposed algorithm. However, the linear λ-D model is derived under the condition that the quality of consecutive frames in the video is consistent. To be compatible with hierarchy coding structure, the linear distortion model is modified before used, as the following,

$$d_i = p_i \cdot \lambda_i \tag{13}$$

where d_i is MSE of the *i*-th CTU in current frame, p_i is the ratio of MSE to the Lagrange multiplier of the *i*-th CTU in the previous same hierarchy frame, and λ_i is the Lagrange multiplier of the *i*-th CTU in current frame.

Combining (10) and (13), the weighted distortion for the *i*-th CTU in current frame is,

$$d_i^s = w_i p_i \lambda_i \tag{14}$$

Our goal is that the total distortion in the whole frame is minimized by proper bit allocation, i.e.,

$$\begin{aligned} d_{pic}^s &= \arg\min_{d_i^s} \sum_{i=1}^{N} n_i d_i^s \\ &= \arg\min_{bpp_i} \sum_{i=1}^{N} n_i w_i p_i \alpha_i bpp_i^{\beta_i} \end{aligned} \tag{15}$$

where d_{pic}^s is the picture level weighted distortion, n_i is the number of pixels in the *i*-th CTU, N is the number of CTUs in a frame, bpp_i is the *bpp* of the *i*-th CTU, and α_i and β_i are λ-R model parameters in CTU level. Besides, the total allocated bits for CTUs in the frame should also satisfy the target bitrate constraint that,

$$\sum_{i=1}^{N} n_i bpp_i \le R_{pic} \tag{16}$$

where R_{pic} is the target bits allocated to the current frame by the frame level rate control mechanism. Equations (15) and (16) constitute the constrained optimization problem, which can be elegantly solved by Lagrange multiplier method [16]. The Lagrange function of the constrained problem is,

$$L(bpp_i, u) = \sum_{i=1}^{N} n_i w_i p_i \alpha_i {bpp_i}^{\beta_i} + u\left(\sum_{i=1}^{N} n_i bpp_i - R_{pic}\right) \tag{17}$$

where u is the Lagrange multiplier. The Karush-Kuhn-Tucker (KKT) condition of equation (17) is,

$$\begin{cases} \frac{\partial L}{\partial bpp_i} = n_i w_i p_i \alpha_i \beta_i {bpp_i}^{\beta_i - 1} + u \cdot n_i = 0 & (i) \\ \frac{\partial L}{\partial u} = \sum\limits_{i=1}^{N} n_i bpp_i - R_{pic} = 0 & (ii) \end{cases} \tag{18}$$

From condition (i), the optimal bits allocated to the *i*-th CTU by the proposed algorithm can be obtained,

$$r_i^{init} = n_i \cdot bpp_i = n_i \left(-\frac{u}{w_i p_i \alpha_i \beta_i}\right)^{\frac{1}{\beta_i - 1}} \tag{19}$$

Taking bpp_i in condition (i) into condition (ii), the single-variable equation related with u can be obtained, as following,

$$\sum_{i=1}^{N} n_i \left(-\frac{u}{w_i p_i \alpha_i \beta_i}\right)^{\frac{1}{\beta_i - 1}} = R_{pic} \tag{20}$$

The above equation can also be viewed as the root of function f which is,

$$\begin{aligned} f(u) &= \sum_{i=1}^{N} n_i \left(-\tfrac{u}{w_i p_i \alpha_i \beta_i}\right)^{\frac{1}{\beta_i - 1}} - R_{pic} \\ &= \sum_{i=1}^{N} coeff_i \cdot u^{\frac{1}{\beta_i - 1}} - R_{pic} \end{aligned} \tag{21}$$

where,

$$coeff_i = n_i \cdot (-w_i p_i \alpha_i \beta_i)^{\frac{1}{1-\beta_i}} \tag{22}$$

Comparing (20) and (21), u in (20) is also the root of f, vice versa. In this paper, we use the Newton iterative method [17] to solve the root of f. The iterative scheme here is,

$$u_{n+1} = u_n - \frac{f(u_n)}{f'(u_n)} \tag{23}$$

where u_n is the *n*-th iteration value. It can be proved that the iteration in (23) necessarily converges with any positive initial value u_0.

With the root solved by the above numerical method, the optimal bits for the *i*-th CTU can be obtained by plugging it into (19).

C. CTU Level Rate Control

With the initially allocated bits above, the Lagrange multiplier for RDO can be calculated by λ-R model [7]. For each CTU, QP is calculated according to the λ-QP model [18]. Then, the CTU is encoded with the calculated λ and QP. Ideally, the actually consumed bits to encode the CTU exactly equal the allocated one. However, there nearly almost exists mismatch between the initially allocated bits and the actually consumed bits for the inaccuracy of λ-R model [7]. To make up this mismatch, bits error generated by the previous CTUs is assigned to all the following uncoded CTUs by weight. The re-allocated bits by the proposed method is,

$$r_i^{re} = r_i^{init} + \frac{\sum_{j=1}^{i-1}\left(r_j^{init} - r_j^{act}\right)}{\sum_{k=i}^{N} r_k^{init}} r_i^{init} \tag{24}$$

where r_i^{re} is the re-allocated bits for the *i*-th CTU, r_i^{re} is the initially allocated bits in Section III-B and r_j^{act} is the actually consumed bits for encoding the *j*-th CTU.

With the re-allocated bits r_i^{re}, the RDO parameters λ_i^{act} and QP_i^{act} for the *i*-th CTU can then be calculated [7], [18],

$$\begin{cases} \lambda_i^{act} = \alpha_i \left(\frac{r_i^{re}}{n_i}\right)^{\beta_i} \\ QP_i^{act} = \lfloor 4.2005 \cdot \ln \lambda_i^{act} + 13.7122 + 0.5 \rfloor \end{cases} \tag{25}$$

where $\lfloor \cdot \rfloor$ is the floor operation. But due to the inaccuracy of RD model, the bits error may accumulate gradually so that there is inadequate or excess remaining bits for the latter

CTUs. To reduce bits error accumulation, λ_i^{act} and QP_i^{act} are clipped around the picture level parameters,

$$\begin{cases} \lambda_i^{act} = clip\left(\lambda_{pic} \cdot 2^{-\frac{dQL(i)}{3}}, \lambda_{pic} \cdot 2^{\frac{dQU(i)}{3}}, \lambda_i^{act}\right) \\ QP_i^{act} = clip\left(QP_{pic} - dQL(i), QP_{pic} + dQU(i), QP_i^{act}\right) \end{cases} \tag{26}$$

where λ_{pic} is the picture level Lagrange parameter, QP_{pic} is the corresponding picture level QP, and $dQL\,(i)$ and $dQU\,(i)$ are saliency related functions. For salient CTUs, $dQL = 8$, $dQU = 2$, for non-salient CTUs, $dQL = 2$, $dQU = 8$. Then, λ_i^{act} and QP_i^{act} are used to encode the *i*-th CTU.

IV. Experimental Results

To evaluate the performance of the proposed algorithm, we implement it into HM [19]. Four bitrates, corresponding to four constant QP encoded bitrates, i.e., QP=22, QP=27, QP=32 and QP=37, are tested. In the following, QP is used to represent the corresponding bitrate for simplicity. In the experiments, data for saliency detection are from two public databases [20], [21], and sequences for rate control are HEVC common test sequences [22].

A. Performance Analysis of Saliency Detection

To evaluate the performance of the proposed saliency detection algorithm, area under the curve (AUC) is tested. The comparison results of AUC are listed in Table I, where LD is *lowdelay* configuration and RA is *randomaccess* configuration. It can be seen from Table I that about half AUCs by the proposed algorithm (LD, QP=22) are bigger than those by the other four algorithms. The average AUCs by the four classical algorithms are 0.1404, 0.0605, 0.1167 and 0.0792 smaller than the one by the proposed algorithm (LD, QP=22). Therefore, the proposed saliency detection algorithm is better than the other four classic algorithms in sense of AUC. The performances under different bitrates and encoder configurations are tested, as shown in Table I. The results show that the biggest AUC difference under four conditions is only 0.0389, therefore the performance of the proposed saliency detection algorithm remains stable for different bitrates and configurations in sense of AUC.

B. Bits Allocation Analysis

The bits ratio allocated to salient areas (SA) at bitrate QP=22 and QP=37 is counted, as shown in Table II. In our experiment, the top 30% CTUs in saliency descending order are regarded as salient, otherwise non-salient. It can be seen from Table II that bits allocated to SA by the proposed algorithm are more than those by HM for all sequences. The bits allocated to SA by the proposed algorithm are 24% more for LD and 30% more for RA at bitrate QP=22; and 27% more for LD and 18% more for RA at bitrate QP=37. As more bits are allocated, the quality in SA can be enhanced by the proposed algorithm.

C. Quality in SA and NSA

Quality in SA and non-salient areas (NSA) by HM, Bai [10] and the proposed algorithm are tested. The results of KristenAndSara are shown in Fig. 1 (a)-(c). From Fig. 1 (a), PSNR in SA by HM is much less than PSNR in NSA, which is improper as distortion in SA is more sensitive. Bais' [10] method increased the quality in SA, but still needs to be further improved, as shown in Fig. 1 (b). In the proposed algorithm, bits are allocated based on saliency, and the quality in SA is effectively increased, as shown in Fig. 1 (c). The more detail results at bitrate QP=22 are shown in Table III, where PSNR-S is PSNR for SA, PSNR-NS is PSNR for NSA. From Table III, the quality in SA by HM is about 1.2269 dB worse than that in NSA. Bais' method improves the quality in SA, but still 0.5887 dB worse than that in NSA. The quality in SA by the proposed algorithm is similar as or even better than that in NSA for every sequence.

D. Quality Comparison

The bitrate quality (RQ) curves by HM, Bai [10] and the proposed algorithm are presented in Fig. 1 (d). It can be seen from Fig. 1 (d) that RQ curve by HM is the best one when PSNR is used as criterion. As PSNR does not correlate well with perceptual quality, we use saliency weighted PSNR (wPSNR) to re-evaluate the quality [27]. The RQ curves of the three methods when wPSNR is used as quality criterion are presented in Fig. 1 (e). As bits are optimally allocated according to saliency, the RD performance by the proposed algorithm is better than the other two algorithms.

The more detail wPSNR comparison is presented in Table IV. From Table IV, the wPSNR gaining by Bai and the proposed algorithm is 0.1894 dB and 0.5093 dB at bitrate QP=22, and 0.3785 dB and 0.4633 dB at bitrate QP=32. Therefore, the proposed algorithm is effective at different bitrate and outperforms HM and Bais' method in sense of wPSNR.

To further evaluate the performance of the proposed rate control algorithm, BD-BR and BD-wPSNR [28] are tested, as shown in Table V. For LD configuration, BD-BR decreasing and BD-wPSNR gaining by the proposed algorithm are 12.61% and 0.4999 dB on average, about 6.32% and 0.2487 dB more than those by Bai [10].

E. Complexity and Bitrate Error Analysis

The bitrate error (BRE) by HM and the encoding time of the proposed algorithm vs. HM are tested, as shown in Table VI where CRC is encoding time ratio excluding saliency detection and CRCS is encoding time ratio including saliency detection. From Table VI, the BRE is 2.70% and 3.55% for HM and the proposed algorithm under LD, and 1.75% and 2.57% for HM and the proposed algorithm under RA. Therefore, the rate control accuracy by the proposed algorithm is slightly worse than that by HM under both LD and RA. The reason is that more bits are allocated in SA by the proposed algorithm, which causes bigger bits error accumulation with the same RD model error. The encoding complexity by the proposed rate

TABLE I
AUC PERFORMANCE COMPARISON

Sequences	Zhang [23]	Itti [24]	Seo [25]	Hou [26]	LD		RA	
					QP=22	QP=32	QP=22	QP=32
HallMonit	0.7071	0.7999	0.8393	0.7912	0.8434	0.8512	0.8723	0.8503
FOREMAN	0.5285	0.6905	0.5696	0.6543	0.7863	0.7654	0.7735	0.7471
HARBOUR	0.5236	0.5773	0.4273	0.4832	0.6403	0.6014	0.6278	0.6107
MOBILE	0.3339	0.4088	0.4287	0.4382	0.7003	0.7109	0.6987	0.7054
BQSquare	0.4693	0.5291	0.4935	0.5423	0.6076	0.6280	0.5867	0.6079
BasketballPass	0.6452	0.7905	0.6748	0.7252	0.7550	0.7520	0.7264	0.7187
Johnny	0.7532	0.8813	0.7915	0.8594	0.8868	0.8847	0.8744	0.8716
FourPeople	0.7350	0.6758	0.7323	0.8095	0.8206	0.8102	0.8211	0.8122
SlideEditing	0.5954	0.8559	0.6491	0.6956	0.8012	0.8182	0.8104	0.8280
SlideShow	0.7880	0.7892	0.7296	0.7284	0.7899	0.7803	0.7753	0.7988
KristenAndSara	0.8193	0.8163	0.8370	0.8643	0.9026	0.8947	0.9072	0.8939
Cactus	0.7158	0.7584	0.7256	0.7566	0.7655	0.7483	0.7746	0.7553
Average	0.6345	0.7144	0.6582	0.6957	0.7749	0.7704	0.7707	0.7667

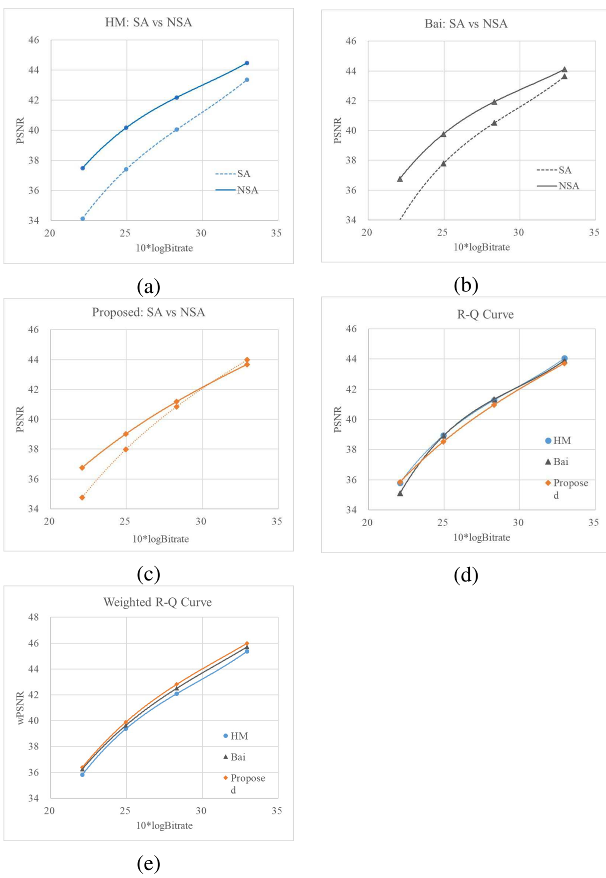

Fig. 1. RQ Curve Comparison of KristenAndSara by Three Methods (LD): (a) RQ curve of SA and NSA by HM (b) RQ curve of SA and NSA by Bai [10] (c) RQ curve of SA and NSA by proposed algorithm (d) RQ curve by four methods (e) weighted RQ curve by four methods.

TABLE II
BITS RATIO OF SALIENT AREAS

	Sequences	LD		RA	
		HM	Proposed	HM	Proposed
QP=22	FourPeople	56%	78%	51%	81%
	Johnny	46%	81%	39%	82%
	KristenAndSara	39%	75%	37%	77%
	BQTerrace	39%	57%	38%	62%
	Cactus	45%	66%	42%	69%
	Kimono	41%	66%	40%	67%
	ParkScene	40%	56%	41%	63%
	Average	44%	68%	41%	71%
QP=37	FourPeople	40%	68%	34%	42%
	Johnny	46%	69%	32%	39%
	KristenAndSara	37%	69%	32%	40%
	BQTerrace	31%	62%	30%	50%
	Cactus	47%	70%	38%	69%
	Kimono	42%	69%	40%	67%
	ParkScene	41%	66%	37%	67%
	Average	41%	68%	35%	53%

control algorithm is similar as HM. If take saliency detection into consideration, the overall encoding complexity by the proposed algorithm is about 1.03 times than HM under LD and 1.05 times than HM under RA.

V. CONCLUSION

In this paper, a saliency based CTU level rate control algorithm is proposed to enhance the quality of SA. Besides, a saliency detection algorithm is also proposed. The proposed saliency detection algorithm is superior to four classic algorithms in sense of AUC. By the proposed rate control algorithm, quality gaining in salient areas is up to 0.8658 dB, the gaining of saliency weighted PSNR is up to 1.0318 dB.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (61602383, 61772424), Fundamental Natural Science Research Funds of Shaanxi Province (2018JQ6016, 2017JQ6019), and Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University (CX201716).

TABLE III
PSNR COMPARISON BETWEEN SALIENT AREAS AND NON-SALIENT AREAS (QP=22)

Sequences		HM		Bai [10]		Proposed	
		PSNR-S (dB)	PSNR-NS (dB)	PSNR-S (dB)	PSNR-NS (dB)	PSNR-S (dB)	PSNR-NS (dB)
LD	FourPeople	42.2858	43.9335	42.2210	43.7542	42.2837	42.6391
	Johnny	41.7606	44.0353	41.8581	43.9266	42.2877	43.4869
	KristenAndSara	43.3700	44.4779	43.6454	44.1125	43.9784	43.6793
	BQTerrace	39.2224	40.2919	39.5715	39.5147	40.2805	39.2050
	Cactus	38.6633	39.4146	38.9410	38.9516	39.1404	38.7931
	Kimono	41.6842	42.1148	41.9683	41.5069	42.2903	41.0913
	ParkScene	39.4625	40.3550	39.3121	39.5440	40.1629	39.4123
	Average	40.9213	42.0890	41.0739	41.6158	41.4891	41.1867
RA	FourPeople	42.0497	43.7491	42.1187	43.5783	42.3294	43.1504
	Johnny	42.1096	44.2080	42.2066	44.0756	42.4514	43.7781
	KristenAndSara	43.3017	44.4014	43.3971	44.0855	43.6707	43.7969
	BQTerrace	39.0137	40.1616	39.2646	39.5927	39.8795	39.2652
	Cactus	38.4159	39.1981	38.6131	38.8174	38.7450	38.6104
	Kimono	38.6327	39.7718	38.9334	38.7850	39.2742	38.2692
	ParkScene	41.3236	41.9449	41.5022	41.2220	41.8601	40.8860
	Average	40.6924	41.9193	40.8622	41.4509	41.1729	41.1080

TABLE IV
WPSNR COMPARISON (LD)

Sequences		HM	Bai [10]	Proposed
QP=22	FourPeople	44.1391	44.1976	44.2930
	Johnny	43.9566	44.2282	44.4749
	KristenAndSara	45.3738	45.7281	45.9944
	BQTerrace	41.0747	41.2051	41.5680
	Cactus	40.4160	40.3824	40.9160
	Kimono	43.5802	44.0751	44.2288
	ParkScene	41.1140	41.1638	41.7447
	Average	42.8078	42.9972	43.3171
QP=32	FourPeople	38.3535	38.5571	38.7213
	Johnny	37.9287	37.8757	38.1162
	KristenAndSara	39.3634	39.9360	39.8926
	BQTerrace	35.0909	35.3397	35.4509
	Cactus	35.5392	35.8850	35.9149
	Kimono	37.7606	38.7474	38.7924
	ParkScene	34.9995	35.1676	35.3902
	Average	37.0051	37.3836	37.4684

TABLE V
BD-BR AND BD-WPSNR COMPARISON (LD)

Sequences	Bai [10]		Proposed	
	BD-BR (%)	BD-wPSNR (dB)	BD-BR (%)	BD-wPSNR (dB)
FourPeople	-4.42	0.2072	-8.06	0.3127
Johnny	-0.99	0.0142	-10.47	0.4715
KristenAndSara	-11.19	0.5322	-15.31	0.6592
BQTerrace	-7.41	0.3023	-12.43	0.5219
Cactus	-2.74	0.0211	-11.39	0.3112
Kimono	-16.82	0.6800	-18.73	0.7898
ParkScene	-0.44	0.0012	-11.85	0.4327
Average	-6.29	0.2512	-12.61	0.4999

TABLE VI
COMPLEXITY AND BIT ERROR COMPARISON (QP=27)

Sequences		HM	Proposed		
		BRE	BRE	CRC	CRCS
LD	FourPeople	2.56	2.87	98	105
	Johnny	3.22	4.04	99	104
	KristenAndSara	3.03	3.89	98	103
	BQTerrace	1.87	2.48	101	103
	Cactus	2.88	3.79	100	103
	Kimono	3.81	3.75	99	102
	ParkScene	1.50	4.06	99	103
	Average	2.70	3.55	99	103
RA	FourPeople	2.31	2.33	97	108
	Johnny	2.62	4.28	98	107
	KristenAndSara	2.43	2.37	99	107
	BQTerrace	1.31	2.19	100	104
	Cactus	1.35	3.06	101	104
	Kimono	1.04	1.22	98	103
	ParkScene	1.18	2.56	99	104
	Average	1.75	2.57	99	105

REFERENCES

[1] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for h.264/avc video coding and its application to rate control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1533–1544, 2005.

[2] J. He and F. Yang, "Efficient frame-level bit allocation algorithm for H.265/HEVC," *Iet Image Processing*, vol. 11, no. 4, pp. 245–257, 2017.

[3] S. Li, M. Xu, Z. Wang, and X. Sun, "Optimal bit allocation for CTU level rate control in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 11, pp. 2409–2424, 2017.

[4] L. Li, B. Li, H. Li, and C. W. Chen, "λ domain optimal bit allocation algorithm for high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 130–142, 2018.

[5] T. Chiang and Y. Q. Zhang, "A new rate control scheme using quadratic rate distortion model," vol. 7, no. 1, pp. 246–250, 1997.

[6] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for dct video coding via ρ-domain source modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 8, pp. 928–940, 2001.

[7] B. Li, H. Li, L. Li, and J. Zhang, "λ domain rate control algorithm for high efficiency video coding," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3841–3854, 2014.

[8] H. Zeng, A. Yang, K. N. Ngan, and M. Wang, "Perceptual sensitivity-based rate control method for high efficiency video coding," *Multimedia Tools and Applications*, vol. 75, no. 17, pp. 10383–10396, 2016.

[9] Z. Zhang, T. Jing, J. Han, Y. Xu, and F. Zhang, "A new rate control scheme for video coding based on region of interest," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2017.

[10] L. Bai, L. Song, R. Xie, J. Xie, and M. Chen, "Saliency based rate control scheme for high efficiency video coding," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2016, pp. 1–6.

[11] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[12] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Transactions on Circuits*

and Systems for Video Technology, vol. 21, no. 7, pp. 971–982, 2011.

[13] M. C. Farias, "Video quality assessment using visual attention computational models," *Journal of Electronic Imaging*, vol. 23, no. 6, pp. 061107, 2014.

[14] C. P. Le and E. Niebur, "Visual attention and applications in multimedia technologies," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2058–2067, 2013.

[15] M. Wang, K. N. Ngan, and H. Li, "Low-delay rate control for consistent quality using distortion-based lagrange multiplier," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 2943–2955, 2016.

[16] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic Press, 1982.

[17] E. Sli and D. F. Mayers, *An introduction to numerical analysis*, pp. 104–126, Cambridge University Press, 2003.

[18] B. Li, J. Xu, D. Zhang, and H. Li, "Qp refinement according to lagrange multiplier for high efficiency video coding," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, 2013, pp. 477–480.

[19] C. Rosewarne, B. Bross, M. Naccari, K. Sharman, and G. Sullivan, "High efficiency video coding (HEVC) test model 16 (HM 16) improved encoder description update 9," ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-AB1002, July 2017.

[20] H. Hadizadeh, M. J. Enriquez, and I. V. Baji, "Eye-tracking database for a set of standard video sequences," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 898–903, 2012.

[21] M. Xu, L. Jiang, X. Sun, Z. Ye, and Z. Wang, "Learning to detect video saliency with hevc features," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 369–385, 2017.

[22] K. Sharman, "Common test conditions," ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC-Z1100, January 2017.

[23] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "Sun: A bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8, no. 7, pp. 32.1–20, 2008.

[24] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision Research*, vol. 49, no. 10, pp. 1295, 2009.

[25] H. J. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," *Journal of Vision*, vol. 9, no. 12, pp. 15, 2009.

[26] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 194, 2012.

[27] W. Zhang, R. R. Martin, and H. Liu, "A saliency dispersion measure for improving saliency-based image quality metrics," *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.

[28] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T SG/16/Q6 Doc. VCEG-M33, April 2001.