IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 30, 2019, accepted December 14, 2019, date of publication December 19, 2019,
date of current version December 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2960807

HEVC/H.265 Intra Coding Based on the Human Visual System

DONGHUI FU[1,2], YANJIE WANG[1,3], BO FAN[1], AND NANNAN DING[1]
[1]Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
[2]University of the Chinese Academy of Sciences, Beijing 100049, China
[3]Key Laboratory of Airborne Optical Imaging and Measurement, Chinese Academy of Sciences, Changchun 130033, China

Corresponding author: Yanjie Wang (wangsybupt@163.com)

ABSTRACT High efficiency video coding (HEVC) is standard in the video compression field and performs well not only in video compression coding, but also in image compression. Regions of interest (ROIs) and just noticeable difference (JND), two human visual models, can accurately quantify human visual system (HVS) characteristics as pixel values. Using ROIs and JND to assist in evaluating image distortion can effectively reduce human visual redundancy and reflect authentic perceptual distortions. However, they are not readily applicable to the HEVC test model (HM) in the pixel domain. It is difficult to secure a suitable Lagrange multiplier λ and quantization parameter (QP) for the JND model in particular. This paper proposes different solutions for the use of rate control (RC) or not where an appropriate λ value is available for perceptual models. In RC, the proposed approach centers on a robust relationship between QP and achieved λ. And we also established a bit allocation technique using the related λ and expression for the RC model. Experimental results validate the rationality and effectiveness of the proposed method.

INDEX TERMS HEVC, ROI, JND, Lagrange multiplier λ, QP, bits allocation.

I. INTRODUCTION

Advancements in communications and semiconductor technology have brought about a new information era accompanied by revolutionary voice, text, picture, animation and video applications. Today, smartphones are regarded as a quotidian necessity. The sheer quantity of data that is transmitted across a single mobile app is enormous, and image information occupies a considerable proportion of said data. Transmitting the maximum quantity of information possible over limited resources is a problem which demands innovative compression algorithms. Scholars have proposed several image processing methods in effort to resolve this problem including JPEG [1] and JPEG2000 [2]. Video coding standards such as H.264/AVC [3] and HEVC [4] perform even better in terms of image compression.

Coding parameters, including the division of coding units (CUs), angles, and residual quad-trees (RQTs) are selected under HM methodology according to the relationship between bits and distortion [5]. The selection of any one parameter directly affects the others and all affect the overall image compression performance. Extant methods involve measuring distortion by SAD (The differences among SAD, SSE, MSE, and similar techniques are not distinguished in this paper for the sake of brevity). Extant methods thus fail to effectively reflect human vision characteristics. The distortion calculated in this manner has abundant visual redundancy and does not allow for effective compression performance.

HVS techniques center on the JND and ROI. Changes in a given image are revealed only if the difference between two corresponding pixels exceeds a certain threshold, the so-called JND [6]. If the difference is less than the value of the threshold, there is technically no distortion. Extant JND models fall into two main categories: Sub-band domain [7] and pixel-wise [8]. Only a portion of any given image is attractive to the human eye [10], namely, the ROI [11].

Bits and distortion are the two main elements of HEVC image compression, and are closely related to coding efficiency. To improve the coding efficiency, it is possible to enhance distortion while reducing bits or to maintain the bit quantity while raising image quality. The JND model can reveal accurate perceptual distortion values, while the ROI model enhances imaging quality in locations of strong visual interest under bit constraints. In terms of HVS from the two perspectives, the correlation coefficient between them is zero.

In other words, either can be applied to measure image distortion without accounting for the impact of the other.

There is indeed an intrinsic relationship between HVS and distortion in compressed perceptual images. ROI and JND

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaoqing Pan.

(a) (b) (c)

FIGURE 1. (a) Original image, (b) JND threshold image, (c) Saliency (ROI) image.

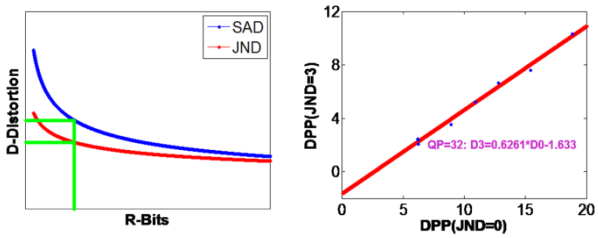
models can improve the image compression efficiency. There are numerous JND models which can precisely reflect the threshold of HVS [9], [12] and several widely accepted algorithms for obtaining image ROIs. Researchers have attempted a variety of models and techniques for image compression and made remarkable achievements.

Early scholars attempted to place more bits in ROIs and less in non-ROIs [34], [35] to manage the HVS. Li *et al.* [36] found, however, that low quality in non-ROIs can significantly decrease the image quality. To ensure the correct ratio between ROIs and non-ROIs, they proposed a method to detect face ROIs and assign optimal weight saliency values based on the GMM. They also established a closed-form bit allocation approach to minimize the perceptual distortion for the appropriate allocation of bits. They effectively adopted the saliency weighted PSNR (SWPSNR), but considered only one factor and did not make full use of the HVS.

The JND, as discussed above, can also be utilized to compute distortion. As per JND models [14] in the discrete cosine transform (DCT) domain, Zhang *et al.* [13] calculated distortion and coding bits with quantization steps from 1 to 255 for multiple bands in a JPEG compression experiment; they then calculated λ values and placed them into a JPEG quantization table corresponding to the HVS. This method allows for effective coding, but comes at a significant computational cost as there are different JNDs in multiple images with tables that must be calculated separately.

Bae *et al.* [9], [15] attempted to do more than simply estimate the JND value of a transform coefficient for a fixed-sized DCT kernel (e.g., 8×8), as many small transform coefficient values in mid- and high-frequency regions are otherwise not sufficiently suppressed. They proposed a HEVC-compliant local distortion detection probability (LDDP)-based PVC scheme which does perform relatively well. They also analyzed the effects of quantization distortion in JND models to build a new DCT-based energy-reduced JND (ERJND) model. This model consists of LR-JNQD and CNN-JNQD sub-models and has favorable bits saving performance [18], but still suffers a large computational burden.

Pixel-wise probability models are more convenient and readily operable than sub-band domain models, which require complex transformations encompassing various factors [16], [17]. However, the pixel-wise λ is difficult to achieve in practice. Zeng *et al.* [20] obtained a simplified fitting weight parameter according to neurophysiological



(a) (b)

FIGURE 2. (a) is RD cost curves and (b) is distortion fitting line (DPP: Distortion per pixel).

experiment results and Weber's ratio, but failed to take the quantization parameter (QP) into account as only the image content element was considered, the λ value obtained by this method is inaccurate.

In this study, we integrated ROI and JND models for HEVC intra perceptual image compression in the pixel domain. We propose the perceptual calibrated PSNR (PCPSNR) to supersede the traditional PSNR. A parameter change typically alters all relationships in the system. We replaced SAD (PSNR) with a new distortion parameter, so we also matched the distortion model for all parameters and established a new method for guiding the bit allocation process when rate control (RC) is executed in HM. The goal of this work is to achieve the optimal image compression performance.

Our contributions can be summarized as follows.

1) We use ROI and JND models in the pixel domain to measure distortion simultaneously which fully utilizes HVS characteristics, effectively reduces human visual redundancy, and accurately reflects perceptual image quality.

2) We recommend a general experimental method to obtain the λ value for the rate distortion optimization (RDO) of perceptual model.

3) We derived a stable relationship between QP and λ based on the perceptual model.

4) We developed a bit allocation approach using the above parameters to ensure RC accuracy.

The remainder of this paper is organized as follows. Section II discusses the related works on HEVC intra compression. Section III presents our methods and theoretical support thereof. Our experimental results are discussed in Section IV. Section V summarizes the paper and proposes future research directions.

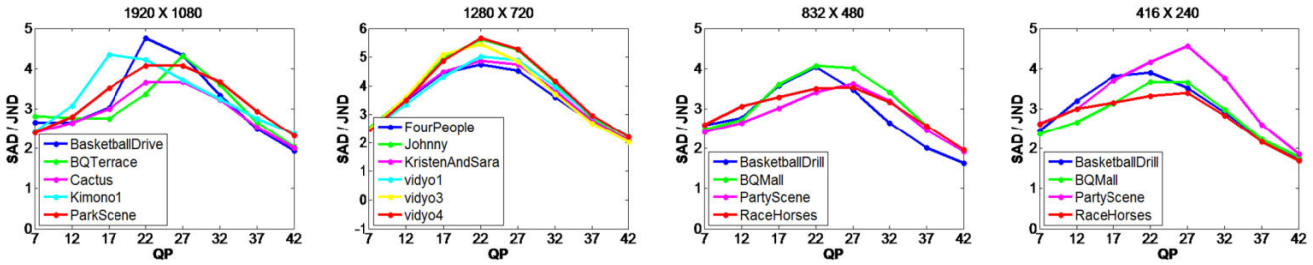


FIGURE 3. Distortion ratio in different QP.

II. RELATED WORK

Good tools make for good work. A suitable JND model and a high-quality saliency algorithm are necessary to secure accurate theoretical results.

A. JND MODELS

We selected a pixel-wise JND model for transplanting to the HM. A pixel-wise JND model [21], [22] reveals the JND threshold for every pixel. The subjective distortion between an original and a reconstructed pixel can be calculated conveniently. However, obtaining the threshold value over several critical factors is an overly complex process. The application conditions and weights of each factor form a certain relationship with each other after they are comprehensively analyzed.

The mechanisms of the human visibility are effectually a ''breakthrough point'' in this type of system [12], as the HVS is very sensitive to the repeatable visual patterns under which the JND threshold is evaluated [23]. Consider an image with two components, luminance contrast and pattern complexity, which determine the final pixel threshold. These components are the decisive factor for contrast-masking and pattern-masking, as their respective maxima are responsible for spatial-masking effects. Luminance adaptation can be used to determine the JND threshold as shown in Fig. 1(b).

Here, the pixel gradient is regarded as the fundamental factor affecting all external performance. Mathematical models were established with the gradient as argument for all components.

B. ROI MODELS

The number of bits allocated to a largest coding unit (LCU) (64×64) depends on the accuracy of the ROI region outline. Previous researchers have used various algorithms, methodologies, and influencing factors to determine saliency values. At present, deep learning is the most widely used and effective approach [42], [43].

To guarantee a correct ROI, images are usually divided into two categories: Face images [37] and generic images [24]. In this study, we predicted saliency to verify the performance of the proposed algorithm rather than test the proper ROI partition. Our method can be considered an image post-processing technique. We adopted a fixation prediction and saliency modeling framework based on inter-image similarities and ensemble of Extreme Learning Machines (ELM) [41]. In this framework, attention is modulated by both the contextual information of a scene and low-level visual cues. The influence of scene memorability on eye movement patterns is dependent on the resemblance of a scene to a former visual experience.

The fixation prediction algorithm uses images similar to the input image to train the extreme learners, which requires a ''retrieving set'' for the input images. The ELM can generate a set of predicted saliency values from the retrieved image set, then the saliency (Fig. 1(c)) of the given image can be measured in terms of the mean predicted saliency value per the ensemble's members. These values guide the subsequent perceptual distortion calculations.

C. R-D CURVE MODELS

The fitting model plays a vital role in any RC algorithm. In the quantization (Q)-domain model [25], Q is the determinant parameter to select the target bitrate R. The ρ-domain model [26], logarithmic model [27], exponential model [28], segmented model [29] and others can also be helpful to a certain extent for RC. According to Li *et al.* [5], exiting algorithms utilize a one-to-one correspondence relationship between R and Q, but as per the flexible coding unit (CU) size in HEVC, R should be decided by Q and other parameters not only by Q. They established a more accurate relationship for R-λ based on the hyperbolic function model [30], [31].

As stated above, bits and distortion are the key to image compression. λ is the slope of the R-D curve. The JND curve must fall under the SAD curve in the first quadrant. The blue and red curves in Fig. 2(a) represent the R-D curve of SAD and JND models, respectively. If SAD is simply replaced with the JND model, distortion reduction will be the only variable that changes between the two parameters – this is reflected in the green lines in Fig. 2(a).

There is a fixed relationship between QP and λ [32] when SAD is used and thus an inevitable connection between them in JND or ROI+JND models. In a single ROI model [36], the relationship change between QP and λ may be ignored, experiments have shown that the effect of this improvement is not particularly significant.

By contrast, as the JND model dramatically alters the distortion value, the conventional relational equation does not

TABLE 1. All tests list that we have done.

Test Model	Test Content (A_B [1])			
JND	Def_SAD	Def _JND	LUT_JND	Pre_JND
ROI	Def_SAD	Def _ROI	No	Pre_ROI
JR	Def_SAD	Def _JR	LUT_JR	Pre_JR

[1]About the format A_B, A is the way to get λ and B is the distortion used in RDO. Def means that λ is obtained by default method in HM, LUT and Pre mean that λ is obtained by lookup table and pre-processing method, respectively. JR means that it is a JND+ROI model.

hold. JND threshold values are content-related, so correlation coefficients differ for different images. The relationship between the two is closely related to their content. Here, we use ROI and JND models to describe subjective distortion. The goal of this study is to acquire the specific relationships for any image in JND and ROI+JND models.

III. DETERMINING QP AND LAMBDA

In HEVC, QP can be regarded as the most decisive parameter influencing the RDO and RC mode selection. If an effective RC is required, we need to know the optimal distortion; however, the distortion value can only be obtained after RDO. If we utilize RDO to select the mode to calculate the distortion, knowing QP is a prerequisite and λ is deduced by QP. In turn, QP is also derived from λ which can only be determined after RC. This is, effectively, a "chicken and egg" dilemma. Regardless of whether we calculate QP to λ or the λ to QP, the SAD distortion equation is based on a large number of experimental data statistics which are not applicable here. Various other strategies are available to determine the QP and λ, as discussed below.

A. λ FOR PERCEPTUAL RDO IN FIXED QP

As mentioned above, a known QP is the prerequisite for RDO. HM provides a ready-made equation to determine the λ for a general SAD model,

$$\lambda = c \cdot 2^{(QP-12)/3}, \tag{1}$$

where c is a fixed value.

Naturally, we expect a similar relationship for perceptual models. However, as the perceptual models are related to the image content, c must be a variable for different images in a subjective model, such an equation seems to not exist. In this paper, we propose that the λ value can be determined via pre-processing or a lookup table.

The R-λ model is important in terms of the value of λ. HEVC employs a complex coding structure encompassing a sequence, group of pictures (GOP), frame, and coding tree unit (CTU), and other elements. New R-λ approaches [5] have emerged as scholars seek to improve the processing ability for complex coding structures. The R-λ model is derived from the R-D model, but is more robust. Ardestani [34] and Li *et al.* [5] proved that the hyperbolic model performs exceptionally well. Our method is also based on the hyperbolic model. The R-D relationship is formulated as follows,

$$d = c \cdot r^{-k}, \tag{2}$$

then,

$$\lambda = -\frac{\partial d}{\partial r} = c \cdot k \cdot r^{-k-1} = k\frac{d}{r}, \tag{3}$$

where c and k are parameters related to the image content.

1) PRE-PROCESSING METHOD

Before formal compression, we can estimate the λ suitable for the image to be compressed by pre-processing. Here, we perform RDO with SAD and a default λ by setting the maximum CU depth to zero for all LCUs [11]. We also count the distortion and bits of the image for a perceptual model in this and the adjacent QP (minus or plus one) according to the previous mode and partition choices as D_1, D_2, R_1, and R_2, respectively. Equation (2) can thus be rewritten as follows,

$$\begin{aligned} D_1 &= c \cdot R_1^{-k} \\ D_2 &= c \cdot R_2^{-k}, \end{aligned} \tag{4}$$

so,

$$k = \frac{\ln\left(D_1/D_2\right)}{\ln\left(R_2/R_1\right)}. \tag{5}$$

Next, we can obtain the λ which is applied to the perceptual distorted model dependent on Eq. (3),

$$\overline{\lambda_p} = \frac{D_1 \ln\left(D_1/D_2\right)}{R_1 \ln\left(R_2/R_1\right)}. \tag{6}$$

2) LOOKUP TABLE METHOD

We built a one-to-one lookup table corresponding to the JND threshold and λ. Inspired by the multi-layer neural network, we divided this into two parts in order to obtain the most suitable λ: The constant JND threshold and the image JND threshold. We first made the JND threshold independent of

FIGURE 4. Test images. (a) is for getting the lambda of lookup table; (b) is used in subjective test.

the image, namely, we set the threshold as a constant. For simplicity, the JND threshold was uniform across all pixels, and the SAD model is a special case where all the thresholds are zero. Given a QP, for different images, the content and resolution may vary greatly, but for the distortion corresponding to the equal threshold, they have an incomplete but strong linear relationship, as shown in Fig. 2(b) and Eq. (7),

$$D_y = \mu \cdot D_x + t. \tag{7}$$

Taking the partial derivative of both sides with respect to R yields the following,

$$\lambda_y = \mu \cdot \lambda_x, \tag{8}$$

where are coefficient constants while x, y are the values of the JND threshold.

Equation (8) can be rewritten as,

$$\lambda_n = \mu_n \cdot \lambda_0, \tag{9}$$

where n is 0, 1, 2, …, N, and N is related to QP (presumably, $\mu_n = \mu_{n+1} = \ldots = \mu_{2^n1}$ when $D_N/D_{N\text{-}1} \leq 0.02$, and distortion is determined by QP). λ_0 can be obtained from Eq. (1) and by fitting the corresponding distorted experimental data as shown in Fig. 2(b). The distorted experimental data is obtained similarly to D_1 in Eq. (4), except that we set the maximum depth to 3 for all LCUs and the JND threshold of all pixels is n for D_n. μ_n is the value of the lookup table built previously.

We next associate the constant JND threshold with the image JND threshold. The constant JND threshold step we selected is one, so rounding is an effective method to analyze the distribution of the image JND threshold under the constant JND threshold. If M is the number of pixels in the image, M_n is the number of pixels under the image JND threshold equal to n after rounding. M_n/M can be regarded as the weight of λ_n in $\overline{\lambda_p}$. So the target is,

$$\overline{\lambda_p} = \sum_{n=0}^{N} \frac{M_n}{M} \lambda_n = \sum_{n=0}^{N} \mu_n \frac{M_n}{M} \lambda_0, \tag{10}$$

B. λ AND QP FOR PERCEPTUAL MODELS IN RC

As discussed above, pre-processing is a workable approach to complex problems. Here, we set the maximum CU depth to 0 for all CTUs first. Unlike the pre-processing discussed in Section III-A, the QP is not known. Instead, an appropriate value of bits is specified.

We continued to use the default HM method for pre-processing at this point in our attempt to secure necessary distortion and bits information. We first solved for λ (Eq. (4)) then derived the following parameters from Eq. (2) and Eq. (3),

$$k = \frac{\overline{\lambda} \cdot \overline{r}}{\overline{d}}, \tag{11}$$

$$c = \frac{\overline{d}}{\overline{r}^{-k}}, \tag{12}$$

where $\overline{d}$, $\overline{r}$, $\overline{\lambda}$ represents the pre-compressed distortion, bits, and λ as-obtained after pre-processing. As discussed in Section II-C, we analyzed the necessity of establishing a new λ-QP derivation relationship. By re-analyzing the experimental conditions for pre-processing and the data shown in Fig. 2(a), we found that an invariant R (bits) may be used as a bridge between the following two equations.

After pre-processing, we obtained necessary parameters such as c and k which can be considered valid for any coding level. We are interested in the frame and LCU level here, so we consider the following equation,

$$\lambda_{iSAD} = c_{iSAD} \cdot k_{iSAD} \cdot (r_{iSAD})^{-k_{iSAD}-1}, \tag{13}$$

$$\lambda_{iJR} = c_{iJR} \cdot k_{iJR} \cdot (r_{iJR})^{-k_{iJR}-1}, \tag{14}$$

where the subscript i represents the i-th LCU, SAD and JR (JND + ROI) reflect the method for evaluating the distortion. Under the given pre-processing conditions, it is obvious that $r_{iSAD} = r_{iJR}$. Therefore, we can integrate Eq. (13) and Eq. (14) as follows,

$$\lambda_{iSAD} = \alpha \cdot (\lambda_{iJR})^{\beta}, \tag{15}$$

where $\alpha = \frac{c_{iSAD} \times k_{iSAD}}{(c_{iJR} \times k_{iJR})^{\frac{k_{iSAD}+1}{k_{iJR}+1}}}$, $\beta = \frac{k_{iSAD}+1}{k_{iJR}+1}$ (The values of alpha and beta for a frame of image in Table 16, Appendix).

Thus,

$$\ln(\lambda_{iSAD}) = \beta \cdot \ln(\lambda_{iJR}) + \ln \alpha, \tag{16}$$

The Joint Collaborative Team on Video Coding (JCT-VC) determines QP by λ value as follows,

$$QP = 4.2005 \ln(\lambda_{iSAD}) + 13.7122, \tag{17}$$

We combined Eqs. (16) and (17) to suit the ROI and JND models as follows,

$$QP = 4.2005\beta \cdot \ln(\lambda_{iJR}) + (4.2005 \ln \alpha + 13.7122). \tag{18}$$

The case we illustrate suits the LCU level, but may also fit the conditions in any level.

C. BITS ALLOCATION

For formal compression, we considered the RDO between perceptual distortion and bits. First, we associated saliency and threshold values with perceptual distortion. The saliency and JND threshold values are implemented in the pixel domain, which makes our application very direct. Saliency values can be converted to weighted distortion [38]. As the threshold values well reflect the perceptual tolerance to distortion for every pixel, we can measure the perceptual distortion as follows,

$$d(x,y) = \begin{cases} 0, \\ \quad if\ |F_{ORG}(x,y) - F_{REC}(x,y)| < JND(x,y) \\ \omega(x,y) \cdot ||F_{ORG}(x,y) - F_{REC}(x,y)| - JND(x,y)| \\ \quad if\ others, \end{cases} \tag{19}$$

TABLE 2. Performance of different lambda in JND models.

Test Image	BD-PCPSNR/dB(JND)			BD-BR/%(JND)			Time/Second(JND)		
	Def _JND	LUT_JND	Pre_JND	Def _JND	LUT_JND	Pre_JND	Def _JND	LUT_JND	Pre_JND
PeopleOnStreet	0.070153	0.343788	0.486857	-1.205979	-5.502993	-7.628789	234.4	248.3	315.1
Traffic	0.106740	0.278748	0.337563	-2.158832	-5.469289	-6.595417	235.5	249.8	316.8
BQTerrace	0.095891	0.270685	0.363523	-1.732728	-4.331562	-5.980068	121.2	128.8	164.3
Cactus	0.089129	0.202878	0.295830	-2.519541	-4.712805	-6.640364	113.9	118.6	152.4
BasketballDrill	0.093844	0.193983	0.295224	-1.928774	-4.922680	-4.316916	22.4	24.2	30.6
BQMall	0.104263	0.343679	0.451396	-1.85645	-5.126989	-6.590664	23.7	25.1	31.5
BlowingBubbles	0.123886	0.285847	0.340968	-2.659077	-4.635842	-5.357627	6.3	6.6	8.5
RaceHorses	0.091185	0.294332	0.328010	-1.658177	-4.130583	-4.571345	6.2	6.5	8.2
Johnny	0.053566	0.259986	0.508083	-1.180666	-4.867373	-9.092498	42.9	44.9	58.9
KristenAndSara	0.051834	0.312687	0.481399	-0.976568	-4.766209	-7.206698	44.6	47.4	60.4
vidyo1	0.081594	0.349655	0.542139	-1.514364	-6.387380	-9.031059	46.8	49.8	63.4
vidyo4	0.024680	0.310191	0.589813	-0.563061	-5.446240	-9.765661	45.7	48.5	61.9
average	**0.082230**	**0.287205**	**0.418400**	**-1.662851**	**-4.895875**	**-6.898092**	**78.6**	**83.2**	**106.0**

TABLE 3. Performance of different lambda in ROI and JND+ROI models.

Test Image	BD-PCPSNR/dB					BD-BR/%				
	Def _ROI	Pre_ROI	Def _JR	LUT_JR	Pre_JR	Def _ROI	Pre_ROI	Def _JR	LUT_JR	Pre_JR
PeopleOnStreet	0.263019	0.337384	0.637947	0.748319	0.884455	-7.440547	-6.057751	-9.835097	-11.593710	-13.420231
Traffic	0.238140	0.213974	0.466431	0.542562	0.605744	-4.716247	-4.255392	-8.797534	-10.361499	-11.367714
BQTerrace	0.489475	0.434636	0.741006	0.795006	0.876430	-8.639245	-7.783370	-12.417289	-12.709807	-13.574052
Cactus	0.174309	0.144204	0.405535	0.434101	0.506613	-4.232563	-3.577706	-9.303278	-10.126350	-11.225031
BasketballDrill	0.584504	0.496046	0.780452	0.806177	0.833128	-11.966164	-10.300575	-13.079140	-13.039949	-13.826074
BQMall	0.299314	0.278849	0.706798	0.733754	0.845528	-4.770547	-4.412005	-10.155101	-10.174327	-11.666697
BlowingBubbles	0.310220	0.247880	0.574553	0.583012	0.684205	-5.701572	-4.583544	-8.885324	-8.847686	-10.366008
RaceHorses	0.285360	0.236336	0.543332	0.559415	0.600187	-3.691204	-3.734110	-7.328601	-7.509503	-7.967027
Johnny	0.291789	0.350819	0.972382	1.027902	1.038112	-5.849980	-7.003785	-13.388072	-15.726956	-15.941697
KristenAndSara	0.244025	0.259317	0.745493	0.782367	0.891607	-4.337609	-4.587824	-10.337523	-10.923153	-12.187815
vidyo1	0.375542	0.356835	0.851464	0.882977	1.032565	-7.447285	-7.055535	-13.372244	-14.478080	-15.944850
vidyo4	0.433999	0.394398	0.773809	0.848196	1.101594	-9.221060	-8.398764	-13.030651	-14.450265	-18.018893
average	**0.332475**	**0.312557**	**0.683267**	**0.728649**	**0.825014**	**-6.556221**	**-5.979197**	**-10.827488**	**-11.661774**	**-12.958841**

$$\omega_i = \frac{N \cdot s_i}{\sum_{i=0}^{N} s_i}, \tag{20}$$

where F_{ORG} (x, y), F_{REC}(x, y), JND(x, y), d(x, y), and ω(x, y) represent the original pixel, reconstructed pixel, JND threshold, subjective distortion, and weight in position (x, y) values, respectively, s_i is the average saliency value of the i-th LCU, ω_i is the weight of the i-th LCU ($\omega(x, y)$ and ω_i correspond one by one according to the pixel position), and N is the number of LCUs in an image.

After clarifying the distortion question, we operated the proposed bit allocation method.

Eq. (3) continues to be used here, but we wrote another formula for the whole image,

$$\lambda = C_F K_F \cdot R_F^{-K_F - 1}, \tag{21}$$

the subscript F indicates that the parameter is associated with a frame image. We assume the target bitrate of an image is R and plug this R into Eq. (21). The Lagrange multiplier for one frame is calculated as,

$$\lambda_{FR} = C_F K_F \cdot R^{-K_F - 1}, \tag{22}$$

If the RC is not used in HM, the Lagrange multiplier in the RDO is obtained by QP (Eq. (1)). The value is valid throughout the entire calculation process and never changes. Here, we regard the λ_{FR} as a parameter and apply it to each LCU in the system. Owing to the one-to-one correspondence between R and λ, when the Lagrange multiplier is plugged into Eq. (14), then,

$$r_{iJR} = \left(\frac{\lambda_{FR}}{c_{iJR} k_{iJR}}\right)^{-\frac{1}{k_{iJR}-R+1}}, \tag{23}$$

$$R_{iJR-\omega} = \frac{r_{iJR}}{\sum_{i=0}^{N} r_{iJR}} \cdot R, \tag{24}$$

where $R_{iJR-\omega}$ reflects the weighted bits for each LCU dependent on the ROI and JND models. So,

$$\lambda_{iJR-\omega} = c_{iJR} \cdot k_{iJR} \cdot R_{iJR-\omega}^{-k_{iJR}-1}. \tag{25}$$

Next,

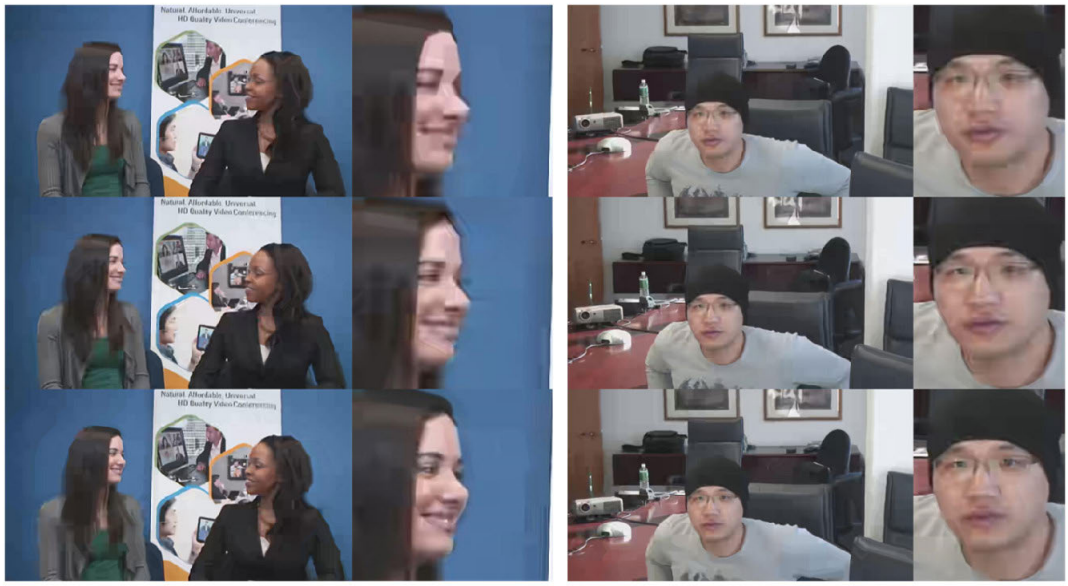
$$QP_i = 4.2005\beta \cdot \ln(\lambda_{iJR-\omega}) + (4.2005 \ln\alpha + 13.7122), \tag{26}$$

where $\lambda_{iJR-\omega}$ and QP_i are the value for every LCU in the perceptual model.

Regardless of how precise the bit allocation is, there are slight deviations in the actual coding process. To reduce the influence of unavoidable differences on RC performance, we can update the remaining bits at a certain frequency and

TABLE 4. Accuracy of bit allocation.

Test Content	Bits Cost/bit						Bits Errors/%					
	BPPP	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0	average
Def_SAD	BPP (actual)	0.2029	0.3999	0.5982	0.7983	0.9895	1.5038	1.9445	2.2795	1.7217	1.9634	**1.8826**
Def_JND		0.1998	0.3848	0.5717	0.7834	0.9631	2.4261	5.1093	5.8565	3.0929	4.2587	**4.1487**
Def_JR		0.2017	0.3881	0.5749	0.7838	0.9647	1.6163	4.4345	5.5796	2.9034	4.0923	**3.7252**
Pre_JND		0.2023	0.4041	0.6064	0.8182	1.0143	2.4893	1.3656	1.2053	2.2746	1.4320	**1.7534**
Pre_JR		0.2033	0.4041	0.6029	0.8097	1.0074	1.6494	1.0594	0.6169	1.3671	1.0375	**1.1460**



(a) (b)

FIGURE 5. Perceptual performance of face image in SAD (up)/JND (middle)/JND+ROI models (down) under RC condition (BPPP = 0.03 (a), 0.03 (b)).

TABLE 5. Performance in JND and JND+ROI models with our pre-processing method.

Test Image	BD-PCPSNR/dB				BD-BR/%			
	Def_JND	Pre_JND	Def_JR	Pre_JR	Def_JND	Pre_JND	Def_JR	Pre_JR
PeopleOnStreet	0.185147	0.447414	0.472616	0.741767	-4.475397	-8.535290	-6.254770	-10.080133
Traffic	0.215257	0.527676	0.413026	0.887033	-4.396385	-10.676939	-7.479380	-16.170776
BQTerrace	0.168459	0.522510	0.699717	1.441815	-3.013772	-7.120568	-9.837524	-20.194302
Cactus	0.185616	0.384200	0.401568	1.074158	-5.574648	-7.763706	-8.758369	-21.904357
BasketballDrill	0.324888	0.586549	0.851836	1.295044	-5.382508	-9.835011	-13.593751	-20.501414
BQMall	0.121855	0.435306	0.517021	1.214616	-1.846946	-5.350024	-7.410188	-16.517336
BlowingBubbles	0.113651	0.724375	0.366761	0.709948	-1.861762	-9.541168	-6.114249	-12.442009
RaceHorses	0.082749	0.266925	0.449689	0.736997	-1.178324	-4.478314	-6.536050	-10.373095
Johnny	0.252789	0.588829	0.943564	1.027556	-9.288810	-9.039213	-13.159810	-14.473593
KristenAndSara	0.272934	0.542210	0.683195	1.132011	-10.869215	-7.572422	-8.732636	-14.671629
vidyo1	0.261430	0.563385	0.600777	1.093749	-9.591982	-13.597687	-7.754503	-14.663874
vidyo4	0.198992	0.632756	0.982589	1.894641	-4.801789	-9.251774	-13.481379	-25.084586
average	**0.198647**	**0.518511**	**0.615197**	**1.104111**	**-5.190128**	**-8.563509**	**-9.092717**	**-16.423092**

redistribute them. The frequency is typically set to 4 [5], [11], so we updated the actual bits every 4 LCUs.

IV. EXPERIMENTAL RESULTS

We conducted a series of experiments to test the performance of the proposed method. As mentioned above, we mainly focused on the mutual deductive relation between QP and λ.

A. NUMERICAL DIFFERENCE

The set of test images comes from HEVC test videos. We selected one frame from each sequence as the image source as-realized on the HM 16.20 platform.

We assigned statistics to the data in a non-RC experiment at fixed QPs: 7, 12, 17, 22, 27, 32, 37, and 42. All parameters were consistent in this case except apart from measurement distortion. As shown in Fig. 3, we calculated the ratio of

TABLE 6. Objective distortion of different models.

Test Image	PSNR/dB (default)	PSNR/dB JND	PSNR/dB JR
PeopleOnStreet	41.0419	40.7528	40.5815
Traffic	40.9753	40.6177	40.4500
BQTerrace	36.6588	36.2585	35.7552
Cactus	38.5754	38.1759	37.1784
BasketballDrill	39.0369	38.7716	38.5414
BQMall	37.2268	36.8640	36.3878
BlowingBubbles	33.8913	33.2661	32.7296
RaceHorses	33.9748	33.5196	33.2037
Johnny	46.3562	45.7949	45.7756
KristenAndSara	44.2748	43.7709	43.5207
vidyo1	45.5890	44.9674	44.8104
vidyo4	46.5412	45.6534	45.8382
average	**40.3452**	**39.8677**	**39.5644**

TABLE 7. Test information for MOS.

Item	Information
Display	Dell E2414Ht × 3
Type and size	LCD 24”
Resolution	1920×1080
Number of subjective	15 (male: 11, and female: 4, not professional in image processing, age: 18 to 33)
Viewing distance	4.0 H (H: image high)

TABLE 8. Evaluation criterion for original and distortion image.

Quality	Impairment	Score values
Excellent	Imperceptible	5
Good	Perceptible, but not annoying	4
Fair	Slightly annoying	3
Poor	Annoying	2
Bad	Very annoying	1

SAD to JND distortion at the eight QPs. As per the polygonal line, the value of SAD distortion is much greater than that of JND distortion. Considering the maximum point, if the SAD distortion of most pixels is close to the JND threshold, the subjective distortion value must be significantly reduced. In addition, if the QP decreases, each pixel distortion value will also decrease. The effect of the JND model is reduced correspondingly, namely, the ratio of SAD to JND decreases. We finally considered a case of extreme distortion, where the QP is sufficiently large and almost any amount of pixel distortion is greater than the corresponding threshold. In this case, the SAD and JND distortion are nearly identical.

We were able to clearly recognize the divergence among different distortion measurement methods, which reveals whether the default λ suitably reflects the relationship between the new distortion and bits. We also found that the broken lines in these four images have almost the same change trends. Previous conclusions do not relate to the resolution or content of the given image.

TABLE 9. Evaluation criterion for two distortion images.

Evaluation	Score values
Much worse	-3
Worse	-2
Slightly worse	-1
the same	0
Slightly better	+1
Better	+2
Much better	+3

TABLE 10. MOS scores.

Test Image	BPPP	MOS Scores S1 [2]	S2 [3]	S3 [4]	S4 [5]	S5 [6]
BasketballDrill (832x480)	0.40	3.13	3.13	3.13	0.00	0.00
	0.30	3.07	3.07	3.07	0.00	0.00
	0.20	2.80	2.80	2.80	0.00	0.00
	0.10	1.87	2.07	2.07	1.07	1.07
BQMall (832x480)	0.50	5.00	5.00	5.00	0.00	-0.04
	0.35	4.00	4.00	3.50	-0.07	-0.53
	0.20	3.07	3.07	2.47	0.00	-1.00
	0.10	1.40	2.13	1.07	1.87	0.07
BlowingBubbles (416x240)	0.40	4.00	4.00	4.00	0.00	0.00
	0.30	3.93	3.93	3.93	0.00	0.00
	0.20	3.07	3.07	3.50	0.00	1.00
	0.10	2.13	2.13	3.07	0.00	0.00
RaceHorses (416x240)	0.40	4.00	4.00	4.00	0.00	0.00
	0.30	3.93	3.93	3.93	0.00	0.00
	0.20	3.07	3.07	3.07	0.00	0.07
	0.10	2.13	2.13	2.13	0.00	1.07
Johnny (1280x720)	0.15	4.13	4.13	4.00	0.00	0.00
	0.10	3.93	3.93	3.93	0.07	0.00
	0.05	2.93	2.93	3.13	-0.73	0.80
	0.03	1.00	1.00	2.07	0.20	1.07
KristenAndSara (1280x720)	0.20	4.00	4.00	4.00	0.00	0.00
	0.10	3.47	3.47	3.47	0.27	0.00
	0.05	1.47	1.67	2.93	0.87	1.20
	0.03	1.00	1.00	2.00	0.87	1.43
vidyo1 (1280x720)	0.20	4.00	4.00	4.00	0.00	0.00
	0.10	3.50	3.50	3.50	0.00	0.00
	0.05	2.53	2.53	2.80	0.07	1.93
	0.03	1.00	1.00	1.87	0.13	1.07
vidyo4 (1280x720)	0.20	4.93	4.93	4.93	0.00	0.00
	0.10	4.00	4.00	4.00	0.00	0.00
	0.05	2.47	2.47	3.47	1.00	1.80
	0.03	1.00	1.13	2.07	1.00	1.40

[2]The score of SAD distorted image compared with original image. [3]The score of JND distorted image compared with original image. [4]The score of JR distorted image compared with original image. [5]The score of JND distorted image compared with SAD distorted image. [6]The score of JR distorted image compared with SAD distorted image.

B. PERCEPTUAL MODEL PERFORMANCE IN NON-RC

The goal of quoting JND and JND+ROI models to describe distortion serves to minimize human visual redundancy. These models are also more accurate than SAD for RDO. When the objective quality of the image does not change, the distortion obtained by different evaluation methods may widely differ (Section IV-A). In order to cope with the changes caused by perceptual distortion, we used two methods we proposed to obtain subjective λ, and a lot of necessary

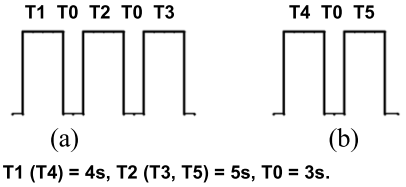


FIGURE 6. Time sequence.

tests have been done to evaluate the performance of the proposed λ. Table 1 lists all the tests we have done.

As discussed in Section III-A (2), we secured a perceptual λ by weighting the JND threshold via a lookup table. The test images used to build the lookup table are shown in Fig.4 (a). As to enhancing the credibility of our method, we selected different images for the subjective tests as shown in Fig.4 (b). About all the tests, we selected five fixed QPs (Fig.3), 22, 27, 32, 37, and 42 to verify the necessity of a new λ.

In the case of BD-BR and BD-PCPSNR, our goal was to minimize the bitrate and subjective distortion while maximizing the PCPSNR. Here we define that,

$$PCPSNR = 10\log\frac{(2^n-1)^2\cdot MN}{\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} d^2(x,y)}, \tag{27}$$

where M × N is the spatial resolution of the image, n is the bit depth of the pixel, and d (x, y) has the same denotation as in Eq. (19).

Next, the performance of proposed λ in different models will be described in detail.

The JND model: Table 2 shows the performance comparison among the default, lookup table, and pre-processing λ techniques. The default λ only contributes slightly to performance improvement, which justifies the need to secure a better λ. The pre-processing method has the best results: its BD-PCPSNR increased 0.42 dB and BD-BR decreased 6.90 %. If we adopt the lookup table method to obtain the λ, the BD-PCPSNR will increase 0.29 dB and BD-BR will decrease 4.90 %. This method does appear to remit some performance improvement with little time consumption. Comparing the two proposed approaches, we can make a reasonable choice according to the application conditions.

The ROI model: We also try to use the perceptual λ in the ROI model, but the data in Table 3 shows that there is no performance difference between default and pre-processing λ, and the default λ is the best choice.

The JND+ROI model: The data in Table 3 also show that, the default λ can also improve the subjective image quality in the JND+ROI model. It makes the BD-PCPSNR and BD-BR reach 0.68 dB and −10.83 %, respectively. However, compared with the default λ, the recommended λ can further improve the performance of a JND+ROI model especially the pre-processing method, PCPSNR promotion and rate savings reaches 0.83 dB and 12.96 %, respectively. This told us that the proposed λ is closer to the Lagrange operator in perceptual RDO.

According to the above analysis and data, we can conclude that: It is effective to improve the subjective image quality by using perceptual models in the pixel domain to measure distortion. In a JND model, we must consider a new λ, while in a ROI model, the default λ maybe the best. For a JND+ROI model, a proposed λ will perform better than the default one. In addition, we can also realize that a combination performs much better than employing a sole JND or ROI model.

In a non-RC case, both the bits and distortion are uncertain variables. The differential mean opinion score (DMOS) only describes the distortion (PSNR), so it cannot be used as a criterion for performance evaluation with a fixed QP.

The lookup table replaces a great deal of workload and only adds an additional time of 6 % (or less) on average. The pre-processing stage, conversely, is fairly time-consuming, and the total additional time may even exceed 30 %.

C. PERCEPTUAL MODEL PERFORMANCE IN RC

We also tested the performance of the proposed method when executing the RC. Five discrete values were selected which represent the average bits pre-allocated per pixel (BPPP): 0.2, 0.4, 0.6, 0.8 and 1.0. More factors must be assessed here than under non-RC condition.

Adopting the conclusion of the previous section, we only attempted to verify the recommended method in a JND and a JND+ROI model.

The restricted bit value is a decisive condition in the RC mode. It is crucial to guarantee accurate bit coding. Table 4 provides us with two pieces of information: 1) Comparison of the proposed method against the conventional method in JND and JND+ROI models. The proposed approach ensures similar even more excellent bit allocation accuracy as the traditional method. 2) If we use the default λ and QP-λ relationship in the perceptual model, the accuracy of bit allocation is seriously reduced. The data from Table 11 to Table 15 in the appendix provide us more detailed bit information.

Table 5 and Table 6 show the performance in JND and JND+ROI models with our proposed method in RC. In the JND model, The BD-PCPSNR is 0.51 dB and the BD-BR reaches −8.56 %, but the PSNR degraded from 40.35 dB to 39.87 dB. And in the JND+ROI model, these factors are further improved. The BD-PCPSNR is 1.10 dB and the BD-BR reaches −16.42 %, the PSNR also degraded from 40.35 dB to 39.56 dB. However, Table 5, Table 6 and Fig.5 areenough to show that compared with the default SAD model, the perceptual evaluation model yields better subjective image quality. This fully shows that the method we proposed to evaluate the distortion in the pixel domain can effectively reduce human visual redundancy, and accurately reflect perceptual image quality. At the same time, we see that if the default λ and QP-λ relationship are used, the perceptual quality of the image will also be improved, BD-PCPSNR and BD-BR reach 0.20 dB and −5.19 % in a JND model, and they reach 0.62 dB and

TABLE 11. Accuracy of bit allocation for a SAD model by default lambda.

Test Image	Bits Cost/bit						Bits Errors/%				
	BPPP	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
PeopleOnStreet		822976	1640648	2398232	3275224	3716352	0.4609	0.1372	-2.4157	-0.0481	-9.2688
Traffic		825232	1646488	2466992	3285656	4104704	0.7363	0.4937	0.3822	0.2703	0.2125
BQTerrace		418280	833456	1247728	1662512	2077440	0.8584	0.4842	0.2868	0.2189	0.1852
Cactus		417976	832408	1247776	1662232	2079000	0.7851	0.3578	0.2906	0.2021	0.2604
BasketballDrill		81776	162040	241920	322232	401800	2.3838	1.4373	0.9615	0.8589	0.6110
BQMall	BPP	81608	161712	241040	320824	400280	2.1735	1.2320	0.5943	0.4182	0.2304
BlowingBubbles	(Actual)	20912	41792	64008	83624	101968	4.7276	4.6474	6.8510	4.6975	2.1314
RaceHorses		20656	40880	61296	81504	101496	3.4455	2.3638	2.3237	2.0433	1.6587
Johnny		183848	349728	546184	738400	921568	-0.2561	-5.1302	-1.2254	0.1519	-0.0035
KristenAndSara		185648	352888	501320	657464	862480	0.7205	-4.2730	-9.3388	-10.8257	-6.4149
vidyo1		185544	359584	539272	732128	899480	0.6640	-2.4566	-2.4754	-0.6988	-2.4002
vidyo4		185856	369824	554112	738952	923296	0.8333	0.3211	0.2083	0.2268	0.1840
Average of ABS							**1.5038**	**1.9445**	**2.2795**	**1.7217**	**1.9634**
Average(BPP)		**0.2029**	**0.3999**	**0.5982**	**0.7983**	**0.9895**	**Average of Total ABS = 1.8826**				

TABLE 12. Accuracy of bit allocation for a JND model by default lambda.

Test Image	Bits Cost/bit						Bits Errors/%				
	BPPP	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
PeopleOnStreet		820616	1532112	2180560	3056064	3440352	0.1729	-6.4873	-11.2728	-6.7363	-16.0070
Traffic		824032	1647544	2467224	3286096	4098936	0.5898	0.5581	0.3916	0.2837	0.0717
BQTerrace		418592	833248	1247120	1660824	2075456	0.9336	0.4591	0.2379	0.1172	0.0895
Cactus		417960	832648	1246384	1660288	2076376	0.7813	0.3868	0.1788	0.0849	0.1339
BasketballDrill		75624	154444	240176	321840	401168	-5.3185	-3.3178	0.2337	0.7362	0.4527
BQMall	BPP	81536	161480	240848	320136	398832	2.0833	1.0867	0.5142	0.2028	-0.1322
BlowingBubbles	(Actual)	20792	41160	61848	81648	101160	4.1266	3.0649	3.2452	2.2236	1.3221
RaceHorses		20792	40848	61144	81048	101192	4.1266	2.2837	2.0700	1.4724	1.3542
Johnny		169752	322800	477264	736488	900736	-7.9036	-12.4349	-13.6892	-0.1074	-2.2639
KristenAndSara		184920	324656	450368	577728	775000	0.3255	-11.9314	-18.5532	-21.6406	-15.9071
vidyo1		180816	320040	482472	713088	840360	-1.9010	-13.1836	-12.7474	-3.2813	-8.8151
vidyo4		185888	346088	513456	738968	879624	0.8507	-6.1176	-7.1441	0.2289	-4.5547
Average of ABS							**2.4261**	**5.1093**	**5.8565**	**3.0929**	**4.2587**
Average(BPP)		**0.1998**	**0.3848**	**0.5717**	**0.7834**	**0.9631**	**Average of Total ABS = 4.1487**				

TABLE 13. Accuracy of bit allocation for a JND+ROI model by default lambda.

Test Image	Bits Cost/bit						Bits Errors/%				
	BPPP	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
PeopleOnStreet		817896	1555312	2205296	3080224	3462384	-0.1592	-5.0713	-10.2663	-5.9990	-15.4691
Traffic		824992	1647104	2467616	3287808	4095432	0.7070	0.5313	0.4076	0.3359	-0.0139
BQTerrace		419320	833648	1245944	1659592	2074976	1.1092	0.5073	0.1434	0.0429	0.0664
Cactus		417968	832640	1246288	1660632	2077040	0.7832	0.3858	0.1710	0.1056	0.1659
BasketballDrill		78496	156648	242240	321992	401752	-1.7228	-1.9381	1.0951	0.7838	0.5990
BQMall	BPP	81672	161392	240888	320008	398544	2.2536	1.0317	0.5308	0.1628	-0.2043
BlowingBubbles	(Actual)	20872	41432	61960	81448	100992	4.5272	3.7460	3.4322	1.9732	1.1538
RaceHorses		20792	40984	61464	81136	101200	4.1266	2.6242	2.6042	1.5825	1.3622
Johnny		179640	328264	484304	736632	906664	-2.5391	-10.9527	-12.4161	-0.0879	-1.6207
KristenAndSara		184888	330112	457216	587928	781896	0.3082	-10.4514	-17.3148	-20.2572	-15.1589
vidyo1		183968	326416	487024	713296	839328	-0.1910	-11.4540	-11.9242	-3.2530	-8.9271
vidyo4		186104	351976	516192	739176	881360	0.9679	-4.5204	-6.6493	0.2572	-4.3663
Average of ABS							**1.6163**	**4.4345**	**5.5796**	**2.9034**	**4.0923**
Average(BPP)		**0.2017**	**0.3881**	**0.5749**	**0.7838**	**0.9647**	**Average of Total ABS = 3.7252**				

TABLE 14. Accuracy of bit allocation for a JND model by processing lambda.

Test Image	Bits Cost/bit						Bits Errors/%				
	BPPP	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
PeopleOnStreet		813616	1610872	2436616	3282768	4101680	-6.8164	-1.6801	-0.8538	0.1821	0.1387
Traffic		825152	1648808	2471512	3291560	4113096	0.7266	0.6353	0.5660	0.4504	0.4174
BQTerrace		417504	833088	1249520	1665928	2083176	0.6713	0.4398	0.4308	0.4249	0.4618
Cactus		418488	838664	1278688	1733944	2095448	0.9086	1.1121	2.7752	4.5250	1.0536
BasketballDrill		81800	163112	242848	322688	401864	2.4139	2.1084	1.3488	1.0016	0.6270
BQMall	BPP	81776	161144	240824	320656	400120	2.3838	0.8764	0.5041	0.3656	0.1903
BlowingBubbles	(Actual)	21280	41888	61672	82944	103784	6.5705	4.8878	2.9514	3.8462	3.9503
RaceHorses		20848	40856	61432	81296	101648	4.4071	2.3037	2.5507	1.7829	1.8109
Johnny		180088	369784	555720	782520	935992	-2.2960	0.3103	0.4991	6.1361	1.5616
KristenAndSara		181952	367440	555120	739096	926968	-1.2847	-0.3255	0.3906	0.2463	0.5825
vidyo1		183128	370776	556696	745392	925200	-0.6467	0.5794	0.6756	1.1003	0.3906
vidyo4		185696	372800	558032	790616	976888	0.7465	1.1285	0.9172	7.2342	5.9991
Average of ABS							**2.4893**	**1.3656**	**1.2053**	**2.2746**	**1.4320**
Average(BPP)		**0.2023**	**0.4041**	**0.6064**	**0.8182**	**1.0143**	**Average of Total ABS = 1.7534**				

TABLE 15. Accuracy of bit allocation for a JND+ROI model by processing lambda.

Test Image	Bits Cost/bit						Bits Errors/%				
	BPPP	0.2	0.4	0.6	0.8	1.0	0.2	0.4	0.6	0.8	1.0
PeopleOnStreet		818232	1635592	2448848	3264240	4076824	-0.1182	-0.1714	-0.3561	-0.3833	-0.4682
Traffic		824616	1655064	2458920	3277640	4092640	0.6611	1.0171	0.0537	0.0256	-0.0820
BQTerrace		416656	830320	1242016	1656448	2070864	0.4668	0.1061	-0.1723	-0.1466	-0.1319
Cactus		417192	834480	1268160	1723752	2150224	0.5961	0.6076	1.9290	3.9106	3.6952
BasketballDrill		81360	162248	241728	321744	400568	1.8630	1.5675	0.8814	0.7061	0.3025
BQMall	BPP	81872	161568	240880	319328	398528	2.5040	1.1418	0.5275	-0.0501	-0.2083
BlowingBubbles	(Actual)	20992	41608	60848	81928	103248	5.1282	4.1867	1.5759	2.5741	3.4135
RaceHorses		21152	40720	60688	80792	100608	5.9295	1.9631	1.3088	1.1518	0.7692
Johnny		185040	369064	552616	746176	924952	0.3906	0.1150	-0.0622	1.2066	0.3637
KristenAndSara		185816	368664	552232	735568	919208	0.8116	0.0065	-0.1317	-0.2322	-0.2595
vidyo1		185160	370048	552728	736672	915752	0.4557	0.3819	-0.0420	-0.0825	-0.6345
vidyo4		185920	373976	554960	781040	941152	0.8681	1.4475	0.3617	5.9353	2.1215
Average of ABS							**1.6494**	**1.0594**	**0.6169**	**1.3671**	**1.0375**
Average(BPP)		**0.2033**	**0.4041**	**0.6029**	**0.8097**	**1.0074**	**Average of Total ABS = 1.1460**				

−9.09 % in a JND+ROI model. But, the proposed way has better performance, obviously.

We conducted subjective tests by means of the single stimulus continuous quality score (SSCQS) based on the adjectival categorical judgment (ACJ) [40] to determine the mean opinion score (MOS). The test conditions and comparison scale are shown in Table 7, Table 8 and Table 9. We ran two MOS experiments to verify the effectiveness of the JND and JND+ROI models. For the first, we displayed reference images (left) and test images (right) at the same time according to the time sequence shown in Fig. 6(a).

At T1, the original image is on the left and the SAD and JND (JND+ROI) images alternately displayed 3 times per second are on the right. The goal in this case was to recognize the difference between two distorted images before evaluating them. At T2, we evaluated the original and one of the two distorted images randomly. At T3, the other distorted image was handled in the same manner. A mid-gray image was observed at T0.

As shown in Table 10, although there were differences between the two distorted images in a JND model, the dissimilarity is not well reflected by comparing the distorted versus original image separately. We instead observed the reference images (left) and test images (right) at the same time according to the time sequence shown in Fig. 6(b). T4 is similar to T1, so we compared the two distorted images at T5. As shown in Table 10, when BPPP is sufficiently large, there is little difference visible between the two distorted images despite an improvement in PCPSNR. When the BPPP is sufficiently small, the JND model much more effectively improves the perceptual image quality than SAD model. Without a doubt, the JND+ROI model has the best perceptual performance among them as shown in Table 10.

TABLE 16. The values of alpha and beta of different contents and resolutions in a JND model for one frame.

BPPP	1.0		0.8		0.6		0.4		0.2	
	Alpha	Beta	Alpha	Beta	Alpha	Beta	Alpha	Beta	Alpha	Beta
BasketballDrive	9.78838	1.22515	7.75062	1.12701	6.51278	1.05591	5.02048	0.93917	4.71446	0.87732
BQTerrace	5.02728	1.14873	4.71167	1.01966	4.56745	0.95426	4.26718	0.89632	3.39511	0.90604
Cactus	4.65270	1.12453	4.33572	1.08939	4.07898	1.05828	3.88619	1.02916	3.58575	0.95084
Kimono1	6.29901	1.11881	5.13346	1.04237	3.99367	0.94423	3.57013	0.88980	3.75185	1.09674
ParkScene	5.10105	1.14573	4.94427	1.16099	4.92165	1.20984	4.48112	1.16910	3.74294	1.06983
BasketballDrill	4.09464	1.02030	3.84530	0.94794	3.74439	0.90583	3.41137	0.93826	2.97224	0.90901
BQMall	4.57406	1.11860	4.29682	1.05231	4.34321	0.94548	4.07635	0.90052	3.66373	0.88327
PartyScene	3.61420	1.00111	3.73205	0.91623	3.56648	0.91033	3.05413	0.92305	2.55825	0.92322
RaceHorses	3.81444	1.16519	3.59964	1.07693	3.44906	1.09693	3.45887	0.97427	3.16565	0.91771
BasketballPass	4.25721	1.15764	3.84743	1.04214	3.68863	0.96620	3.57800	0.94223	2.78084	0.96501
BlowingBubbles	3.88298	0.92754	3.77277	0.94188	3.42979	0.97732	3.02344	0.93658	2.61905	0.92318
BQSquare	5.44250	0.79102	4.89842	0.81854	4.33868	0.81957	3.03930	0.89323	2.36034	0.92349
RaceHorses	3.55547	0.98472	3.40029	1.02612	3.83456	0.87252	3.74734	0.86299	2.79242	0.89751
FourPeople	9.15153	1.25741	7.76375	1.20137	6.43240	1.15580	5.30133	1.04537	5.23941	0.84513
Johnny	18.8154	1.40963	18.0230	1.39304	11.0218	1.24211	6.80400	1.05101	5.60896	0.94979
KristenAndSara	11.6893	1.32618	9.11730	1.24415	6.98818	1.14741	5.97485	1.10763	4.69756	0.97298
vidyo1	12.6019	1.34603	12.2006	1.32536	8.94147	1.22484	7.37893	1.17246	5.42659	0.95647
vidyo3	16.1127	1.37330	12.4735	1.27895	8.85070	1.15322	6.24420	0.98420	5.28868	0.84629
vidyo4	23.8359	1.47179	19.0080	1.41207	11.6796	1.26614	7.83740	1.11659	6.20351	0.92427

The time consumption is the same as the pre-processing method in non-RC.

V. CONCLUSION

In this study, the ROI and JND models are combined to evaluate perceptual image quality in the pixel domain for HEVC intra coding. For the successful implementation of the scheme, we established a new method for obtaining QP and λ values in perceptual models under the non-RC or RC conditions. We also proposed an approach to allocate bits according to the subjective distortion. The results show that the subjective PSNR improvement can be gained in a JND, a ROI, and a ROI+JND model under both non-RC and RC conditions. At the same time, it is easy to find that the combination performs much better than employing a sole JND or ROI model.

Still, this study has several limitations. One hand: 1) The limited sample set and large sampling step, for example, may have caused substantial error in the lookup table. 2) The pre-compression method is also overly time-consumptive due to the costly pre-compression stage. On the other hand: Pay more attention to the current image processing methods, and deep learning can better complete a lot of work. In HEVC, scholars used deep learning method to achieve fast intra mode and CU partitioning, [44] not only improves BD-BR, but also greatly reduces coding time. Meanwhile, Li *et al.* [45] used deep learning method to excavate the continuous multi-frame image information in HEVC and applied it to the design of loop filter, so as to significantly improve the filtered image quality and further improve the efficiency of HEVC coding.

So in the future work, 1) we plan to further improve the performance of perceptual coding by improving the ROI and JND models based on our proposed approach. 2) We hope to use deep learning method to obtain perceptual lambda in a subjective model to solve the accuracy and time consumption problems. 3) As per the supra-threshold level concept, the image may indeed have a more precise JND scale and/or multiple scales [39] that can further minimize human visual redundancy. 4) We will also further study HEVC and continue to strive for more innovative approaches.

APPENDIX

Table 11 to Table 15 present detailed bit information for all the test images in different BPPP.

Table 16 presents the values of alpha and beta with different contents and resolutions for one frame.

REFERENCES

[1] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, 1991.
[2] D. Taubman and M. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards and Practice*, vol. 642. Berlin, Germany: Springer, 2012.
[3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
[5] B. Li, H. Li, L. Li, and J. Zhang, "λ domain based rate control for high efficiency video coding," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841–3854, Sep. 2014.
[6] F. Ernawan and S. H. Nugraini, "The optimal quantization matrices for JPEG image compression from psychovisual threshold," *J. Theor. Appl. Inf. Technol.*, vol. 70, no. 3, pp. 566–572, Dec. 2014.
[7] S.-H. Bae and M. Kim, "A novel DCT-based JND model for luminance adaptation effect in DCT frequency," *IEEE Signal Process. Lett.*, vol. 20, no. 9, pp. 893–896, Sep. 2013.
[8] X. Zhang, R. Xiong, W. Lin, S. Ma, J. Liu, and W. Gao, "Video compression artifact reduction via spatio-temporal multi-hypothesis prediction," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6048–6061, Dec. 2015.
[9] J. Kim, S.-H. Bae, and M. Kim, "An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1786–1799, Nov. 2015.
[10] E. P. Simoncelli, *Foundations of Vision*. Sunderland, MA, USA: Sinauer, 1996.

[11] S. Li, M. Xu, Y. Ren, and Z. Wang, "Closed-form optimization on saliency-guided image compression for HEVC-MSP," *IEEE Trans. Multimedia*, vol. 20, no. 1, pp. 155–170, Jan. 2018.
[12] J. Wu, L. Li, W. Dong, G. Shi, W. Lin, and C.-C. J. Kuo, "Enhanced just noticeable difference model for images with pattern complexity," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2682–2693, Jun. 2017.
[13] X. F. Zhang, S. Wang, K. Gu, W. Lin, S. Ma, and W. Gao, "Just-noticeable difference-based perceptual optimization for JPEG compression," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 96–100, Jan. 2017.
[14] Z. Wei and K. N. Ngan, "Spatial just noticeable distortion profile for image in DCT Domain," in *Proc. IEEE Int. Conf. Multimedia Expo*, Hannover, Germany, Jun./Apr. 2008, pp. 925–928.
[15] S.-H. Bae, J. Kim, and M. Kim, "HEVC-based perceptually adaptive video coding using a DCT-based local distortion detection probability model," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3343–3357, Jul. 2016.
[16] X. Zhang, R. Xiong, S. Ma, and W. Gao, "Reducing blocking artifacts in compressed images via transform-domain non-local coefficients estimation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 836–841.
[17] X. Zhang, R. Xiong, X. Fan, S. Ma, and W. Gao, "Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4613–4626, Dec. 2013.
[18] S. Ki, S.-H. Bae, M. Kim, and H. Ko, "Learning-based just-noticeable-quantization-distortion modeling for perceptual video coding," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3178–3193, Jul. 2018.
[19] H. Wang, L. Wang, Q. Tu, and A. Men, "Perceptual video coding based on saliency and just noticeable distortion for H.265/HEVC," in *Proc. Int. Symp. Wireless Pers. Multimedia Commun.*, Jan. 2015.
[20] Z. Zeng, H. Zeng, J. Chen, J. Hou, C. Cai, and K.-K. Ma, "A novel direction-based JND model for perceptual HEVC intra coding," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Xiamen, China, Nov. 2017.
[21] K.-C. Liu, "An improvement of just noticeable color difference estimation," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, Gdansk, Poland, Sep. 2016.
[22] M. Uzair and R. D. Dony, "Estimating just-noticeable distortion for images/videos in pixel domain," *IET Image Process.*, vol. 11, no. 8, pp. 559–567, Aug. 2017.
[23] C. Yuan, X. Sun, and R. Lv, "Fingerprint liveness detection based on multi-scale LPQ and PCA," *China Commun.*, vol. 13, no. 7, pp. 60–65, Jul. 2016.
[24] J. Zhang and S. Sclaroff, "Exploiting surroundedness for saliency detection: A Boolean map approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 889–902, May 2016.
[25] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.
[26] Y. Pitrey, M. Babel, and O. Deforges, "One-pass bitrate control for MPEG-4 scalable video coding using ρ-domain," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, May 2009, pp. 1–5.
[27] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2012.
[28] W. Ding and B. Liu, "Rate control of MPEG video coding and recording by rate-quantization modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 1, pp. 12–20, Feb. 1996.
[29] J. Ribas-Corbera and S. Lei, "Rate control in DCT video coding for low-delay communications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 172–185, Feb. 1999.
[30] S. Mallat and F. Falzon, "Analysis of low bit rate image transform coding," *IEEE Trans. Image Process.*, vol. 46, no. 4, pp. 1024–1042, Apr. 1998.
[31] M. Dai, D. Loguinov, and H. Radha, "Rate-distortion modeling of scalable video coders," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 2, Oct. 2004, pp. 1093–1096.
[32] B. Li, D. Zhang, H. Li, and J. Xu, *QP Determination by λ Value*, document Rec. JCVVC-I0426, Geneva, Switzerland, Apr./May 2012.
[33] M. R. Ardestani, A. A. B. Shirazi, and M. R. Hashemi, "Rate-distortion modeling for scalable video coding," in *Proc. IEEE 17th Int. Conf. Telecommun. (ICT)*, Apr. 2010, pp. 923–928.
[34] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.
[35] J. Li and H.-H. Sun, "On interactive browsing of large images," *IEEE Trans. Multimedia*, vol. 5, no. 4, pp. 581–590, Dec. 2003.
[36] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image Vis. Comput.*, vol. 29, no. 1, pp. 1–14, Jan. 2011.
[37] M. Xu, Y. Ren, and Z. Wang, "Learning to predict saliency on face images," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3907–3915.
[38] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185–1198, May 2011.
[39] H. R. Wu and D. M. Tan, "Subjective and objective picture assessment at supra-threshold levels," in *Proc. PCS*, 2015, pp. 312–316.
[40] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT.500-13, Jan. 2012.
[41] H. R. Tavakoli, A. Borji, J. Laaksonen, and E. Rahtu, "Exploiting inter-image similarity and ensemble of extreme learners for fixation prediction using deep features," *Neurocomputing*, Mar. 2017, pp. 10–18.
[42] J. Sainio, A. Ylä-Outinen, M. Viitanen, J. Vanne, and T. D. Hämäläinen, "Eye-controlled region of interest HEVC encoding," in *Proc. IEEE ISM*, Dec. 2018, pp. 186–187.
[43] A. Stergiou, G. Kapidis, G. Kalliatakis, C. Chrysoulas, R. Veltkamp, and R. Poppe, "Saliency tubes: Visual explanations for spatio-temporal convolutions," 2019, *arXiv:1902.01078*. [Online]. Available: https://arxiv.org/abs/1902.01078
[44] W. Kuang, Y. Chan, S. H. Tsang, and W. C. Siu, "Online-learning-based Bayesian decision rule for fast intra mode and CU partitioning algorithm in HEVC screen content coding," *IEEE Trans. Image Process.*, vol. 29, pp. 170–185, 2020.
[45] T. Li, M. Xu, C. Zhu, R. Yang, Z. Wang, and Z. Guan, "A deep learning approach for multi-frame in-loop filter of HEVC," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5663–5678, Nov. 2019.

DONGHUI FU received the B.S. degree in electronic information engineering from Jilin University, Changchun, China, in 2015. He is currently pursuing the Ph.D. degree with the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun. His research interest includes image compression.

YANJIE WANG received the B.S. degree in computer science and applications from the Jilin University of Technology, Changchun, China, in 1988, and the M.S. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, in 1998. He is a Doctoral Tutor and a Scientific Researcher. His research interest includes real-time image processing.

BO FAN received the B.S. degree in information engineering from Jilin University, Changchun, China, in 2011, and the M.S. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, in 2014, where he is currently pursuing the Ph.D. degree. His research interests include high-image acquisition and real-time image processing.

NANNAN DING received the B.S. degree in information engineering from Jilin University, Changchun, China, in 2007, and the Ph.D. degree from the Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, in 2012. His research interests include image signal processing, image registration, and image stitching fusion.