IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 7, 2021, accepted September 1, 2021, date of publication September 3, 2021, date of current version September 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3110292

Fast CU Partition Decision Strategy Based on Human Visual System Perceptual Quality

JINCHAO ZHAO, TENGYAO CUI, AND QIUWEN ZHANG, (Member, IEEE)
College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China

Corresponding author: Qiuwen Zhang (zhangqwen@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771432 and Grant 61302118, and in part by the Basic Research Projects of Education Department of Henan under Grant 21zx003 and Grant 20A880004.

ABSTRACT A fast Coding Unit (CU) partition decision strategy based on Human Visual System (HVS) perception quality is proposed in this paper. Considering that it is difficult for existing fast algorithms to further improve compression efficiency, perceptual coding technology has been tried to remove visual redundancy for achieving the purpose of reducing the bit rate on the basis of maintaining subjective visual quality. However, the existing perceptual coding model is still insufficient to reflect the characteristics of the HVS, which has limited improvement in coding efficiency, especially in fast algorithms. In this method, the characteristics of the limited human capacity for spatial-temporal resolution and the visual sensory memory are used to improve coding performance. First, the color complexity is used as a control factor to optimize the Just Noticeable Difference (JND) model to remove visual redundancy in a way that is more in line with the characteristics of visual perception. Second, a classification model of motion patterns based on human visual saliency is designed to provide a basis for CU classification, which effectively improves the coding accuracy. Finally, an offline Decision Tree (DT) classifier is designed based on the above model, and texture features are incorporated into the classifier as another key attribute to further reduce the computational complexity. The results of performance evaluation confirm that the proposed method achieves significantly improved coding performance compared with original Versatile Test Model (VTM). Compared with existing algorithms, our method not only improves coding efficiency, but also improves subjective visual quality.

INDEX TERMS VVC, fast partition decision, just noticeable difference, motion pattern classification, texture information, decision tree.

I. INTRODUCTION

With the vigorous development of multimedia, video services with high visual quality and low transmission cost are urgently needed in many fields. In order to promote the development of video coding technology, the Joint Video Exploration Team (JVET) composed by the Video Coding Experts Group (VCEG) and the Moving Picture Experts Group (MPEG) released a new generation of video coding technology standard, H.266/Versatile Video Coding (VVC) [1], which provides a new platform for the development of video technology. A series of new coding tools, such as the Quad Tree with nested Multi Type Tree (QTMT) partition structure, are introduced into VVC [2]. The Coding Unit (CU) has obtained the asymmetry and directionality due to the introduction of QTMT. These technical improvements have brought nearly half of the compression gain to VVC with the expense of obviously increased computational complexity. The optimal partition is determined by recursively checking the Rate Distortion Optimization (RDO) cost of all partition modes, and this process takes about 95% of the coding time [3]. The VTM stipulates the impermissible partition method to avoid redundant division, as shown in Figure 1. Nevertheless, it is still challenging to significantly reducing the computational cost of finding the best CU partition structure. Therefore, it is necessary to find a coding scheme with low computational complexity and high efficiency.

Most of the existing methods are aimed at reducing the bit rate and maintaining the objective quality at the pixel-wise, while ignoring the value of subjective visual perception. In fact, as the final recipient of video content, the visual perception of HVS is not based the pixel-wise. In other

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

words, the subjective visual perception cannot be accurately reflected by the objective quality evaluation standard, such as Peak Signal-to-Noise Ratio (PSNR). Recently, the concept of optimizing coding quality based on visual perception has been discussed by some scholars, and the purpose is to remove visual redundancy that cannot be perceived pixel-wise. This idea is not only possible to significantly reduce the coding complexity, but also to maintain or even improve the quality of visual perception, which is a novel and valuable technical path. Perception-based methods to improve coding efficiency have been tried in [4] and [5]. However, some studies have pointed out that these schemes still do not fully reflect the characteristics of the HVS. Just Noticeable Difference (JND) is one of the cores of the Perceptual Video coding technology (PVC), which reflects the threshold of HVS to perceive the differences. The visual redundancy is converted into an accurate threshold by JND. Well known, regions with complex textures are easier to hide small texture changes, and texture changes in darker background luminance are more likely to be ignored. When the difference between pixels exceeds the threshold, it will be detected by HVS. if the distortion is controlled within the range that the HVS cannot detect, the bit rate will be effectively reduced without changing the subjective quality.

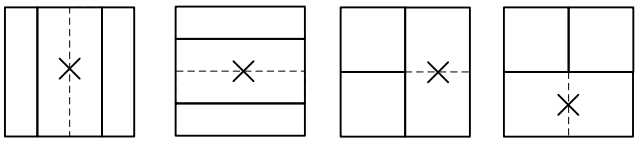

FIGURE 1. The case of redundant partition structure.

Another way to improve coding efficiency is to appropriately reduce the coding quality of non-Regions of Interest (ROI) based on visual saliency. Different from the perceived distortion value revealed by JND, the goal of the visual saliency model is to enhance the imaging quality of ROIs with the bit constraint. In existing studies, the motion area is considered to be more visually significant. The encoder allocates more bits to the region containing the moving object to achieve more accurate encoding. However, this solution does not take into account that the spatial-temporal resolution capacity and visual-memory capacity of the HVS are limited. In fact, when the speed of the moving target exceeds a certain level, the target areas will have an aliasing effect [6]. Although these areas are assigned higher visual saliency, they cannot make humans perceive better visual quality, even if the area is given more bandwidth. Therefore, the allocation method of visual saliency needs to be optimized to make it more in line with the perceptual characteristics of HVS.

Based on the above observations, we propose a fast CU partitioning algorithm based on visual perception quality for the complex computational complexity and high-quality coding requirements of VVC. First, we analyzed the principle and nature of CU division in VVC to find a more efficient partition strategy. Then, based on the visual characteristics of the HVS, a JND model that effectively removes visual redundancy is proposed. Third, based on the visual saliency characteristics of the HVS, a motion classification model is proposed to allocate bits reasonably. Finally, a fast CU partition decision method based on decision tree is proposed. Experimental results show that the proposed algorithm achieves a satisfactory compromise between coding efficiency and subjective visual quality. In this paper, the contributions can be summarized as follows:

1) The color features and local structure are not fully utilized by the existing JND model based pixel-wise. We proposed an optimization scheme for JND based on color complexity, which uses color complexity as a visual weight factor to improve the accuracy of the CM model and the overall scheme.

2) The difference between the visual saliency model and visual perception is analyzed, and a motion pattern model closer to HVS visual perception is proposed to optimize the allocation of bits in regions with different motion patterns.

3) As the first method of CU partitioning beads on perceptual quality in VVC, we combine the above two models with texture features as the attributes of DT. This method maintains a good subjective visual experience while accelerating CU partition.

This paper is organized as follows. In Section II, related work is analyzed. In Section III, the proposed method is introduced in detail. The experimental results of the proposed is discussed in Section IV, and conclusion is given in Section V.

II. RELATED WORK

In order to improve the coding efficiency, the existing research has done a lot of work on the two aspects of improving the quality and controlling the bit rate, many scholars have made contributions to this. Yang *et al.* proposed a cascade decision architecture, in which directionality is used as a classification index to determine CU partition [7]. Zhang *et al.* used Gray-Level Co-occurrence Matrix to describe texture complexity to terminate CU partition early [8]. A similar approach was adopted in [9]. The difference between the above two is that the latter terminates the Multi-Tree partition with the calculated texture direction. Zhang *et al.* proposed a fast decision method that incorporates neighboring block information and local features as a reference. The Bayesian classification decision and the early termination algorithm based on the conditional random fields are executed in two stages, which has achieved considerable efficiency gains [10]. An intelligent classifier based on Convolutional Neural Networks (CNN) is adopted in [11], [12]. Grellert *et al.* proposed a Support Vector Machine model to disassemble the complex CU partition problem into a combination of multiple binary classification problems [13]. Yang *et al.* proposed a DT dual classifier framework. One classifier is used to determine whether a block is partitioned, and the other is used to distinguish between natural content and screen content. The CU partitioning and intra-frame mode, which have non-promising, are skipped in this frame [14]. The above method

is based on objective indices. However, trying to better coding efficiency has been more difficult achieved by improving these methods [15].

A. IMPROVE CODING EFFECIENCY BY JND

In view of the fact that the coding ability of traditional algorithms has almost reached the limit, perceptual visual coding has begun to be studied and high hopes are placed. To explore perceptual coding, the first problem that needs to be solved is to establish the mapping relationship between objective indicators and subjective perception, that is, how to find an evaluation indicator that can describe the real visual perception in an extremely close way. Ban *et al.* analyzed the distortion difference between PSNR and structural similarity index SSIM, which is more suitable for the perception characteristics of HVS, is used as an evaluation index, and an R-D cost model based on SSIM is proposed in [16]. Zhou *et al.* proposed a description of the relationship between visual perception and bits based on the divisive normalization method, which further optimized R-D control and bit allocation [17]. These methods reflect real perception by combining with traditional objective indicators such as PSNR, SSIM and Mean Square Error. Compared with the above indicators, JND has unique advantages in reflecting visual perception.

JND is to obtain the minimum visual threshold by simulating the perception characteristics of HVS, which reflects the difference in visual sensitivity to different features under physiological and psychological factors [18]. To further improve the coding performance, JND can be embedded in the fast algorithm to further compress the bit rate. Bae *et al.* proposed a joint JND model based on Discrete Cosine Transform (DCT) to deal with Temporal Masking (TM) and Foveated Masking (FM) effects [19]. A flexible discrete cosine variation kernel JND model based on probability summation theory is proposed in [20]. Ki *et al.* proposed a local distortion detection probability-based method and a DCT-based energy-reduced model, which can effectively estimate the transform coefficients of a fixed-size discrete cosine transform kernel [21]. However, all of the above transform-domain models do not take the quantization effects into consideration. Therefore, the JND suppression cannot be accurately performed when performing the quantization process. Compared with the transform-domain model, the pixel-wise model eliminates the process of transforming frequency domain, which means lower complexity [22]. Luminance adaption (LA) and CM are two fundamental factors of the pixel-wise model. The above two were not connected in the early model until a nonlinear additivity model for masking is proposed in [23]. On this basis, Yang *et al.* proposed to integrate LA and CM by the nonlinear additively masking model [24]. Liu *et al.* proposed a model based on edge contour decomposition in [25], where luminance, contrast and contour are evaluated separately. In [26] and [27], the local texture, color and motion are used as the influencing factors of the saliency model to achieve the purpose of local low-pass filtering. Zhou *et al.* proposed to embed the R-D model based on the pixel-wise as a weighting factor into the rate control framework to reduce the bit [28]. Zhang *et al.* proposed a fast CU size decision based on the JND model for 3D-HEVC, the Coding Tree Block (CTB) feature was analyzed to early skip the non-promising partition depth and intra prediction mode [29].

B. REASONABLY ALLOCATE BITS BY ROI

Another important issue for improving coding efficiency is how to allocate bits more reasonably. Guo *et al.* tried to allocate fewer bits for non-ROIs [30]. Kim *et al.* pointed out that the quality of the non-ROIs also plays a non-negligible effect on the overall visual perception quality. It is unreasonable to blindly reduce the bit of non-ROIs [31]. For this reason, a closed-form bit allocation approach was proposed to determine a reasonable ratio of non-ROIs to ROIs. References [32] and [33] configure the saliency model based on color attributes and luminance characteristics. Ma *et al.* used motion pattern and edge distortion as a reference for judging the ROIs, and allocated more bits to the region where distortion is more likely to be felt [34]. This approach limitedly improves the rationality of allocating bits. H. Oh *et al.* took the limited spatial-temporal resolution of humans into account and embedded the motion classification model based on the Hedge algorithm into the visual saliency model, which is more in line with the real perception of HVS [35].

III. PROPOSED METHOD

On the basis of Quad Tree (QT) type, the of the Multi Type Tree (MTT) partition structure is introduced. The QTMT structure includes five partition types, namely Binary Tree_Horizontal type (BT_H), Binary Tree Vertical type (BT_V), Ternary Tree_Horizontal type (TT_H), Ternary Tree_Vertical type (TT_V) and QT type. VVC not only realizes the transition from single partition to flexible partition, but also gives sub-CUs directivity and asymmetry [36]. In order to find the statistical law of CU partition, we calculated the distribution of CU partition mode. The test conditions are as follows: under the All-Intra configuration, six videos with different resolution are used as test sequences, each sequence is 200 frames, and the Quantization Parameters (QP) are set to 22, 27, 32, and 37, respectively. We define the CTU size as 128×128, the allowed minimum QT size and the minimum MTT size are 8×8 and 4×4, respectively. The allowed maximum MTT root node is 32×32 and the maximum MTT depth is 3. We define the QT depth and MTT depth of the CU as QTi and MTj, respectively, where $i \in \{0,1,2,3,4\}$ and $j \in \{0,1,2,3\}$. The CU depth distribution are shown in Table 1.

It is observed that QT1 partition type occupies the highest percentage, 19.6%. The QT2MT0 and QT2MT1 partition types also account for a large average of CUs, which are 16.7% and 14.1%, respectively. It can be seen that the size of the CUs is also affected by the resolution. Higher resolutions often encode with larger sizes, while lower resolutions are the opposite. In high-resolution sequences, the proportion of

TABLE 1. The CU depth distribution.

Sequences	QP	QT1	QT2				QT3				QT4		
			MT0	MT1	MT2	MT3	MT0	MT1	MT2	MT3	MT0	MT1	MT2
	22	13.4	16.8	20.7	15.4	13.4	4.9	5.1	4.9	3.3	1.1	0.6	0.3
CatRobot	27	20.1	20.5	21.3	13.1	10.7	3.9	3.8	2.6	2.1	1.2	0.4	0.2
3840×2160	32	26.1	23.2	20.3	13.5	9.1	3.4	2.1	1.2	0.6	0.3	0.1	0
	37	32.1	25.3	21.7	11.2	5.8	2.1	1.2	0.3	0.2	0.1	0	0
	22	35.6	33.1	20.8	6.7	2.2	1.2	0.2	0.1	0	0	0	0
FoodMarket	27	43.6	36.4	15.1	3.2	0.6	1	0	0	0	0	0	0
3840×2160	32	53.4	33.9	9.5	2.4	0.5	0.2	0.1	0	0	0	0	0
	37	62.6	31.1	5.2	0.6	0.4	0.1	0	0	0	0	0	0
	22	0.7	8.2	15.7	32.5	24.4	4.5	4.3	3.9	2.8	2.1	0.7	0.2
BasketballDrie	27	8.7	16.1	25.5	26.3	13.1	3.4	2.5	2.1	1.4	0.5	0.3	0.1
1920×1080	32	13.4	25.8	29.5	15.7	9.7	2.1	1.8	1.2	0.5	0.3	0	0
	37	35.2	22.4	19.8	12.1	6.7	1.6	1.2	0.6	0.3	0.1	0	0
	22	12.0	27.6	20.3	9.9	9.1	5.3	4.9	4.5	4.3	1.2	0.7	0.3
Johnny	27	32.4	20.3	14.1	9.2	8.9	4.2	3.5	3.4	2.7	0.6	0.4	0.2
1080×720	32	37.9	17.4	13.7	9.3	8.1	4.4	3.7	2.6	1.9	0.7	0.2	0.1
	37	41	18.6	13.7	9.8	8.3	2.9	2.4	1.3	0.6	0.4	0.1	0
	22	0	0.3	0.9	2.3	3.3	1.9	11	24.1	31.8	6.7	9.1	8.6
PartyScene	27	0	0.9	1.6	2.5	3.4	3.7	15.1	22.8	28.9	7.2	7.9	6.0
832×480	32	0.1	1.8	2.8	3.7	5.6	5.1	16	21.9	23.1	6.7	7.2	6.0
	37	0.2	2.1	5.6	7.7	14.9	7.5	16.7	18.4	14.7	6.1	4.0	2.1
	22	0	2.7	8.1	6.5	7.3	8.5	16.9	20.1	16.2	5.8	4.6	3.2
RaceHorses	27	0	5.3	9.4	5.7	8.5	11.9	16.3	16.1	15.6	5.1	3.5	2.6
416×240	32	0	4.9	9.6	12.1	13.5	12.8	15.1	12.1	10.6	5.1	3.0	1.2
	37	0	5.8	14.2	16.9	17.8	11.5	11.1	9.3	6.9	4.2	1.5	0.8
Average		19.6	16.7	14.1	10.3	8.6	4.5	6.5	7.2	7.0	2.3	1.8	1.3

CU is positively correlated with the size of CU. However, in low-resolution sequences, small-sized CUs are used more frequently to improve coding accuracy.

The intuitive statistical results are shown in Figure 2, the CU with large size (QT1, QT2MT0, QT2MT1, QT2MT2, QT2MT3) has reached 69.3%, while the CUs with small size (QT4MT0, QT4MT1, QT4MT2) have only 5.9%. This means that most of CU (69.3%) can be terminated early without affecting the coding quality.

Areas with flat textures are more likely to be encoded by QT and retain larger sizes. The distribution ratio of QT and QTMT as shown in Figure 3. The statistical results show that the CU divided by QT reaches 43.1%, which indicates that nearly half of the areas can ignore the recursive process of MTT. Considering the above results, the following strategies are proposed.

A. VISUAL PERCEPTION MODEL BASED ON HVS

1) JND MODEL BASED ON COLOR COMPLEXITY

In this section, on the basis of the classic pixel-wise JND model, the color complexity is used as a visual weighting factor to improve the CM model. The JND model is composed of CM based on color complexity and LA, which is defined as follows,

$$JND = L + C - \alpha \min\{L, C\} \tag{1}$$

where L and C are the threshold of LA and CM, respectively. α is used to solve the overlap effect between the two masking coefficients, and the value is 0.3. The calculation method of LA is consistent with [25].

When the traditional CM model performs texture analysis, the pixel value change of a single channel is merely taken into consideration, and it has limitations in reflecting area where the texture change is not significant. However, the visual stimulation of color is more direct for HVS. Compared with the texture, the color complexity is performed in the three-channel LAB color space, and the intensity of color change is reflected clearly. Greater color complexity reflects drastic color changes, which means that the area has a stronger ability to mask noise. In addition, color changes are often accompanied by texture changes, so more image

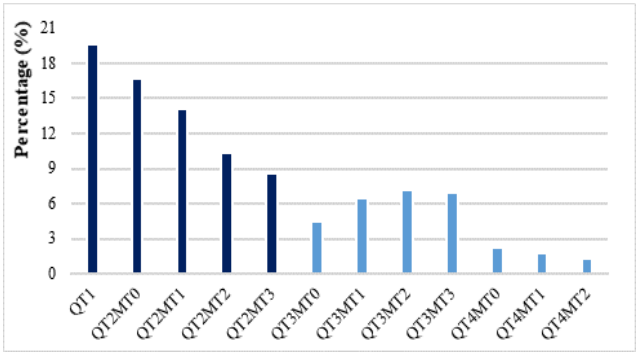

FIGURE 2. The statistical results of partition mode.

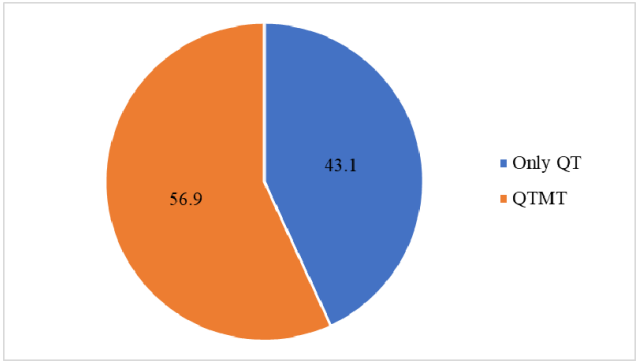

FIGURE 3. The distribution ratio of QT and QTMT.

information can be used for JND estimation to make up for the ineffectiveness of occlusion caused by blurred texture features [37]. We define the improved color complexity contrast masking model as follow,

$$C = \gamma \times \beta \times W_{(x,y)} \times I_{(x,y)} \tag{2}$$

where γ is the visual weighting factor of color complexity. β is the control factor, taking an empirical value of 0.12. $W_{(x,y)}$ is the edge weighting factor of the pixel (x, y), which is generated of the edge image extracted by the canny operator and a Gaussian low-pass filter. $I_{(x,y)}$ is the maximum weighted average gradient value, which are expressed as,

$$I_{(x,y)} = \max_{k=1,2,3,4} \left| \frac{1}{16} \sum_{j=1}^{5} \sum_{i=1}^{5} Y_{(x-3+i,y-3+j)} \times M_{i,j} \right| \tag{3}$$

where $Y_{(x,y)}$ is the pixel value of (x, y), $M_{i,j}$ is the high-pass filter. The distortion tolerance of the region is effectively reflected by the color complexity, but the color complexity is used as a CM factor to participate in the calculation of the JND threshold has not been proven feasible. To solve this problem, after a lot of experiments and corrections, the color complexity is used as a visual weighting factor to control the CM effect, which is defined as follows,

$$\gamma = \varepsilon \times \left(1 - \exp\left(\frac{-h_{(x,y)}}{\rho}\right)\right) + \mu \tag{4}$$

where ε, ρ, μ are control factors. $h_{(x,y)}$ is the color complexity of (x, y), which are calculated as follows,

$$h_{(x,y)} = \sum_{i,j \in \Omega} Ga \times \left(1 - \exp\left(-\frac{Eu(f_{(x,y)}, \bar{f}_{(x,y)})}{V}\right)\right) \tag{5}$$

where Ω is the 8 adjacent pixels surrounding the pixel (x, y), Ga is the Gaussian weight, Eu is the Euclidean distance. $f_{(x,y)}$ is the value of the pixel (x, y) in the LAB color space, and $\bar{f}_{(x,y)}$ is the average value obtained from pixel values of (x, y) and its neighboring pixel in the LAB color space. According to subjective quality evaluation, the value of V is 13, and the values of ε, ρ and μ are 1.3, 1.0 and 0.8, respectively.

2) A MODEL FOR CLASSIFICATION OF MOTION PATTERN

The proposed motion pattern classification model reflects the visual characteristics of HVS. The visual saliency of HVS is mainly affected by visual stimulation, spatial-temporal resolution capacity and visual-memory capacity. Moving objects are the main source of visual stimulation, and the central part of the picture serves as the center of the viewpoint, providing equally important visual stimulation. Visual memory forces the visual system to keep its attention to these areas and provides a basis for the human brain to predict the next movement in this area. The spatial-temporal resolution capacity reflects the range of the HVS to maintain clear and accurate vision. When the movement speed of object exceeds this range, temporal aliasing occurs, causing motion blurring. On the contrary, when the movement speed is within the range, improving the coding quality of this area will greatly improve the subjective visual experience. In addition, the human brain has the ability to adaptively adjust the visual focus based on the subconscious judgments made by the video content. Inspired by the above, if the redundancy of low visual saliency area is removed and the coding quality of the high visual saliency area is improved, the coding efficiency will be significantly improved.

Existing methods tend to assign more visual saliency to objects that move faster. However, the limited spatial-temporal resolution ability restricts the human visual perception of high-speed moving objects. The temporal-spatial resolution that can be recognized by human decreases as the speed of the moving object increases. We define the motion pattern as three types, namely the stationary pattern, the ordinary pattern and the aliased pattern. The different motion patterns are shown in Figure 4, the green frame in Figure 4 (a) and the blue frame in Figure 4 (b) represent the stationary pattern and the ordinary pattern, respectively. The aliased pattern is displayed in Figure 4 (c) red frame. Figure 5 shows the magnitudes corresponding to the different motion patterns in Figure 4.

According to the existing method, the player with higher movement speed in Figure 4 (c) red frame gets more visual saliency, and low-pass filtering is not performed on the area where the player is located. Therefore, this area has no positive effects on improving the quality of visual perception,

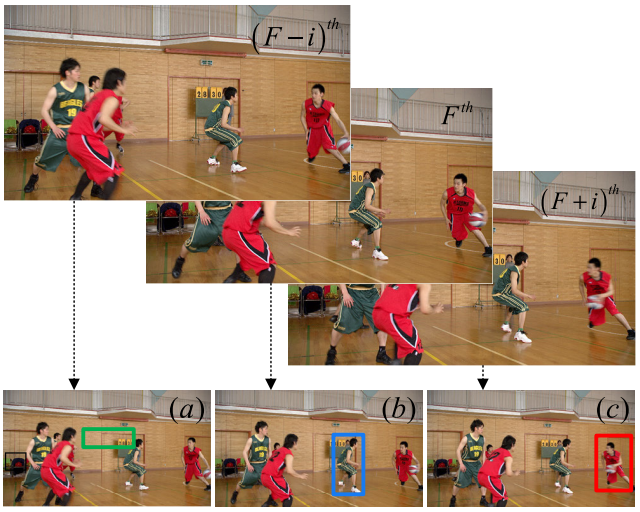

FIGURE 4. Examples of different motion patterns.

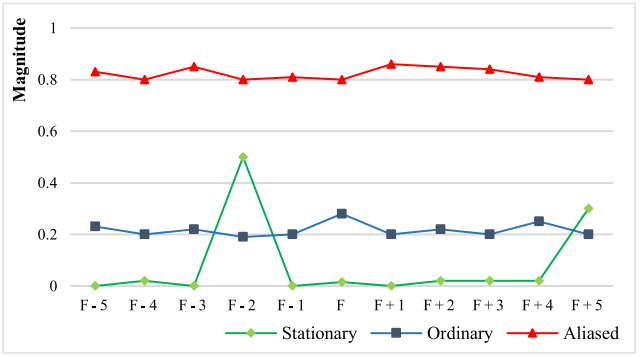

FIGURE 5. Magnitudes of different motion patterns.

even if it allocates more bitrates. Based on the above analysis, we proposed a motion pattern classification model based on HVS.

The proposed algorithm uses energy ratio and motion vector as classification conditions. We define the energy ratio of the current CU as follows,

$$P = \frac{\sum_{q\in U}\left(U_{(q)} - U_{(q+MV)}\right)^2}{\sum_{q\in U}\left(U_{(q)}\right)^2 - \frac{1}{W\times H}\left(\sum_{q\in U}\left(U_{(q)}\right)\right)^2} \tag{6}$$

where P is the energy ratio of Temporal Residual Energy to Texture Energy. $U_{(q)}$ is the current CU, q is the internal pixel of the current CU, MV is the motion vector. The width and height of the current CU are represented by W and H. According to the calculation result, the motion patterns are classified as follows,

$$\begin{cases} Stationary & MV = 0 \\ Ordinary & MV \neq 0,\ P \leq 1 \\ Aliased & MV \neq 0,\ P > 1 \end{cases} \tag{7}$$

According to the proposed method, since the movement speed of the player in Figure 4 (b) blue frame is within the perception threshold range of the HVS, a higher saliency is assigned to the players. Different from existing methods, due to the speed of the player in the red frame in Figure 4 (c) is outside the threshold range, so the area where it is located is smoothed strongly. After processing, the blue frame area in Figure 4 (b) is clearer, and the visual perception of the red frame area is almost unchanged. This reflects that the method achieves the goal of improving subjective visual experience while rationally distributing the limited bit rates.

B. CU PARTITION DECISION STRATEGY BASED ON DT

In the proposed strategy, we use JND threshold, motion pattern and image texture features as classification attributes to construct a DT, so as to achieve a good compromise between encoding accuracy and computational complexity. However, the flexible partition type in VVC makes it difficult to achieve accurate classification of the CU partition structure. It is an effective solution to transform the complicated CU partition problem into a multi-classification problem.

The prediction of the CU partition in the original VTM decision framework is executed layer by layer, and the partition prediction of each depth layer is executed sequentially until the recursive is terminated. Although this scheme achieves the desired accuracy, the computational complexity is unacceptable. On this basis, a scheme in which QT and MTT are executed independently is proposed. When both QT and MTT are determined to be terminated, the partition of CTU is terminated. This scheme has achieved effective coding efficiency, but the classification problem of MTT has not been fundamentally solved. According to the result of CU depth distribution, it is noticed that most CUs select QT as the optimal partition. If QT is decided in the early stage, then the complicated MTT will be terminated, and the calculation cost of RDO will be saved. Based on this principle, a new CU partition decision framework is proposed, as shown in Figure 6.

The partition types are defined as NO Spilt (NS), QT, BT_H, BT_V, TT_H, TT_V. In this framework, the CU classification problem is broken down into a combination of several binary classification problems, and only one partition type is evaluated at each decision level. Considering the advantages of texture features when dealing with large-size CUs, we add texture information as the classification attribute of the DT to further reduce coding time. Therefore, the three measurement features of texture information, JND threshold and motion state are adopted as classification attributes by the proposed method.

1) OVERALL TEXTURE

The overall texture characteristics of CTU are reflected by homogeneity and texture direction. We define the attribute set of the overall texture as follows,

$$D_1 = \left\{D_{NG}, D_{NMGM}, D_{AG\_H}, D_{AG\_V}, D_{CUS}\right\} \tag{8}$$

The size of the current CU is represented by D_{CUS}. The horizontal gradient and vertical gradient of the current CU

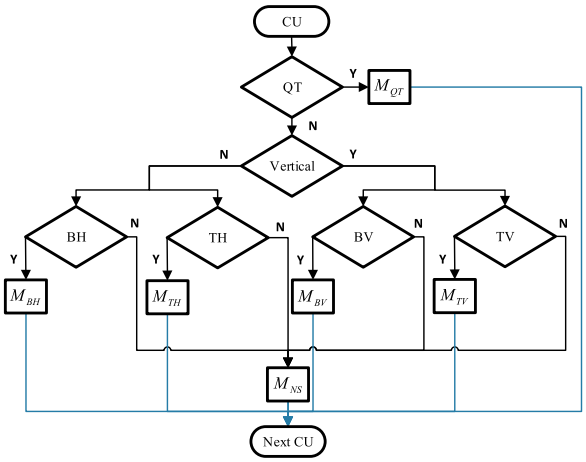

FIGURE 6. Proposed classification decision framework.

are obtained as follows,

$$G_x = \sum_{i=1}^{W}\sum_{j=1}^{H} A_{i,j} \times \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{9}$$

$$G_y = \sum_{i=1}^{W}\sum_{j=1}^{H} A_{i,j} \times \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{10}$$

where G_x and G_y represent the horizontal and vertical gradients, respectively. $A_{i,j}$ represents the 3rd order matrix centered on the current point. D_{NG} and D_{NMGM} represent the normalized gradients and the normalized maximum gradient magnitude, which reflect the homogeneity of the current CU, and the CUs with lower value are prefer not to perform segmentation. D_{NG} and D_{NMGM} are calculated as follows,

$$D_{NG} = \frac{|G_x| + |G_y|}{N_p} \tag{11}$$

$$D_{NMGN} = \frac{G_{\max}}{N_p} \tag{12}$$

where the number of pixels is represented by N_p. $G_{\max}$ is the maximum gradient of current CU, which is the maximum value of G_x and G_y. The average gradients in the horizontal direction and vertical direction are represented by D_{AG_H} and D_{AG_V}, which reflect the direction of the texture and provide a reference for the direction decision of MTT. They are expressed as,

$$D_{AG\_H} = \frac{\sum_{i=1}^{W}\sum_{j=1}^{H} abs\,(G_x)}{N_p} \tag{13}$$

$$D_{AG\_V} = \frac{\sum_{i=1}^{W}\sum_{j=1}^{H} abs\,(G_y)}{N_p} \tag{14}$$

2) JND THRESHOLD

As a metric feature, JND not only clearly describes local texture features, but also makes the coding process more consistent with the visual perception characteristics. Based on the proposed JND model, the attribute set of JND threshold is defined as follows,

$$D_2 = \{D_J\} \tag{15}$$

The accommodate distortion ability of the current CU increases with the increase of the JND value. The proposed scheme tends to reduce the bit allocation of areas with high JND thresholds to remove visual redundancy, and divide the current CU into sub-CUs with large size.

3) MOTION PATTERN

In most cases, the objects and details of the video content are very rich. There is no doubt that these contents cannot be accurately and completely identified by HVS in a short period of time. Therefore, the visual saliency distribution is optimized according to the classification of the motion pattern, and the area that has a positive effect on improving the subjective visual quality is enhanced. At the same time, the depth level of the motion areas that cannot be accurately perceived is appropriately reduced without affecting the subjective visual quality. The attribute set of different motion patterns as follows,

$$D_3 = \{D_{ST}, D_{OR}, D_{AL}\} \tag{16}$$

If the motion pattern of current CU is D_{OR}, the possibility of this CU being divided into higher depth will be significantly increased. Conversely, if the motion pattern of current CU is D_{AL}, the low depth partition tends to be selected. In particular, if the motion pattern of current CU is D_{ST}, this step will be terminated, and the motion pattern is no longer needed as one of the classification attribute which can effectively reduce the computational complexity. Since stationary area has very low visual saliency, the texture attributes and JND attributes are sufficient to provide a classification basis for the DT.

The proposed classifier adopts offline training and pruning process. The standard VVC test sequences with different resolutions are used as training sequences. Considering the objectivity of experimental results, the sequences participating in the test with the proposed method are not used as training sequences, as shown in Table2. To avoid overfitting, the confidence of each leaf-node is set to 90%. When the confidence level is less than 90%, we believe that the best partition of the CUs lying in such leaf nodes cannot be determined. In this case, RDO will be executed to ensure coding accuracy.

IV. EXPERIMENTAL RESULTS

Performance evaluation of the proposed method is demonstrated in this section. Five sets of video sequences with different resolutions are used for testing, and the coding performance of the proposed method is evaluated from two dimensions of subjective quality and objective quality. First, the overall performance of the proposed method is tested and

TABLE 2. Training sequences.

Resolution	Sequences	Frames
3840×2160	ParkRunning3	100
	Tango2	100
1920×1080	EBURainFruits	250
	FlyingGraphics	300
	Desktop	600
1280×720	SlideShow	500
	SlideEditing	300
832×480	BasketballDrilText	500
416×240	BasketballPass	500

TABLE 3. Testing condition.

Parameter	Description
Testing platform	VTM-7.0
Configuration	All-Intra
Frames	200
Hardware	Inter Core i5-8500 CPU 3.00 GHz
Others	The same as default VTM-7.0

analyzed. Then, compared with others, the validity of the proposed JND model is verified. Finally, the proposed method is compared with other fast algorithms to obtain an objective performance evaluation. The experiment was carried out on VTM-7.0, and the Common Test Condition (CTC) under *All-Intra* configuration (the file encoder_intra_vtm.cfg) were adopted. All experiments run on *Inter Core i5-8500 CPU 3.00 GHz* platform, as shown in Table3.

A. OBJECTIVE EVALUATION

The PSNR is used to reflect the objective quality of coding. However, recent studies have shown that there are some differences between PSNR and real human visual perception. It is necessary to add other evaluation indicators to improve the accuracy of quality evaluation. In order to make the objective assessment consistent with the visual perception, we added the Cumulative Probability of Blur Detection (CPBD) to measure the degree of artifacts that affect the perceived quality. The PSNR reflects the level of noise in the sequence. With the same test conditions, lower PSNR means lower distortion. The value of CPBD is between 0-1, and a higher value means better encoding accuracy. In addition, Bjøntegaard Delta Bit Rate (BDBR) and Time Saving (TS) are used to evaluate the bit rate saving and time saving of the proposed method. Lower BDBR and higher TS mean higher coding efficiency.

Table 4 provides the overall performance results. Simulation experiments show that the proposed method achieved an average CPBD of 0.64 bps. In terms of CPBD, this is a satisfactory result, which means higher encoding accuracy. Compared with original VTM, the proposed method achieved a time saving of 48.01% while increasing the BDBR by 0.79% on average. Among them, the minimum gain of B is 0.39%, and the maximum gain of TS is 59.62%. The results show that the proposed method effectively removes redundancy. The results also show that the proposed method performs well for high-resolution sequences and meets the coding expectations of VVC. Meanwhile, the proposed method achieves similar coding performance with different QPs, which means that the performance has better robustness.

Under the same encoding configuration, the PSNR results of the proposed JND model and others are compared, as shown in Figure 7. The result show that the proposed model achieves the lowest PSNR, which is reduced by 3.27%, 2.65%, and 2.54%, compared with MET [25], FEP [38] and PC [39], respectively. The results of CPBD and PSNR both show that the proposed method effectively improves the objective quality of coding.

The overall performance difference with other fast methods is shown in Table 5. According to the result, compared with QPBD and FPDV, the BDBR of the proposed method is reduced by 3.37% and 0.81% respectively under the same encoding times. This means that the proposed method reduces the bit rate even more. Compared with the CNN method, our method achieves the same level of BDBR, but saves 15.59% of the time, which achieves a considerable reduction in complexity.

These results demonstrate that the proposed method has excellent performance in terms of computational complexity and coding quality.

Figure 8 shows the R-D performance of four different resolution test videos processed by the proposed method and original VTM, including ''BQTerrace'', ''PartyScene'', ''BQsquare'', and ''Fourpeople''. Obviously, the proposed method has almost consistent R-D performance as the original VTM.

B. SUBJECTIVE EVALUATION

According to the methodology for the subjective assessment of the quality of television pictures ITU-R BT. 500-14 [43], we selected 20 viewers for the test. Among them, 10 viewers were researchers engaged in image processing. The other 10 people have no research experience, including 6 students and 4 teachers. Before the test, the viewers were informed of the purpose of the test and the scoring criteria. Compared with the video encoded by original VTM, we define the subjective visual quality as 5 levels with a rating range of [0, 5]. The subjective score standards are shown in Table 6.

In order to minimize the errors caused by environmental and display changes, the views were arranged to test one by one with two identically configured displays in the same environment. The sequence encoded by original VTM and the sequence encoded by proposed method are displayed on two adjacent screens. The distance between the screen and the viewer is 85 cm. After the test starts, 3 non-test video sequences are played to the audience to stabilize mood and

TABLE 4. The overall performance of the proposed method.

Class	Sequence	CPBD (bps)		PSNR (dB)		BDBR (%)		TS (%)	
		QP22	QP32	QP22	QP32	QP22	QP32	QP22	QP32
Class A 1920×1080	Cactus	0.59	0.61	29.24	28.82	0.57	0.71	49.75	52.16
	Kimono	0.49	0.56	27.03	26.91	0.48	0.69	56.83	59.62
	ParkSecene	0.50	0.58	29.93	29.68	0.39	0.67	56.01	48.35
	BasketballDrive	0.63	0.74	28.17	27.10	1.13	1.42	54.96	57.81
	BQTerrace	0.64	0.62	29.02	28.84	0.61	0.55	44.37	46.77
Class B 1280×720	KristenAndSara	0.62	0.61	29.59	29.49	1.03	1.24	41.14	45.36
	Johnny	0.51	0.57	27.95	27.86	0.97	1.09	51.98	55.65
	FourPeople	0.64	0.61	29.26	29.11	1.26	1.18	40.23	44.49
Class C 832×480	PartyScene	0.70	0.78	28.35	28.17	0.52	0.62	40.98	42.57
	BasketballDrill	0.59	0.73	27.01	26.74	0.74	0.81	48.16	49.73
	RacehorsesC	0.60	0.65	27.48	27.39	0.35	0.42	45.83	48.61
	BQMall	0.82	0.78	29.03	28.97	1.49	1.73	44.23	45.36
Class D 416×240	BQSquare	0.69	0.67	29.31	29.28	0.59	0.72	37.86	39.42
	RaceHorses	0.71	0.68	28.57	28.47	0.41	0.43	36.65	38.65
Class E 3840×2160	FoodMarket4	0.61	0.65	26.99	27.52	0.57	0.62	52.26	52.76
	CatRobot1	0.59	0.63	27.78	28.29	0.64	0.73	49.78	50.52
	Campfire	0.65	0.68	27.25	27.94	0.62	0.71	51.33	52.02
Average		0.62	0.66	28.35	28.27	0.73	0.84	47.20	48.81

TABLE 5. The overall performance comparison between the proposed and others.

Class	Sequence	QBPD [41]		FPDV [42]		CNN [43]		Proposed	
		BDBR (%)	TS (%)	BDBR (%)	TS (%)	BDBR (%)	TS (%)	BDBR (%)	TS (%)
Class A 1920×1080	Cactus	4.14	49.20	1.84	52.44	/	/	0.71	52.16
	Kimono	1.97	65.40	1.93	59.51	0.87	33.32	0.69	59.62
	ParkSecene	5.08	60.80	1.26	51.84	0.83	35.41	0.67	48.35
	BasketballDrive	4.22	67.98	3.28	59.35	/	/	1.42	57.81
	BQTerrace	3.47	48.57	1.08	45.30	0.95	34.50	0.55	46.76
Class B 1280×720	KristenAndSara	5.15	43.34	2.78	55.11	1.61	34.84	1.24	45.36
	Johnny	7.33	52.27	3.22	56.88	/	/	1.09	55.65
	FourPeople	5.08	43.10	2.70	57.57	1.38	38.01	1.18	44.49
Class C 832×480	PartyScene	1.90	40.25	0.26	38.62	0.55	31.10	0.62	42.57
	BasketballDrill	6.64	44.61	1.82	48.48	1.30	33.39	0.81	49.73
	RacehorsesC	2.42	36.24	0.88	49.05	0.37	23.63	0.42	48.61
	BQMall	4.87	30.19	1.87	52.47	0.98	31.77	1.73	45.36
Class D 416×240	BQSquare	4.53	37.47	0.19	31.95	0.68	30.73	0.72	39.42
	RaceHorses	2.74	29.33	0.54	41.69	0.71	31.79	0.43	38.64
Average		4.25	46.34	1.69	50.01	0.93	32.59	0.88	48.18

TABLE 6. Subjective score standards.

Subjective Visual Perception	Opinion Score
No difference observed	0
Difference just observed	1
Slight improvement	2
General improvement	3
Significant improvement	4
Excellent	5

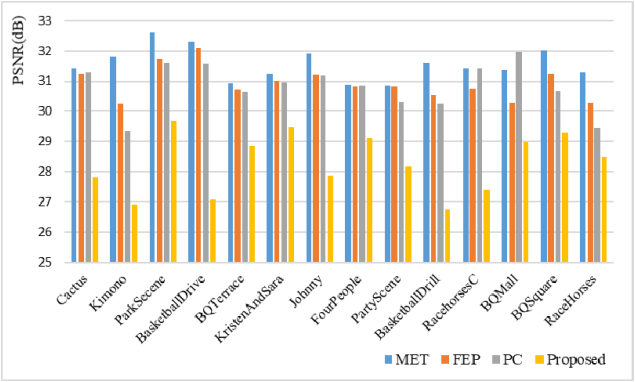

FIGURE 7. PSNR comparison between the proposed JND model and others.

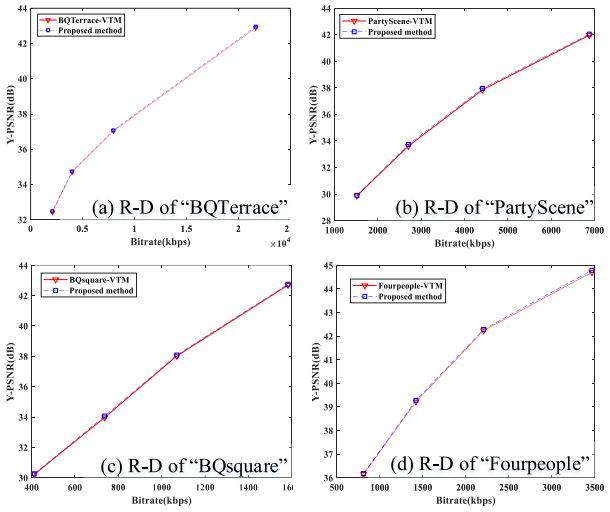

FIGURE 8. R-D performance.

vision first. The test video sequences are played sequentially, with an interval of 3 seconds between each two video sequences for viewers to score. After all video sequences with the same QP value have been tested, the video sequences with another QP are tested.

According to the statistical specifications of ITU-R BT. 500-14, the confidence interval is set to 95%. The Perceived Gain (PG) can be calculated according to the statistical results of MOS and SD in Table 7 as follows,

$$\omega - \frac{2.093\eta}{\sqrt{N_v}} \tag{17}$$

FIGURE 9. The perceptual difference of the "Johnny" sequence. (a) processed by the FPDV, (b) processed by the proposed method.

where N_v is the number of viewer. ω is the Mean Opinion Score (MOS), and η is the Standard Deviation (SD) of opinion score. Both of the above are calculated by the viewer scoring statistics. The subjective quality assessment results are shown in Table 7.

As illustrated in Table 7, the sequence processed by the proposed method achieves an average MOS of 3.36, which means that the subjective perception quality of the test sequence has been improved. Moreover, the difference in SD with different QP values is small, which means that the proposed method has good generalization ability. Compared with the original VTM, the PG of proposed method is between 2.66 and 3.64, which achieves a stable subjective quality improvement for different sequences. What need be noted is that, the MOS of the Class E is slightly lower than other Class under similar objective quality indicators. The reason is that the subjective visual improvement of the proposed algorithm is more difficult to be noticed on the basis that the Ultra High Definition sequence originally has an excellent visual experience.

The perceptual difference between the sequence processed by the proposed method and the other algorithms are shown in Figure 9 and 10. In general, the proposed method achieves the closest overall performance to FPDV. In order to reflect the difference in visual perception with similar objective performance, the test sequence is decoded separately by the FPDV method and the proposed method. The area of ordinary pattern is enlarged as shown in Figure 9. It can be seen that the sequence processed by the proposed method assigns more saliency to the face, so the details and edges of Figure 9 (b) are clearer, which confirms the conclusion that the proposed method effectively improves the subjective visual perception.

(a) (b)

FIGURE 10. The perceptual difference of the "BasketballDrive" sequence. (a) represents the original sequence, (b) processed by the proposed method.

TABLE 7. The subjective quality assessment results.

Class	Sequence	MOS		SD		PG	
		QP22	QP32	QP22	QP32	QP22	QP32
Class A 1920×1080	Cactus	3.74	3.59	0.22	0.27	3.64	3.46
	Kimono	3.53	3.52	0.37	0.40	3.36	3.33
	ParkSecene	3.35	3.43	0.32	0.33	3.20	3.28
	BasketballDrive	3.50	3.49	0.34	0.30	3.34	3.35
	BQTerrace	3.60	3.63	0.34	0.35	3.44	3.47
Class B 1280×720	KristenAndSara	3.66	3.68	0.31	0.24	3.51	3.57
	Johnny	3.50	3.50	0.32	0.31	3.35	3.35
	FourPeople	3.58	3.40	0.51	0.30	3.34	3.26
Class C 832×480	PartyScene	3.62	3.76	0.34	0.29	3.46	3.62
	BasketballDrill	3.35	3.51	0.20	0.30	3.26	3.37
	RacehorsesC	3.53	3.64	0.41	0.35	3.34	3.48
	BQMall	3.36	3.40	0.28	0.23	3.22	3.29
Class D 416×240	BQSquare	3.10	3.23	0.38	0.26	2.92	3.10
	RaceHorses	3.05	2.83	0.52	0.36	2.81	2.66
Class E 3840×2160	FoodMarket4	2.75	2.83	0.34	0.24	2.59	2.72
	CatRobot1	2.93	2.68	0.29	0.36	2.79	2.51
	Campfire	3.05	3.03	0.22	0.29	2.94	2.89
Average		3.36	3.36	0.34	0.30	3.20	3.22

In order to show the overall performance more intuitively, we choose a sequence with multiple motion patterns for testing in Figure 10. It can be observed that the face area reflects the ordinary pattern, while the arm and its adjacent area conform to the aliased pattern. Compared with original VTM, the region of ordinary pattern exhibits a higher visual quality, and the obvious loss of visual quality is not noticed in the region of aliased pattern. Obviously, the video quality produced by the proposed method is more pleased to humans.

V. CONCLUSION

In this paper, a fast CU partition decision strategy is proposed. The existing methods cannot further compress the video in a way that is more in line with the visual characteristics of the HVS. To improve this defect, we use the JND model and the

motion state as classification attributes, and combine them with the decision tree to develop a CU partition decision strategy oriented to the perceived quality of HVS. In this paper, the color complexity is added as a visual weighting factor to the pixel-based JND model, and the value of texture information and color information is further mined to improve the accuracy. Subsequently, according to the principle of visual saliency, the motion pattern of the region is classified, which provides a basis for CU division and removes visual redundancy. Finally, the above two elements are combined with CTU texture features to build a DT. Experimental results show that the complexity of the proposed method is reduced by about 48.01%, while the increase of BDBR is only 0.79%. Considering the subjective and objective quality evaluation results, we can conclude that the proposed method achieves stable visual perception quality improvement while reducing computational complexity.

REFERENCES

[1] Y.-K. Wang, R. Skupin, M. M. Hannuksela, S. Deshpande, V. Drugeon, R. Sjoberg, B. Choi, V. Seregin, Y. Sanchez, J. M. Boyce, W. Wan, and G. J. Sullivan, "The high-level syntax of the versatile video coding (VVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Apr. 5, 2021, doi: 10.1109/TCSVT.2021.3070860.

[2] H. Schwarz, M. Coban, M. Karczewicz, T.-D. Chuang, F. Bossen, A. Alshin, J. Lainema, C. R. Helmrich, and T. Wiegand, "Quantization and entropy coding in the versatile video coding (VVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Apr. 9, 2021, doi: 10.1109/TCSVT.2021.3072202.

[3] Q. Zhang, P. An, Y. Zhang, L. Shen, and Z. Zhang, "Low complexity multiview video plus depth coding," *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, pp. 1857–1865, Nov. 2011, doi: 10.1109/TCE.2011.6131164.

[4] L. Xu, W. Lin, L. Ma, Y. Zhang, Y. Fang, K. N. Ngan, S. Li, and Y. Yan, "Free-energy principle inspired video quality metric and its use in video coding," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 590–602, Apr. 2016.

[5] S. Valizadeh, P. Nasiopoulos, and R. Ward, "Perceptually-friendly rate distortion optimization in high efficiency video coding," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 115–119.

[6] H. Hadizadeh, M. J. Enriquez, and I. V. Bajic, "Eye-tracking database for a set of standard video sequences," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 898–903, Feb. 2012, doi: 10.1109/TIP.2011.2165292.

[7] H. Yang, L. Shen, X. Dong, Q. Ding, P. An, and G. Jiang, "Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 6, pp. 1668–1682, Jun. 2020, doi: 10.1109/TCSVT.2019.2904198.

[8] Q. Zhang, Y. Zhao, B. Jiang, L. Huang, and T. Wei, "Fast CU partition decision method based on texture characteristics for H.266/VVC," *IEEE Access*, vol. 8, pp. 203516–203524, 2020, doi: 10.1109/ACCESS.2020.3036858.

[9] T. Katayama, T. Song, T. Shimamoto, and X. Jiang, "Reference frame generation algorithm using dynamical learning PredNet for VVC," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2021, pp. 1–5, doi: 10.1109/ICCE50685.2021.9427694.

[10] J. Zhang, S. Kwong, and X. Wang, "Two-stage fast inter CU decision for HEVC based on Bayesian method and conditional random fields," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3223–3235, Nov. 2018, doi: 10.1109/TCSVT.2017.2747618.

[11] K. Kim and W. W. Ro, "Fast CU depth decision for HEVC using neural networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 5, pp. 1462–1473, May 2019, doi: 10.1109/TCSVT.2018.2839113.

[12] H.-S. Kim and R.-H. Park, "Fast CU partitioning algorithm for HEVC using an online-learning-based Bayesian decision rule," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 130–138, Jan. 2016, doi: 10.1109/TCSVT.2015.2444672.

[13] M. Grellert, B. Zatt, S. Bampi, and L. A. da Silva Cruz, "Fast coding unit partition decision for HEVC using support vector machines," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 6, pp. 1741–1753, Jun. 2019, doi: 10.1109/TCSVT.2018.2849941.

[14] H. Yang, L. Shen, and P. An, "An efficient intra coding algorithm based on statistical learning for screen content coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2468–2472, doi: 10.1109/ICIP.2017.8296726.

[15] G. Correa, P. A. Assuncao, L. V. Agostini, and L. A. da Silva Cruz, "Fast HEVC encoding decisions using data mining," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 4, pp. 660–673, Apr. 2015, doi: 10.1109/TCSVT.2014.2363753.

[16] J. Ban, H. Lai, and X. Lin, "A novel method rate distortion optimization for HEVC based on improved SSIM," in *Proc. 9th Int. Symp. Comput. Intell. Design (ISCID)*, vol. 2, Dec. 2016, pp. 260–263.

[17] M. Zhou, X. Wei, S. Wang, S. Kwong, C. Fong, P. Wong, W. Yuen, and W. Gao, "SSIM-based global optimization for CTU-level rate control in HEVC," *IEEE Trans. Multimedia*, vol. 21, no. 8, pp. 1921–1933, Aug. 2019, doi: 10.1109/TMM.2019.2895281.

[18] P.-J. Lee, T.-A. Bui, and S.-R. Yao, "Improve the HEVC alogrithm complexity based on the visual perception," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2019, pp. 1–4, doi: 10.1109/ICCE.2019.8662038.

[19] S.-H. Bae and M. Kim, "A DCT-based total JND profile for spatiotemporal and foveated masking effects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 6, pp. 1196–1207, Jun. 2017, doi: 10.1109/TCSVT.2016.2539862.

[20] S.-H. Bae and M. Kim, "A novel generalized DCT-based JND profile based on an elaborate CM-JND model for variable block-sized transforms in monochrome images," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3227–3240, Aug. 2014.

[21] 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.

[22] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, "High dynamic range video compression exploiting luminance masking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 5, pp. 950–964, May 2016, doi: 10.1109/TCSVT.2015.2426552.

[23] C.-H. Chou, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," in *Proc. IEEE Int. Symp. Inf. Theory.*, Jun. 1994, p. 420, doi: 10.1109/ISIT.1994.395035.

[24] X. K. Yang, W. S. Ling, Z. K. Lu, E. P. Ong, and S. S. Yao, "Just noticeable distortion model and its applications in video coding," *Signal Process., Image Commun.*, vol. 20, no. 7, pp. 662–680, Aug. 2005.

[25] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1648–1652, Nov. 2010, doi: 10.1109/TCSVT.2010.2087432.

[26] S. Wang, L. Ma, Y. Fang, W. Lin, S. Ma, and W. Gao, "Just noticeable difference estimation for screen content images," *IEEE Trans. Image Process.*, vol. 25, no. 8, pp. 3838–3851, May 2016, doi: 10.1109/TIP.2016.2573597.

[27] S. Nami, F. Pakdaman, and M. R. Hashemi, "Juniper: A jnd-based perceptual video coding framework to jointly utilize saliency and JND," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2020, pp. 1–6, doi: 10.1109/ICMEW46912.2020.9106036.

[28] M. Zhou, X. Wei, S. Kwong, W. Jia, and B. Fang, "Just noticeable distortion-based perceptual rate control in HEVC," *IEEE Trans. Image Process.*, vol. 29, pp. 7603–7614, 2020, doi: 10.1109/TIP.2020.3004714.

[29] Q. Zhang, Y. Wang, L. Huang, B. Jiang, and R. Su, "Adaptive CU split prediction and fast mode decision for 3D-HEVC texture coding based on just noticeable difference model," *Digital Signal Process.*, vol. 106, Nov. 2020, Art. no. 102851, doi: 10.1016/j.dsp.2020.102851.

[30] 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, doi: 10.1109/TIP.2009.2030969.

[31] J. Kim, D. Y. Lee, S. Jeong, and S. Cho, "Perceptual video coding using deep neural network based JND model," in *Proc. Data Compress. Conf. (DCC)*, Mar. 2020, p. 375, doi: 10.1109/DCC47342.2020.00087.

[32] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004.

[33] 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, doi: 10.1016/j.imavis.2010.07.001.

[34] Y.-F. Ma and H.-J. Zhang, "A model of motion attention for video skimming," in *Proc. Int. Conf. Image Process.*, Sep. 2002, p. 1, doi: 10.1109/ICIP.2002.1037976.

[35] H. Oh and W. Kim, "Video processing for human perceptual visual quality-oriented video coding," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1526–1535, Apr. 2013, doi: 10.1109/TIP.2012.2233485.
[36] Q. Zhang, Y. Zhao, B. Jiang, and Q. Wu, "Fast CU partition decision method based on Bayes and improved de-blocking filter for H.266/VVC," *IEEE Access*, vol. 9, pp. 70382–70391, 2021, doi: 10.1109/ACCESS.2021.3079350.
[37] C. Wang, Y. Wang, and J. Lian, "Just-noticeable distortion model based on color complexity and structure tensor," *J. Shanghai Univ. Natural Sci. Ed.*, vol. 1, pp. 1–11, Feb. 2021, doi: 10.12066/j.issn.1007-2861.2276.
[38] J. Wu, G. Shi, W. Lin, A. Liu, and F. Qi, "Just noticeable difference estimation for images with free-energy principle," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1705–1710, Nov. 2013, doi: 10.1109/TMM.2013.2268053.
[39] 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, doi: 10.1109/TIP.2017.2685682.
[40] Z. Wang, S. Wang, J. Zhang, S. Wang, and S. Ma, "Effective quadtree plus binary tree block partition decision for future video coding," in *Proc. Data Compress. Conf. (DCC)*, Apr. 2017, pp. 23–32, doi: 10.1109/DCC.2017.70.
[41] Y. Fan, J. Chen, H. Sun, J. Katto, and M. Jing, "A fast QTMT partition decision strategy for VVC intra prediction," *IEEE Access*, vol. 8, pp. 107900–107911, 2020, doi: 10.1109/ACCESS.2020.3000565.
[42] G. Tang, M. Jing, X. Zeng, and Y. Fan, "Adaptive CU split decision with pooling-variable CNN for VVC intra encoding," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2019, pp. 1–4, doi: 10.1109/VCIP47243.2019.8965679.
[43] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Standard ITU-R Standard BT.500-14, 2019.

JINCHAO ZHAO received the master's degree in computer technology from the Huazhong University of Science and Technology, Wuhan, China, in 2009. He is currently a Professor with Zhengzhou University of Light Industry, China. His current research interests include video processing, Bayesian estimation, and intelligent computation.

TENGYAO CUI received the B.S. degree in electronic information science and technology from Zhengzhou University of Light Industry, Zhengzhou, China, in 2019, where he is currently pursuing the master's degree in software engineering with the School of Computer and Communication Engineering. His current research interests include image processing, high-efficiency video coding, machine learning, and extensions of the versatile video coding.

QIUWEN ZHANG (Member, IEEE) received the Ph.D. degree in communication and information systems from Shanghai University, Shanghai, China, in 2012. Since 2012, he has been with the Faculty of the College of Computer and Communication Engineering, Zhengzhou University of Light Industry, where he is currently an Associate Professor. He has published over 30 technical articles in the field of pattern recognition and image processing. His major research interests include 3D signal processing, machine learning, pattern recognition, video codec optimization, and multimedia communication.