TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)
31 Oct - 3 Nov 2023. Chiang Mai, Thailand

Perceptual Based Fast CU Partition Algorithm for VVC Intra Coding

Xin-Yi Cui
School of Electronics and Information Technology
Sun Yat-Sen University
Guangzhou, China
cuixy23@mail2.sysu.edu.cn

Fan Liang
School of Electronics and Information Technology
Sun Yat-Sen University
Guangzhou, China
Peng Cheng Laboratory
Shenzhen, China
isslf@mail.sysu.edu.cn

***Abstract*—The introduction of quad-tree with nested multi-type tree (QTMT) brings a significant reduction in bitrate to Versatile Video Coding (VVC) because QTMT enables a more flexible coding unit (CU) partition based on image content. However, the extensive rate distortion optimization (RDO) process induces a massive increase in encoding time. To mitigate this burden, a perceptual-based fast CU partition algorithm for VVC intra coding is proposed in this paper. The just noticeable difference model (JND) is adopted to simulate the human visual system, and JND variance is employed for early partition termination and division mode selection by reflecting the perceptual texture consistency. Experimental results on VTM17.0 show that the proposed method can achieve 31.13% time complexity reduction, with 1.32% Bjontegaard delta bit rate (BDBR) increase.**

***Index Terms*—VVC, Fast CU Partition, Perceptual Coding**

I. Introduction

Internet-based video now accounts for about 80% of global data traffic [1], and this figure is still growing. The Joint Video Experts Team (JVET) has developed Versatile Video Coding (VVC) [2], the latest generation international video coding standard, in response to high demand for better video compression. VVC obtains 50% bit-rate saving [3] over the previous High Efficiency Video Coding (HEVC) standard by adopting some advanced coding tools. However, computational complexity increased dramatically with the introduction of coding tools including Quadtree Plus Multi-Type Tree (QT+MTT), History-Based MV Prediction (HMVP), Affine Motion Compensation [4], etc.

The QTMT structure allows non-square coding units (CUs) in VVC since it extends Quad Tree (QT) with Binary Tree (BT) and Ternary Tree (TT). Fig. 1 shows a coding tree unit (CTU) partition example in VVC. The optimal CU split mode corresponding to the minimum rate distortion (R-D) cost is determined by recursively examining the rate distortion cost. More refined rectangular CUs make VVC block partitions more flexible and suitable for high resolutions. However, the brute-force rate distortion optimization (RDO) process brings a significant increase in encoding time.

Aimed at reducing coding complexity, there has been plenty of research on accelerating CU partition. First, there are some non-learning based strategies. Chen et al. [5] exploit gradient features and difference of sub-CUs to choose QT partition and decide the final splitting mode, respectively. Cui et al. [6] employ directional gradients to skip unnecessary modes. Considering combinations of different divisions may lead to the same CU structure, Cao et al. [7] design a history-based algorithm to accelerate partitioning. Liu et al. [8] use cross-block difference obtained by sub-block content and gradient to terminate redundant partition patterns.

Corresponding author: Fan Liang

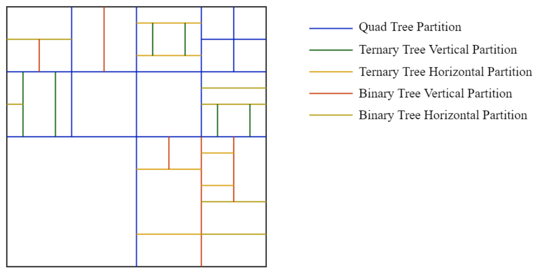


Fig. 1. Example of CTU partition under QTMT structure

Additionally, there are some fast partition algorithms based on learning. The early skip scheme proposed by Fu et al. [9] applies Bayesian decision rule to extract information of splitting modes. Tang et al. [10] employ CU shape-adaptive CNN architecture to determine CU splitting depth. Li et al. [11] propose a fast CU partition strategy based on a tunable decision model. In the fast partition method designed by He et al. [12], CUs are classified into different categories to apply corresponding random forest classifiers. Wu et al. [13] use Support Vector Machine to predict the partition of CU depending on texture features.

Currently, most of existing algorithms are focusing on ensuring objective quality. Because higher quality video is for a better human viewing experience, it is important to take subjective perception into account in video encoding. If the CU splitting process is improved based on perception, not only the computational complexity will be further reduced, but also the subjective quality of encoded videos can be guaranteed. Therefore, a perceptual-based fast CU partition strategy is proposed in this paper. We explore the perceptual characteristics of CUs with the help of the just noticeable difference (JND) model. JND variance and differences between sub-blocks after partition are employed to gauge the perceived consistency of CUs. With our method, which can save 31.13% encoding time with 1.32% Bjontegaard delta bit rate (BDBR) increase, redundant division patterns can be effectively avoided with negligible subjective performance loss.

979-8-3503-0219-6/23/$31.00 ©2023 IEEE

This paper is organized as follows: Section II describes the perceptual algorithm in detail. Section III shows the experimental results based on VTM17.0. The conclusion is presented in Section IV.

II. Proposed Algorithm

Due to the limited visual resolution, the human visual system (HVS) cannot perceive distortion below a certain threshold, which can be well estimated by the JND model. This paper mainly adopts the improved pixel-domain JND model [14] considering the effect of content regularity, luminance adaptation, and contrast masking. This model suggests that HVS is more sensitive to regular visual content than irregular one.

A. Just Noticeable Diffference (JND) Model

JND model in [14] defines $\mathcal{N}$ as quantified orientation differences to measure content regularity. $\mathcal{N}$ is obtained from correlations between orientations of pixels in a local region. Then the visual masking is described below:

$$\begin{aligned} V_M(x,y) &= f(L_c(x,y), \mathcal{N}(x,y)) \\ &= \frac{1.84 \cdot {L_c}^{2.4}}{{L_c}^2 + 26^2} \cdot \frac{0.3 \cdot \mathcal{N}^{2.7}}{\mathcal{N}^2 + 1} \end{aligned} \tag{1}$$

where (x, y) and L_c represents the position and luminance contrast, respectively. According to (1), visual masking is more prominent in regions with higher luminance contrast and lower visual regularity.

Another factor called luminance adaptation, which is modeled as a background luminance ($B(x, y)$)-based segmentation function:

$$L_A(x,y) = \begin{cases} 17 \times (1 - \sqrt{\frac{B(x,y)}{127}}) & B(x,y) \le 127 \\ \frac{3}{128} \times (B(x,y) - 127) + 3 & B(x,y) > 127. \end{cases} \tag{2}$$

Finally, with $L_A(x, y)$ and $V_M(x, y)$, the JND threshold can be calculated as follow:

$$\begin{aligned} JND(x,y) &= L_A(x,y) + V_M(x,y) \\ &\quad - 0.3 \cdot min\left\{L_A(x,y), V_M(x,y)\right\}. \end{aligned} \tag{3}$$

B. Early Termination of JND Based CU Splitting

The perceptual homogeneity of an image region is reflected by the JND variance. A smaller JND variance indicates that the region's visual features are more homogeneous, while a bigger JND variance indicates an image region contains more content fluctuations [15]. A perceptual distortion factor defined by Xiang et al. [16] is employed here:

$$P_F(d,i,j) = \log_2\left(V_j(d,i,j)\right) \tag{4}$$

with

$$V_j(d,i,j) = \sum_{m=1}^{M} \sum_{n=1}^{N} [JND(x,y) - \overline{JND}] \tag{5}$$

where d refers to CU's depth, i and j are CU's index. $V_j(d, i, j)$ refers to JND variance and $\overline{JND}$ means the average of spatial JND. If the perceptual distortion factor is less than a given threshold, then the CU can be considered as smooth and the partition is terminated directly. In this paper, we use $P_F(d, i, j)$ for early termination division judgment for some relatively small CUs.

In addition, for 32×32 CUs, we use the perceptual difference between sub-CUs to make CU size decisions. When we apply QT partition to the current CU, the larger value of perceptual difference between sub-CUs means richer perceptual characteristics of current CU and a larger difference between P_F of sub-CUs. And the maximum difference between P_F of sub-CUs can be calculated as follow [16]:

$$D_p(d,i) = \max_{m,n \in \{0,1,2,3\}} \left\{\left|P_F\left(d,i,m\right) - P_F\left(d,i,n\right)\right|\right\} \tag{6}$$

where m and n are index representing sub-CUs. In this paper, we also use D_P for the determination of early termination. If D_P is less than a certain threshold, then the perceptual characteristics are less sensitive for human eyes.

C. Mode Decision of JND Based CU Splitting

Given that QTMT structure has been applied in VVC, it will be better to further determine the optimal partition. The variance of the perceptual distortion factors of the sub-CUs denoted as V_P can measure the perceptual complexity in a local region. The larger the variance, the greater the perceptual differences between the sub-CUs, and our eyes will be more sensitive to the characteristics of sub-CUs. Then we define the maximum variance as:

$$V_m = \max_{q \in \{BH,BV,TH,TV\}} \left\{V_P\left(d,k,q\right)\right\} \tag{7}$$

where q is the index of splitting modes. The partition corresponding to V_m will be added to the candidate list. According to (7), the variance of P_F of sub-CUs after quadtree partition doesn't need to be calculated. That's because we choose to always add QT to the partition mode candidate list for a better quality of the encoded video.

D. JND Based CU Splitting Algorithm

In the proposed algorithm, we use threshold TL1 for P_F and threshold TL2 for D_P. We obtain the value of TL1 and TL2 from experiments, which are shown in Table I. Since applying the JND model directly to large-size CUs results in large distortion easily and the perceptual characteristic in small-size CUs will not change notably, we only carry the method on CUs with specified sizes shown in the first column of Table I, and S denotes the CU's size ($width\ of\ CU \times height\ of\ CU$). Fig. 2 shows the flowchart of the JND based CU splitting algorithm.

TABLE I
Threshold values for various QPs and CU size

CU Size	QP	TL1	TL2
$S = 256$	[0, 30]	-2.5	/
	[31, 63]	-2	/
$S = 512$	[0, 63]	-2	/
$S = 1024$ (only 32×32)	[0, 63]	/	0.2

(1) For CUs with $S \in [256, 512]$, P_F is calculated and compared with TL1. If P_F is less than TL1, the splitting is terminated and the RDO procedure can be skipped early. Or else the current CU is encoded by default.

(2) For 32×32 CUs, it would be less accurate to use PF for the decision because of the bigger size. D_P is calculated and compared with TL2. If D_P is less than TL2, the region is considered smooth, then the splitting is terminated early.

Or else QT is added to the RDO list and V_m is computed to select another candidate splitting mode.

(3) Otherwise, if CU's size is out of the manageable range, the default encoding process is adopted.

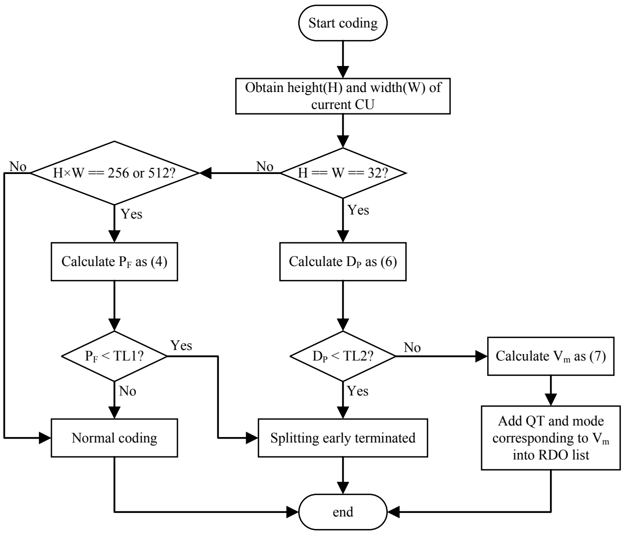


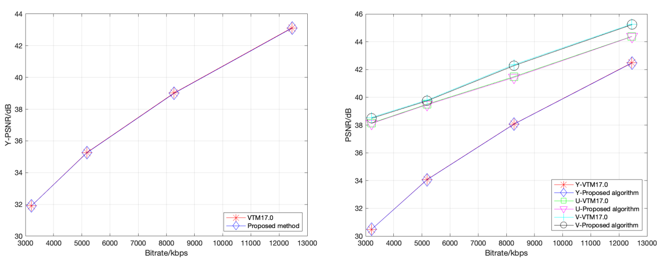
Fig. 2. Flowchart of the proposed method

III. Experimental Results

The experiments of the proposed algorithm are based on VVC reference software VTM17.0. The All-Intra configuration is implemented in this study with QPs set to $\{22, 27, 32, 37\}$. The tests were conducted on the first 100 frames of video sequences from common test condition [17]. The anchor points always keep the same as in the original VTM17.0. BDBR [18] is applied as a quality metric of the proposed algorithm. Time saving is computed as follows to represent encoding complexity.

$$TS = \frac{T_{anchor} - T_{pro}}{T_{anchor}} \times 100\% \qquad (8)$$

where T_{anchor} and T_{pro} denotes encoding time of original VTM17.0 and the proposed algorithm, respectively. Calculation of the judgment conditions brings about 4.94% extra time complexity, and it is included in T_{pro}. Table II demonstrates the experimental results compared with the original VTM17.0 algorithm, achieving an encoding time saving of 31.13% for an average BDBR increase of 1.3168%. In addition, Fig. 3 indicates that our perceptual algorithm maintains the rate distortion performance of sequence *BQSquare* compared with the original VTM17.0 algorithm. As shown in Table III, when comparing the results of our proposed algorithm with the SVM-based algorithm [13], both have similar E and our proposed algorithm has a smaller BDBR.



(a) YUV-PSNR vs Bitrate (b) Y/U/V-PSNR vs Bitrate

Fig. 3. RD curve comparison of *BQSquare*

TABLE II
Performance of proposed algorithm compared with VTM 17.0

Class	Sequence	**BDBR**[%]	**TS**[%]
A1 (3840×2160)	Tango2	2.1083	39.23
	FoodMarket4	1.2298	39.17
	Campfire	1.9055	37.33
A2 (3840×2160)	CatRobot1	2.2037	37.96
	DaylightRoad2	1.4408	38.57
	ParkRunning3	1.2145	39.69
B (1920×1080)	MarketPlace	1.6445	34.61
	RitualDance	2.1365	37.12
	Cactus	1.3653	36.30
	BasketballDrive	2.0784	36.94
	BQTerrace	0.7067	22.48
C (832×480)	RaceHorses	0.7345	29.26
	BQMall	1.1904	29.92
	PartyScene	0.4253	19.39
	BasketballDrill	1.4555	25.95
D (416×240)	RaceHorses	0.5558	19.25
	BQSquare	0.2144	13.20
	BlowingBubbles	0.5003	19.78
	BasketballPass	1.0577	23.84
E (1280×720)	FourPeople	1.4441	36.22
	Johnny	1.8225	36.47
	KristenAndSara	1.5347	32.16
Average value		**1.3168**	**31.13**

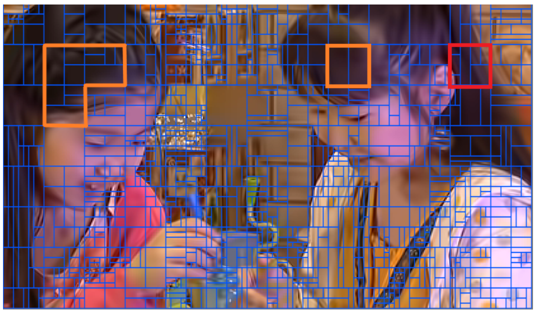
TABLE III
Comparision with related work

	ISCAS [13]	**Proposed**
BDBR[%]	2.71	1.32
TS[%]	63.16	31.13
TS/BDBR	**23.31**	**23.58**

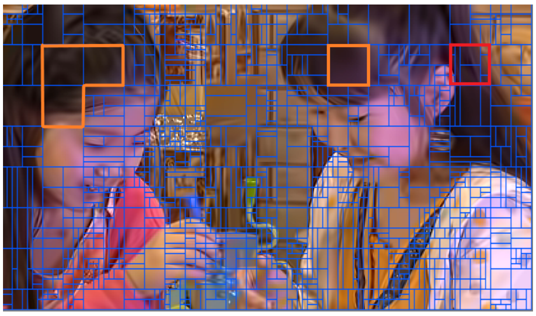
To visualize the performance of the proposed algorithm, the CU partition comparison of the first frame in *BlowingBubbles* between the original VTM17.0 algorithm and the proposed algorithm is presented in Fig. 4. Compared to the anchor, the splitting process of 32 × 32 CUs marked in Fig. 4 with orange borders is terminated early to obtain compression time saving. For some 32 × 32 CUs, the proposed algorithm makes different mode decisions compared with the original algorithm. For example, a representative 32 × 32 region marked with the red box in Fig. 4 is selected. Compared to the anchor which uses BT-V, the proposed algorithm adopts TT-V to separate the ear, hair, and background parts.

IV. Conclusion

In this paper, a fast CU partition algorithm based on the just noticeable difference is proposed for VVC intra coding. To reduce video encoding complexity, the perceptual distortion factor and maximum difference between sub-CUs are used to guide whether to skip unnecessary division. Addi-



(a) Anchor



(b) Proposed algorithm

Fig. 4. CU parition comparison of *BlowingBubbles*

tionally, our proposed algorithm determines the optimal split mode among {BT-H, BT-V, TT-H, TT-V} with maximization for the variance of perceptual distortion factors between sub-CUs. As demonstrated by the experiment results, our method is able to reduce encoding time by 31.13% with about 1.32% increase of BDBR.

References

[1] B. Bross, J. Chen, J. -R. Ohm, G. J. Sullivan and Y. -K. Wang, "Developments in International Video Coding Standardization After AVC, With an Overview of Versatile Video Coding (VVC)," in *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463-1493, Sept. 2021.

[2] B. Bross, J. Chen, S. Liu and Y. Wang, "Versatile Video Coding (Draft 10)," *JVET-S2001*, Jul. 2020.

[3] G. Sullivan, B. Bross, "White paper on the VVC standard," *JVET-W2019-v1*, Jul. 2021.

[4] B. Bross et al., "Overview of the Versatile Video Coding (VVC) Standard and its Applications," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736-3764, Oct. 2021.

[5] J. Chen, H. Sun, J. Katto, X. Zeng and Y. Fan, "Fast QTMT Partition Decision Algorithm in VVC Intra Coding based on Variance and Gradient," *IEEE Visual Communications and Image Processing (VCIP)*, Sydney, NSW, Australia, 2019, pp. 1-4.

[6] J. Cui, T. Zhang, C. Gu, X. Zhang and S. Ma, "Gradient-Based Early Termination of CU Partition in VVC Intra Coding," *Data Compression Conference (DCC)*, Snowbird, UT, USA, 2020, pp. 103-112.

[7] J. Cao, Y. Jia, F. Liang and J. Wang, "Encounter CU Again: History-Based Complexity Reduction Strategy for VVC Intra-Frame Encoder," *IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, Tampere, Finland, 2021, pp. 1-6.

[8] H. Liu, S. Zhu, R. Xiong, G. Liu and B. Zeng, "Cross-Block Difference Guided Fast CU Partition for VVC Intra Coding," *International Conference on Visual Communications and Image Processing (VCIP)*, Munich, Germany, 2021, pp. 1-5.

[9] T. Fu, H. Zhang, F. Mu and H. Chen, "Fast CU Partitioning Algorithm for H.266/VVC Intra-Frame Coding," *IEEE International Conference on Multimedia and Expo (ICME)*, Shanghai, China, 2019, pp. 55-60.

[10] G. Tang, M. Jing, X. Zeng, Y. Fan, "Adaptive CU split decision with pooling-variable CNN for VVC intra encoding," *IEEE Visual Communications and Image Processing (VCIP)*, Sydney, NSW, Australia, 2019, pp. 1-4.

[11] Y. Li, G. Yang, Y. Song, H. Zhang, X. Ding, D. Zhang, "Early intra CU size decision for Versatile Video Coding based on a tunable decision model," *IEEE Transactions on Broadcasting*, vol. 67, no. 3, pp. 710-720, Sept. 2021.

[12] Q. He, W. Wu, L. Luo, C. Zhu, H. Guo, "Random forest based fast CU partition for VVC intra coding," *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Chengdu, China, 2021, pp. 1-4.

[13] G. Wu, Y. Huang, C. Zhu, L. Song, W. Zhang, "SVM based fast CU partitioning algorithm for VVC intra coding," *IEEE International Symposium on Circuits and Systems (ISCAS)*, Daegu, Korea, 2021, pp. 1-5.

[14] J. Wu, G. Shi, W. Lin, C. C. J. Kuo, "Enhanced just noticeable difference model with visual regularity consideration," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016, pp. 1581-1585.

[15] H. Xie, G. Xiang, D. Yu, H. Yu, Y. Li, W. Yan, "Perceptual fast CU size decision algorithm for AVS2 intra coding," *IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, Singapore, 2019, pp. 277-281.

[16] G. Xiang, H. Jia, M. Yang, J. Liu, C. Zhu, Y. Li, X. Xie, "An improved adaptive quantization method based on perceptual CU early splitting for HEVC," *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, 2017, pp. 362-365.

[17] J. Boyce, K. Suehring, X. Li, V. Seregin, "JVET common test conditions and software reference configurations," *JVET-J1010-v1*, Apr. 2018.

[18] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33*, Apr. 2001.