Adaptive Inter CU Depth Decision for HEVC Using Optimal Selection Model and Encoding Parameters

Yue Li, Gaobo Yang, Yapei Zhu, Xiangling Ding, and Xingming Sun, *Senior Member, IEEE*

***Abstract*—High efficiency video coding adopts a new hierarchical coding structure, including coding unit (CU), prediction unit (PU), and transform unit to achieve higher coding efficiency than its predecessor H.264/AVC high profile. However, its hierarchical unit partitioning strategy leads to huge computational complexity. In this paper, an adaptive inter CU depth decision algorithm is proposed, which exploits both temporal correlation of CU depth and available encoding parameters. An optimal selection model of CU depth is established to estimate the range of candidate CU depth by exploiting the temporal correlation of CU depth among current CU and temporally co-located CUs. To reduce the accumulated errors in the process of CU depth prediction, the maximum depth of the co-located CUs and the coded block flag (CBF) of the current CU are used. Moreover, PU size and CBF information are also used to decide the maximum depth for the current CU. Experimental results show that the proposed CU depth decision approach reduces 56.3% and 51.5% average encoding time, and the Bjontegaard delta bit rate increases only 1.3% and 1.1% for various test sequences under the random access and the low delay B conditions, respectively.**

***Index Terms*—High efficiency video coding (HEVC), coding unit (CU), prediction unit (PU), CU depth.**

I. Introduction

With the continuous increase of video resolutions, High Definition (HD) and Ultra-High Definition (UHD) videos provide more realistic and immersive viewing experience for users. The amounts of video data increase enormously with the increase of spatial resolution. To achieve high coding efficiency, High Efficiency Video Coding (HEVC) [1] has been jointly standardized by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). HEVC adopts many advanced feature tools such as flexible quad-tree coding structure to improve coding efficiency. It achieves more than 50% bit rate savings with subjective image quality equivalent to that of the previous video coding standard H.264/AVC [2], [3].

Different from H.264/AVC and MPEG-2, an efficient block partitioning structure is adopted by HEVC to provide more flexibility. Specifically, four hierarchical block concepts are introduced, which include coding tree unit (CTU), coding unit (CU), prediction unit (PU) and transform unit (TU). That is, HEVC adopts a quad-tree block structure, in which the size of CTU is 64×64 pixels. CTU can be recursively partitioned into the smallest CU (SCU) of 8×8 pixels, and correspondingly the CU depth varies from 0 to 3. The quad-tree structure brings additional coding efficiency. However, the encoding complexity is significantly increased as well, since the exhaustive CTU partition, prediction mode decision and motion estimation (ME) [4], [5] are involved in the process of rate-distortion optimization (RDO). The increased complexity is a bottleneck for the practical applications of HEVC, especially for power-constrained devices or real-time applications. Therefore, it is highly desirable to optimize CTU partition for complexity reduction while maintaining high coding efficiency.

In the literature, there are many fast algorithms presented for the block partitioning structure of HEVC. They can be divided into two categories: fast CU depth decision (FCDD) and joint fast CU depth and mode decision (FCDMD). The FCDD methods are proposed to early determine the CU depth, which decides whether the current CU needs to be further partitioned or not. The joint FCDMD approaches reduce the computational complexity by early determining CU depth and mode decision simultaneously. Meanwhile, there are intra-coded frames and inter-coded frames in video compression. For intra FCDD [6]–[10] and intra FCDMD [11]–[13] techniques, they efficiently reduce the computational complexities of intra CU depth decision and/or intra mode decision by exploiting rate distortion (RD) cost correlation, neighboring CU depth correlation and local edge complexities. However, the time consumption of intra-encoding is much less than that of inter-encoding. Moreover, the intra FCDD techniques cannot be directly extended to inter CU depth decision. There also exist many fast algorithms to speed up the inter-frame encoding process, which are also divided into two categories: inter FCDD and inter FCDMD.

Existing inter FCDD algorithms can be divided into three types. The first type directly exploits the encoding parameters

Manuscript received August 24, 2016; revised December 26, 2016 and February 23, 2017; accepted April 16, 2017. Date of publication June 6, 2017; date of current version August 31, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61572183, Grant 61379143, Grant 61232016, and Grant U1405254, and in part by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20120161110014. *(Corresponding author: Gaobo Yang.)*

Y. Li, G. Yang, and X. Ding are with the School of Information Science and Engineering, Hunan University, Changsha 410082, China (e-mail: yueli@hnu.edu.cn; yanggaobo@hnu.edu.cn; xianglingding@163.com).

Y. Zhu is with the Faculty of Physics and Electronic Information Science, Hengyang Normal University, Hengyang 421002, China (e-mail: zyp1016@hynu.edu.cn).

X. Sun is with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: sunnudt@163.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2017.2704423

such as CBF [14], SKIP flag [15], RD cost [16], prediction residuals [17], motion vector (MV) and sample adaptive offset (SAO) [18] to early determine CU depth. For instance, simple second order prediction residuals statistics, which are directly obtained from the current CU independent of its spatially and temporally neighbouring CUs, were used to early determine CU depth [17]. In addition, a novel fast CU encoding scheme was proposed based on spatio-temporal encoding parameters including SAO, MV, TU size and CBF [18]. However, these approaches do not fully exploit the encoding parameter of PU mode in the procedure of CU depth decision.

The second type adopts the encoding information of neighboring or co-located CUs to predict the current CU depth. In [19], a fast inter CU decision algorithm was proposed based on latent sum of absolute differences (SAD) estimation. A two-layer ME scheme is adopted to obtain the SAD cost, and an exponential model is established to model the correlation between RD cost and SAD cost. Then, an adaptive threshold is computed from the last ten encoded CUs to early decide the CU depth. In [20], an early CU size decision method was proposed by adaptively determine the depth range using neighboring and co-located CUs depth information. In addition, a fast CU depth decision was proposed in [21]. Firstly, depth level "0" is early determined based on the depth correlation among spatial and temporal neighboring CTUs. Secondly, depth level "3" is also skipped for some CUs by exploiting the correlation between the optimal mode of prediction unit and the best CTU depth selection. However, these methods still do not fully exploit the temporal correlation of CU depth.

The third type establishes CU splitting decision rules by online or offline learning the spatio-temporal encoding information for fast CU depth decision. The most-widely used learning decision methods include Bayesian [22], [23], Markov Random Field (MRF) [25], *K* nearest neighbors [26], Support Vector Machine (SVM) [27]. In [23], a fast CU partitioning scheme was proposed by Bayesian decision rule. Firstly, online learning is designed by using the error Bayesian decision rule; Secondly, the joint online and offline learning is presented based on the minimum risk Bayesian decision rule. A machine learning based CU depth decision was presented in [24], which reduces the computational complexity with a given RD degradation constraint for each CU depth by using the offline training mode. A MRF based fast CU decision was proposed in [25], in which the CU depth decision thresholds are obtained by online training. In [26], a fast CU size selection method was proposed based on pyramid motion divergence (PMD), which is calculated as the variance of the down-sampled optical flow. In addition, a *K* nearest neighbors like method is presented to determine the CU splitting. In [28], a data mining based fast CU decision algorithm was proposed, in which the features of SKIP flag, Merge flag and RD cost are trained to create a decision tree. However, the lower CU depth in fast motion regions can be further early skipped, which is not fully exploited in these methods.

Most existing fast CU depth decision methods still have two drawbacks. First, it is intuitive that the depths are relatively higher for the CUs in fast motion region than those CUs in static region and smooth region. However, most existing fast CU depth decision approaches do not fully consider this to skip lower depth for the CUs in fast motion regions. Second, the temporal correlation of CU depth is not fully exploited. To address these two issues, an adaptive inter CU depth decision approach is proposed to further reduce the computational complexity of HEVC encoder. The basic idea behind this is to estimate the depth range by jointly exploiting the temporal CU depth correlation among neighboring frames and the encoding parameters. Specifically, the main contributions are two-folds: 1) An optimal prediction model is established to predict the range of CU depth, which can early skip the lower depth in motion-complex regions and early determinate the CU depth in smooth regions as well. 2) The correlation between the CU depth and the encoding parameter of PU mode is also exploited for early CU depth decision.

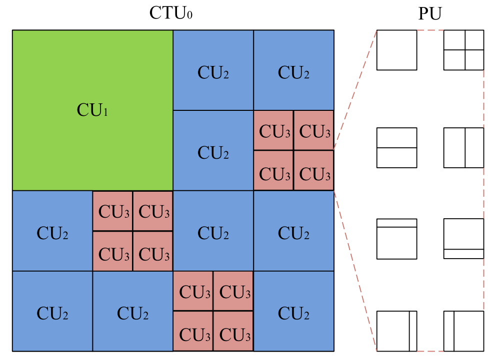


Fig. 1. CU and PU structures.

The rest paper is organized as follows. Section II presents our motivation and statistical analyses. Section III presents the prediction model of CU depth selection, and the proposed approach is discussed in detail. Experimental results are reported in Section IV. We conclude this paper in Section V.

II. Motivation and Statistical Analyses

Fig. 1 is the hierarchical block coding structure of HEVC. CTU can be recursively partitioned into four CUs until the CU is the SCU. For each PU, HEVC supports eleven candidate modes including two Intra modes (Intra_2N×2N, Intra_N×N), eight Inter modes (Inter_2N×2N, Inter_2N×N, Inter_N×2N, Inter_N×N, Inter_2N×nU, Inter_2N×nD, Inter_nL×2N and Inter_nR×2N), and Merge mode. To achieve higher coding efficiency, an exhaustive RDO process is applied for each CU depth to obtain the optimal prediction mode with the minimum RD cost. That is, the optimal mode is determined based on the RD cost, which is calculated as

$$J_m = D + \lambda \cdot B_m \tag{1}$$

where J_m is the RD cost function, D is the distortion between the original CU and the reconstructed CU, B_m represents the bit cost and λ is the Lagrangian multiplier. Then, the total RD cost of four partitioned sub-CUs is compared with the RD cost of its parent CU. Then, the current CU is decided whether to

TABLE I
TEST CONDITIONS

Search range	64
CTU size	64
Depth levels	"0", "1", "2", "3"
Configuration	Low Delay B
Number of encoded frames	100
Basis quantization parameter (QP)	24, 28, 32, 36

TABLE II
DISTRIBUTION OF CU DEPTH (UNIT:%)

Sequences	QP	Depth0	Depth1	Depth2	Depth3
Basketballpass (416x240)	24	16.89	34.96	30.50	17.65
	28	20.11	35.75	30.65	13.49
	32	23.32	37.99	29.92	8.77
	36	27.25	40.61	26.58	5.56
	average	21.89	37.33	29.41	11.37
Traffic (2560x1600)	24	34.60	32.30	24.18	8.92
	28	48.28	29.82	16.92	4.98
	32	58.79	27.03	11.45	2.73
	36	67.83	23.35	7.46	1.36
	average	52.38	28.12	15.00	4.50
Johnny (1280x720)	24	59.38	27.11	12.16	1.35
	28	70.78	20.23	8.25	0.74
	32	76.88	15.97	6.70	0.45
	36	81.72	12.35	5.65	0.28
	average	72.19	18.91	8.19	0.71
average		48.82	28.12	17.54	5.52

be further divided into sub-CUs or not. The partition rule can be expressed by

$$p = \begin{cases} un\_partition, & J_{cu} \leq J_{sub\_cus} \\ partition, & J_{cu} > J_{sub\_cus} \end{cases} \tag{2}$$

where $p \in \{un\_partition, partition\}$ represents the flag of a CU, which indicates whether it will be further split or not, J_{cu} and $J_{sub-cus}$ are the RD cost of CU and its four sub-CUs, respectively. Since the optimal CU depth is selected by recursively dividing CU into sub-CUs, it is very computation-intensive. If the optimal CU depth can be early and accurately predicted, it is likely to skip some unnecessary CU depth.

Three typical video sequences including *Basketballpass* (416×240), *Traffic* (2560×1600) and *Johnny* (1280×720) are encoded to analyze the optimal CU depth distribution because they have different motion activities. *Basketballpass* is a fast motion sequence, *Traffic* is a medium motion sequence, and *Johnny* is a head-shoulder sequence with slow motion. Table I summarizes the test conditions, and Table II reports the CU depth distribution under different quantization parameters (QPs). From Table II, the average percentages are 48.82%, 28.12%, 17.54% and 5.52% for the depth levels of "0" to "3", respectively. Especially, for *Johnny* sequence with slow motion, more than 72.19% CUs select level "0" as the optimal depth, and the percentage decreases with the increase of depth level. The reason is that there are a large proportion of homogenous and static background regions in nature videos, and most CUs in these regions choose lower depths. In this paper, the depth levels "0" and "1" are defined as the lower depths, while "2" and "3" are defined as the higher depths. If we can early determine whether a CU will select lower depths as its optimal depth or not, it is possible to greatly reduce computational cost by skipping the unnecessary process of further recursive CU partition. However, for fast motion sequences such as *Basketballpass*, the probability is only 21.89% for depth level "0" to be selected as the optimal depth. Apparently, it is much lower compared with the 72.19% of *Johnny* sequence. However, the more complex the motion of video sequences, the bigger the percentage of depth levels of "1", "2" and "3". Thus, we infer that complex motion videos are more likely to select higher CU depths. For videos with medium or fast motions, if they are treated with the same strategy for slow motion videos, the time saving will be limited. That is, for the regions with complex motion in medium and fast motion sequences, lower CU depth can be early skipped to reduce computational costs.

The depth of current CU is strongly related to that of its co-located CUs in previously encoded frames. Fig. 2 shows some optimal CU depths of three adjacent frames for *Basketballpass* and *Johnny* sequences when they are encoded with HM12.0 under Low Delay configuration. Apparently, static background regions are likely to select lower depths as the optimal depth, while medium or fast motion regions are more likely to choose higher depths. Fig. 2(g) and 2(h) show the absolute differences of CU depth between the first frame and the third frame under the same QP. Fig. 2(i) and 2(j) show the absolute differences of CU depth between the second frame and the third frame under different QPs. In Fig. 2(g), 2(h), 2(i) and 2(j), four colors including white, green, blue and red denote that the absolute difference values are zero, one, two and three, respectively. We further observe that most background regions are marked with white or green, which means that there are high CU depth correlations among neighboring frames. However, for the successive frames encoded with different QP, their CU depths are different in some medium or fast motion regions. This implies that QP has some influence on CU depth decision as well. To model the temporal correlation of CU depth, we define the probability of CU depth correlation as follows.

$$P_c = \frac{N_c}{\sum_{c=0}^{3} N_c} \tag{3}$$

$$c = \left| d_{cur} - d_{pre} \right|, c \in \{0, 1, 2, 3\} \tag{4}$$

where d_{cur} and d_{pre} are the depths of the current CU and its co-located CU in previous frame, respectively. c is the absolute difference of CU depth and N_c is the number of c. Thus, P_c is the probability of absolute depth difference.

Table III reports the results of CU depth correlation. From it, we observe that the depth of co-located CU provides a good prediction reference for current CU. When c equals 0, it means that the current CU has the same optimal depth with its co-located CU. The average probability is 76.34%, and the probabilities of depth correlation do not exhibit great difference among video sequences with diverse motions, even when the QPs are changed. Meanwhile, when c is equal to 3, it means that the current CU has the weakest depth correlation with its co-located CUs. Actually, the average probability is only 0.23%. In summary, there is a strong correlation of optimal depth among the current CU and its temporally co-located CUs, simply because temporally neighboring CUs usually

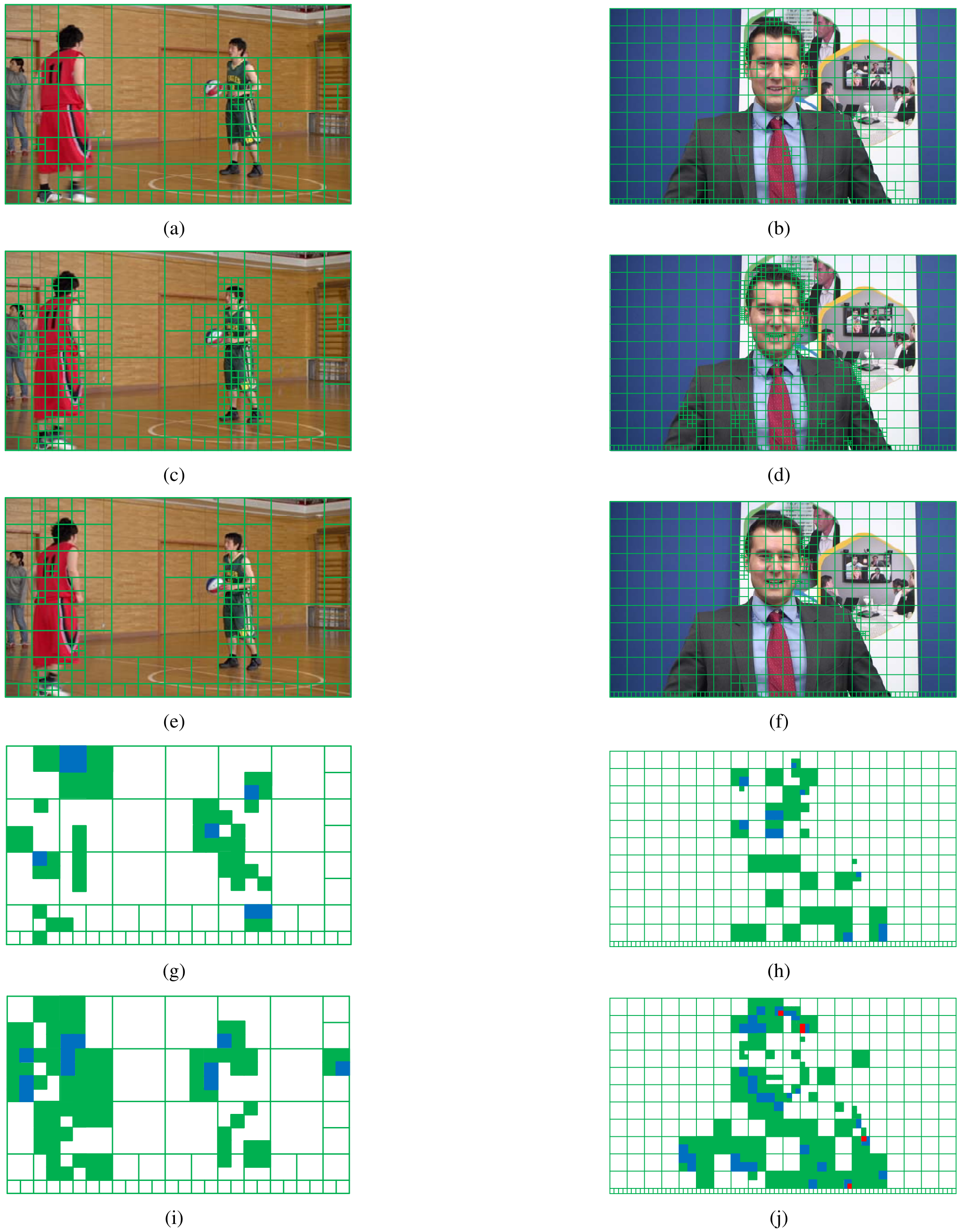


Fig. 2. Three successive frames of *Basketballpass* and *Johnny* sequences and their optimal CU depths and the absolute differences of depth among frames. (a) and (b) are the 8^{th} fame with QP 35, (c) and (d) are the 9^{th} fame with QP 33, (e) and (f) are the 10^{th} fame with QP 35, (g) and (h) are the CU depth absolute difference between the 8^{th} frame and the 10^{th} frame, (i) and (j) are the CU depth absolute difference between the 9^{th} frame and the 10^{th} frame.

share similar motion and textures in natural videos. That is, the absolute differences of CU depth are usually small for temporally neighboring CUs. Moreover, with the increase of the absolute difference of CU depth, the probability dramatically decreases as well.

III. Proposed Adaptive CU Depth Decision Method

A. Optimal Selection Model of CU Depth (OSMCD)

The depth of temporally co-located CU may be directly exploited to predict the depth of current CU. Motivated by this observation, we firstly establish an optimal selection model of CU depth (OSMCD). Then, an OSMCD-based adaptive CU depth decision approach is proposed.

Fig. 3 shows the correlation between the current CU and its temporally co-located CUs. t_{cur} is the interval of the current CU. Co_CU_0 and Co_CU_1 are two temporally co-located CUs in the intervals t_{pre0} and t_{pre1}, respectively. OSMCD is established by exploiting the temporal correlation of CU depth. Specifically, the depths of two co-located CUs in the previously encoded frames are adopted to estimate the depth for current CU. From Table III, the sum of P_0 and P_1 is 96.9% on average, which means that there is only small depth

TABLE III
CU DEPTH CORRELATION IN THE TEMPORAL (UNIT:%)

Sequences	QP	c=0	c=1	c=2	c=3	$c\leq 1$
Basketballpass (416×240)	24	74.05	22.74	2.96	0.25	96.79
	28	74.35	22.14	3.25	0.26	96.49
	32	75.06	21.52	3.19	0.23	96.58
	36	74.59	21.69	3.45	0.27	96.28
	Average	74.51	22.02	3.21	0.26	96.53
Traffic (2560×1600)	24	60.77	33.81	5.08	0.34	94.58
	28	68.12	27.40	4.15	0.33	95.52
	32	73.18	22.88	3.62	0.32	96.06
	36	76.82	20.08	2.85	0.26	96.90
	Average	69.72	26.04	3.93	0.31	95.76
Johnny (1280×720)	24	77.86	19.54	2.48	0.11	97.40
	28	83.51	14.87	1.51	0.11	98.38
	32	87.33	11.41	1.15	0.10	98.74
	36	90.49	8.58	0.83	0.10	99.07
	Average	84.80	13.60	1.49	0.11	98.40
Average		76.34	20.55	2.88	0.23	**96.90**

variation between the current CU and its temporally co-located CU. That is, most CUs′ depths exhibit strong correlation with their co-located CUs in the intervals t_{pre0} and t_{pre1}. As shown in Fig. 3, the depths of co-located CUs are used to predict the depth of current CU, which is defined as follows.

$$D_{cur} = D_{pre0} + \Delta D \tag{5}$$

where D_{cur} is the depth of the current CU, D_{pre0} is the maximum depth of the co-located CU in the interval t_{pre0}, and ΔD is the depth variation between the current CU and its co-located CU in the intervals t_{pre0}. As shown in Fig. 2(g), 2(h), 2(i) and 2(j), the depth differences are different among successive frames under different QPs. Therefore, we should further consider the fact that the optimal CU depth exhibits a strong dependence on the QP. Thus, ΔD is defined as

$$\begin{aligned}\Delta D = &\left|D_{pre0} - D_{pre1}\right| + \left|QP_{cur} - QP_{pre0}\right| \\ &- \left|QP_{pre0} - QP_{pre1}\right|\end{aligned} \tag{6}$$

where D_{pre1} is the maximum depth of co-located CU in the interval t_{pre1}. QP_{cur}, QP_{pre0} and QP_{pre1} are the QPs in the intervals t_{cur}, t_{pre0} and t_{pre1}, respectively. In this paper, the nearest neighboring frames of the current frame are used to compute ΔD, in which QP_{pre0} is equal to the QP of the current frame and QP_{pre1} is equal to QP-1 of the current frame. Especially, if the encoded frame with QP-1 does not exist, the encoded frame with QP+1 is chosen as the nearest neighboring frame. Thus, equation (6) is rewritten as

$$\Delta D = \left|D_{pre0} - D_{pre1}\right| - 1 \tag{7}$$

The variation of CU depth, i.e., ΔD, may also be affected by motion characteristic. Especially for the CUs in medium or fast motion regions, ΔD is larger than those CUs in motion-less regions. Thus, a weighting factor is introduced into equation (5), which is rewritten as

$$D_{cur} = D_{pre0} + w \cdot |\Delta D| \tag{8}$$

where w is the weighting factor. Since the CU depth correlation is related with the locations of neighboring frames, w is defined in terms of the encoding time correlation between the current CU and the adjacently encoded CUs. That is,

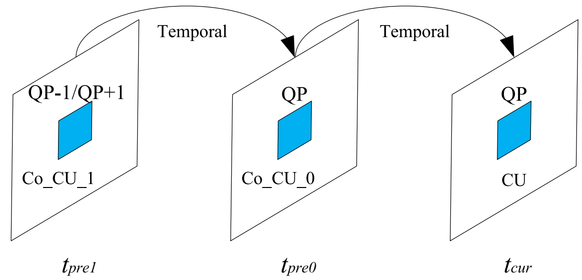


Fig. 3. Temporal correlation between current CU and its co-located CUs.

$$w = \frac{1}{t_{cur} - min\left\{t_{pre0}, t_{pre1}\right\}} \times \left|\frac{t_{cur} - t_{pre0}}{t_{pre1} - t_{pre0}}\right| \tag{9}$$

where t_{cur}, t_{pre0} and t_{pre1} are the current interval and two previously encoded intervals, respectively. In this study, the basic unit of intervals is assumed to be 1. From the above analysis, the OSMCD is established to predict the depth of the current CU as follows.

$$D_{pre0} \leq D_{cur} \leq D_{pre0} + w \cdot |\Delta D|. \tag{10}$$

B. Early Skipping Lower CU Depth With OSMCD

In equation (10), the minimum depth of the current CU can be predicted from D_{pre0}, and the maximum depth of the current CU can be predicted from D_{pre0} and D_{pre1}. However, we know from Table III that there is up to 76.34% CUs whose depth is equal to D_{pre0}, which is the depth of the co-located CU. That is, there is still CUs (23.66%) left whose depth is quite different from D_{pre0}. To reduce the error prediction rate of the depth, OSMCD is further adjusted as follows.

$$min\left\{D_{pre0}, \left|D_{pre0} - D_{th}\right|\right\} \leq D_{cur} \leq D_{pre0} + w \cdot |\Delta D| \tag{11}$$

where D_{th} is the threshold for depth adjustment. In equation (11), the depth of the current CU is predicted within a range of "$min\{D_{pre0}, |D_{pre0} - D_{th}|\}$" and "$D_{pre0} + w \cdot \Delta D$". If the depth of the current encoding CU is estimated by equation (11), some unnecessary checking of candidate depths can be avoided. In this paper, we only skip depth levels "0" and "1" to balance between visual quality and encoding time saving. The skipping of depth levels "0" and "1" is expressed as

$$D^0_{Skip} = \begin{cases} \text{skip}, & min\left\{D^0_{pre0}, \left|D^0_{pre0} - D_{th}\right|\right\} \geq 1 \\ \text{no\_skip}, & \text{else} \end{cases} \tag{12}$$

$$D^1_{Skip} = \begin{cases} \text{skip}, & min\left\{D^1_{pre0}, \left|D^1_{pre0} - D_{th}\right|\right\} \geq 2 \\ \text{no\_skip}, & \text{else} \end{cases} \tag{13}$$

where D^0_{pre0} is the maximum depth of CU with size 64×64, and D^1_{pre0} is the maximum depth of CU with size 32×32.

C. Early CU Depth Decision Based on OSMCD and CBF

Based on the analysis in Sections III-A and III-B, the maximum depth for the current CU is predicted by using the depths of previously encoded CUs. In this paper, the depth of

TABLE IV
THE PERCENTAGE OF CUs WITH CBF IS ZERO (UNIT:%)

	QP	Basket...	Traffic	Johnny	average
$P_{(CBF=0\mid D_{opt})}$	24	71.75	74.60	87.50	77.95
	28	78.96	84.20	92.01	85.06
	32	85.33	89.29	94.61	89.74
	36	89.39	92.09	95.96	92.48
	average	81.35	85.05	92.52	86.31

TABLE V
THREE DIFFERENT CONDITIONS FOR EARLY DETERMINATION CU DEPTH

	condition A	condition B	condition C
Parameters	Merge|Inter_2N×2N	CBF=0	(condition A) &&(condition B)

co-located CUs with size 64×64 is used to predicted the maximum depth of current CU to reduce prediction error. However, there may still exist some accumulated errors. The inherent reason is that if some depths are wrongly classified as the optimal depth, which might be further used to predict the depth of other CUs. That is, the depth prediction errors might be accumulated in this way. To address this issue, the CBF information of the current CU is further exploited to refine the prediction depth of the current CU. To investigate the correlation between the CBF and the optimal depth of CU, we defined a probability $P_{(CBF=0|D_{opt})}$ as follows.

$$P_{(CBF=0|D_{opt})} = \frac{N_{(CBF=0)}}{\frac{W}{8} \times \frac{H}{8}} \tag{14}$$

where W and H are the width and height of video frame, respectively. $N_{(CBF=0)}$ is the number of 8×8 encoded blocks, whose CBF is zero in the optimal depth. D_{opt} is the optimal depth among four candidate depth levels for current CU.

Three typical videos including *Basketballpass*, *Traffic* and *Johnny* are tested under the same conditions summarized in Table I. And Table IV reports the experimental results. From it, we observe that the probability $P_{(CBF=0|D_{opt})}$ is more than 71%, which indicates that for most encoded CUs with the optimal depth, their CBFs are zero. This phenomenon inspires us to infer that when a CU is to be encoded with the depth of "$D_{pre0} + w \cdot \Delta D$", which is estimated by equation (10), if its CBF is zero, the prediction depth "$D_{pre0} + w \cdot \Delta D$" is actually the optimal depth for the current CU, and thus the partitioning of the current CU can be early terminated. Otherwise, the maximum prediction depth of the current CU should be increased by 1 until the CBF is zero.

D. Early CU Depth Decision Based on Encoding Parameters

HEVC supports multiple CU sizes from 64 × 64 to 8 × 8, and the symmetric and asymmetric prediction units should be computed for each CU. For static background region and slow motion regions, it is more likely to choose large prediction mode as the optimal mode of CU. To early determine the CU depth, we define three conditions, which are summarized in Table V. Condition A represents that the optimal mode is either Merge or Inter_2N×2N (P_2N×2N) for the current CU, the current depth is the optimal depth and the current CU will not be split any more. Condition B refers to the fact that if the CBF of the current CU is zero, the current depth is also the optimal depth and the current CU will not be split. Condition C is the combination of both conditions A and B.

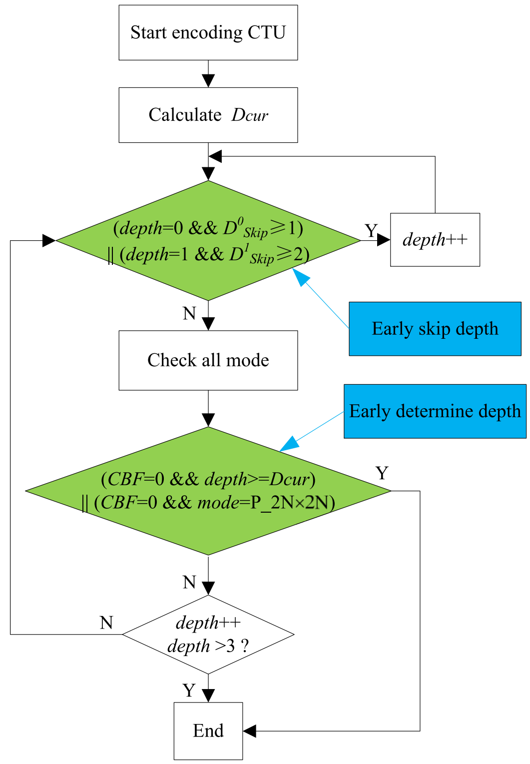


Fig. 4. Flowchart of the proposed algorithm.

To testify the false probability of condition A, B and C for early terminating CU depth decision, we define False Rate (FR) as follows.

$$FR_d = \frac{N_d}{\sum_{d=0}^{3} N_d} \tag{15}$$

$$d = \left| d_{opt} - d_{pre} \right|, d \in \{0, 1, 2, 3\} \tag{16}$$

where d_{opt} and d_{pre} are the optimal CU depth selected by the original HM12.0 and the prediction CU depth under the above early determination condition, respectively. d is the absolute difference between d_{opt} and d_{pre}, and N_d is the number of d. FR_d is the probability of false depth decision. Meanwhile, to evaluate the effectiveness of the early determination condition under different QPs (24, 28, 32, 36), the decrease of peak signal-to-noise ratio (ΔPSNR) and bit rate increase (ΔBR) are adopted as the metrics of RD performance, which are computed as

$$\Delta PSNR = PSNR_{Pre} - PSNR_{Ori} \tag{17}$$

$$\Delta BR = (BR_{Pre} - BR_{Ori})/BR_{Ori} \times 100\% \tag{18}$$

where $PSNR_{Pre}$, BR_{Pre} are PSNR and bit rate of early CU depth decision under early determination condition, respectively. And $PSNR_{Ori}$, BR_{Ori} are PSNR and bit rate of the original HM, respectively.

TABLE VI
THE PROBABILITY OF FALSE PARTITION RATE UNDER THE EARLY CU DEPTH DECISION CONDITION (UNIT:%)

Sequences	QP	Condition A			Condition B			Condition C		
		d=1	*d*=2	*d*=3	*d*=1	*d*=2	*d*=3	*d*=1	*d*=2	*d*=3
Basketballpass (416×240)	24	17.35	2.94	1.07	15.04	1.28	0.08	14.03	0.91	0.04
	28	16.11	3.02	0.76	15.31	1.56	0.13	13.47	1.08	0.05
	32	16.22	2.78	0.54	16.75	2.22	0.16	13.72	1.25	0.05
	36	16.65	2.96	0.37	18.05	2.66	0.18	14.96	1.71	0.08
	Average	16.58	2.93	0.69	16.29	1.93	0.14	15.05	1.24	0.06
Traffic (2560x1600)	24	28.17	6.54	1.25	26.50	3.65	0.25	25.41	2.92	0.15
	28	22.15	4.53	0.70	21.25	3.08	0.23	19.27	2.20	0.11
	32	18.41	3.40	0.44	19.03	2.89	0.24	15.99	1.95	0.12
	36	15.94	2.41	0.25	17.02	2.31	0.18	14.10	1.55	0.09
	Average	21.17	4.22	0.66	20.95	2.98	0.23	18.69	2.16	0.12
Johnny (1280×720)	24	12.05	3.18	0.81	10.07	1.02	0.06	9.75	0.88	0.04
	28	10.86	2.51	0.63	8.78	0.89	0.05	8.33	0.75	0.04
	32	9.61	2.16	0.47	8.13	0.91	0.06	7.35	0.69	0.03
	36	8.68	1.82	0.31	7.80	1.01	0.06	6.58	0.66	0.03
	Average	10.30	2.42	0.56	8.70	0.96	0.06	8.00	0.75	0.04
Average		16.02	3.19	0.63	15.31	1.96	0.14	13.58	1.38	0.07

TABLE VII
THE RD PERFORMANCE UNDER EARLY CU DEPTH DECISION CONDITION

Sequences	QP	Condition A		Condition B		Condition C	
		$\Delta PSNR$	ΔBR	$\Delta PSNR$	ΔBR	$\Delta PSNR$	ΔBR
Basketballpass (416×240)	24	-0.076	6.05	-0.046	-0.17	-0.013	-0.11
	28	-0.086	5.04	-0.072	-0.35	-0.012	-0.23
	32	-0.108	4.09	-0.169	-1.15	-0.046	-0.24
	36	-0.145	2.97	-0.212	-2.15	-0.061	-1.62
	Average	-0.104	4.54	-0.125	-0.95	-0.033	-0.55
Traffic (2560×1600)	24	-0.123	2.79	-0.071	-1.72	-0.033	-0.93
	28	-0.114	3.93	-0.099	-1.82	-0.034	-1.04
	32	-0.127	3.67	-0.143	-1.95	-0.061	-1.45
	36	-0.131	2.54	-0.150	-2.17	-0.067	-1.62
	Average	-0.123	3.23	-0.116	-1.92	-0.049	-1.26
Johnny (1280×720)	24	-0.061	0.09	-0.055	-2.33	-0.027	-1.49
	28	-0.066	1.37	-0.066	-1.21	-0.030	-0.94
	32	-0.066	2.24	-0.079	-1.94	-0.043	-1.76
	36	-0.083	1.72	-0.104	-1.19	-0.032	-0.76
	Average	-0.069	1.36	-0.076	-1.67	-0.033	-1.24
Average		-0.099	3.04	-0.106	-1.51	-0.038	-1.02

Table VI reports the false probabilities of CU partitioning. Table VII is the corresponding RD performance for the CU depth decision under early determination condition. We observe that FR_d is about 20% ($d \in \{1, 2, 3\}$) in Table VI. From Table VII, the PSNR decreases 0.099 dB and the bit rate increases 3.04% on average under the condition A. This means that there are still some CU depths which are not early determined effectively. For the condition B, we can observe that FR_d is also large. Especially for *Basketballpass* sequence with fast motion, the PSNR decreases 0.125 dB, which means that the condition B is not effective for videos with fast motion. In addition, as shown in Table VI, the FR_d under the condition C is less than that under condition A and B, respectively. And the $\Delta PSNR$ and ΔBR in Table VI is negligible compared with those under condition A and B. In summary, the FR_d under condition A, B and C can be represented as follows.

$$P\left(FR_d \middle| C\right) < P\left(FR_d \middle| B\right) < P\left(FR_d \middle| A\right), d \in \{1, 2, 3\} \quad (19)$$

For different d values ($d \in \{1, 2, 3\}$), equation (19) still holds. From the RD performance under condition C, which is reported in Table VII, we observe that both $\Delta PSNR$ and ΔBR are negligible for video sequences with various motion activities. Thus, condition C is more preferable for early CU depth decision.

E. Proposed Overall Algorithm

The basic idea behind the proposed fast CU depth decision approach is to adaptively adjust the range of CU depth prediction by simultaneously utilizing both the depth correlation among temporally neighboring CUs and the encoding parameters of CU. Fig. 4 is the flowchart, and the procedure is summarized as follows:

1) Start inter CU depth prediction.
2) Compute the adaptive minimum and maximum CU depths using equations (11), (12) and (13). For the CU with size 64×64, if the minimum depth is larger or equal to 1, the depth level “0” is skipped. For the CU with size 32×32, if the minimum depth is larger or equal to 2, the depth level “1” is skipped.
3) Calculate each prediction mode for the current CU in the RDO process and derive the coding parameters CBF and the mode of prediction unit. If CBF is zero, and the prediction mode is P_2N×2N or the CU depth is larger

TABLE VIII
PERFORMANCE OF PROPOSED ADAPTIVE CU DEPTH DECISION UNDER RA TEST CONDITION

Resolution	Sequences	Case1		Case2		Case3		Case4		Overall Proposed	
		BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A (2560×1600)	PeopleOnstreet	0.7	27.4	1.0	24.2	1.0	41.1	0.7	21.4	1.4	47.5
	Traffic	1.1	12.5	1.0	55.0	1.1	57.0	0.7	52.8	1.5	63.8
Class B (1920×1080)	Kimono1	1.0	8.3	0.6	45.5	1.1	46.4	0.5	43.2	1.3	51.4
	ParkScene	1.0	13.8	0.7	50.8	0.9	52.4	0.6	49.0	1.2	61.2
	Cactus	0.9	13.0	1.1	45.3	1.3	49.8	0.8	43.1	1.7	55.3
	BasketballDrive	1.7	12.2	0.7	42.6	1.8	47.2	0.5	40.1	1.9	56.1
	BQTerrace	1.0	12.4	1.1	53.0	1.2	55.6	0.8	55.9	1.5	64.1
Class C (832×480)	BasketballDrill	1.6	17.6	0.6	34.3	1.6	42.3	0.3	32.3	1.8	48.1
	BQMall	1.2	16.6	1.1	43.2	1.4	49.1	0.8	40.8	1.8	56.2
	PartScene	0.5	20.9	0.9	35.4	0.7	45.2	0.7	33.5	1.1	52.4
	RaceHorseC	1.7	22.2	1.1	26.2	1.5	38.6	0.5	23.2	1.7	44.7
Class D (416×240)	BasketballPass	0.6	22.3	0.9	27.9	0.9	39.5	0.5	25.7	1.2	46.4
	BQSquare	0.4	14.3	0.7	46.3	0.4	49.5	0.5	43.9	0.7	57.5
	Blowingbubbles	0.4	18.7	0.9	36.2	0.5	43.3	0.8	35.0	0.9	51.8
	RaceHorses	0.7	25.2	1.0	22.6	1.0	37.9	0.6	20.0	1.3	44.0
Class E (1280×720)	FourPeople	0.7	6.5	0.4	65.2	0.7	66.0	0.2	64.1	0.8	69.9
	Johnny	0.7	3.1	0.2	71.1	0.4	70.8	0.1	69.7	0.4	73.1
	KristenAndSara	0.8	4.5	0.4	67.3	0.7	67.3	0.2	65.9	0.8	70.2
	Average	0.9	15.1	0.8	44.0	1.0	50.0	0.5	42.2	1.3	56.3

or equal to the maximum prediction depth, the current CU depth is selected as the optimal depth. Otherwise, the maximum prediction depth will be increased by 1 until the CBF of CU is zero.

4) Encode the next CTU.

IV. EXPERIMENTAL RESULTS AND ANALYSES

A. Test Conditions

To verify the validity of the proposed CU depth decision approach, HM12.0 and the common test conditions [31] of HEVC standardization are adopted under random access (RA) and low-delay (LD) B configurations, respectively. The search range is 64 in both horizontal and vertical directions, the size of CTU is 64×64 and correspondingly the maximum CU depth is 4. QP is set with 22, 27, 32 and 37, respectively. Fast encoder decision, fast decision for Merge, and transform skip are enabled. For performance evaluation metrics, the BDBR defined in [32] and the encoding time reduction are adopted. The encoding time reduction is defined as follows.

$$TS = \frac{T_{HM} - T_p}{T_{HM}} \times 100\% \tag{20}$$

where T_p and T_{HM} are the total encoding time of the proposed approach and the original HM encoder, respectively. Since we observe from Table III that the probability of absolute depth difference less than or equals 1 reaches up to 96.9% on average, D_{th} in equations (12) and (13) are set with 1.

B. Experimental Results

Table VIII reports the experimental results of the proposed approach. To observe the detailed contributions, the results are individually measured under five cases: 1) the early skipping lower CU depth method in Section III-B; 2) the early CU depth determination in Section III-C and Section III-D; 3) Joint early skipping lower CU depth and early CU depth determination in Sections III-B and III-C; 4) the early CU depth determination in Section III-D; 5) the overall adaptive CU depth decision approach. From Table VIII, the four methods under the cases 1), 2), 3) and 4) effectively reduce the encoding time with acceptable BDBR increase when they are compared with the original HM12.0. Specifically, case 1) achieves 15.1% encoding time reduction with 0.9% BDBR increase on average. For videos with fast motion, it reduces more encoding time with negligible BDBR increase compared with slow motion videos. The reason behind this is that for video sequences with fast motion, there are more motion-complex regions, which are more likely to choose high CU depth as the optimal depth. That is, the lower depths can be early skipped. However, videos with slow motion have much more static background or homogeneous regions, which are very likely to select low depth as the optimal CU depth. For example, the total encoding time saving is only 3.1% for *Johnny* sequence with slow motion. That is, the early skipping CU depth approach under case 1) can efficiently reduce the encoding time for videos with complicated motion.

From Table VIII, we can also observe that under case 2), the early CU depth decision method achieves 44% encoding time saving with 0.8% BDBR increase on average, respectively. Moreover, it reduces more encoding time for slow motion videos than for video sequences with medium or fast motion. Actually, for videos with medium or fast motion, higher CU depth is more likely to be selected as the optimal depth. This means that it can not early determine the CU depth, and thus the encoding time reduction is limited. For the case 3), the OSMCD-based CU depth decision approach saves about 50% encoding time with only 1% BDBR increase. For the case 4), the early CU depth decision method which exploits the encoding parameters saves 42.4% encoding time with 0.5% BDBR increase. For the overall adaptive CU depth decision method (under case 5), which combines two methods in the cases 1) and 2), achieves 56.3% encoding time reduction with 1.3% BDBR increase on average. Apparently, the combined method reduces more encoding time for videos with various motions, while maintaining acceptable loss of RD performance as the original HM encoder. This indicates that the proposed

TABLE IX
PERFORMANCE COMPARISON OF PROPOSED METHOD WITH RECENT WORKS FOR RA CONDITION

Resolution	sequences	ShenTCSVT [29]		ShenTMM [20]		AhnTCSVT [18]		ECU+CBF+ESD		Proposed	
		BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A (2560x1600)	PeopleOnstreet	0.2	42.5	4.0	23.3	0.9	26.9	2.4	32.5	1.4	47.5
	Traffic	1.1	60.5	2.1	44.1	0.8	61.6	2.4	64.4	1.5	63.8
Class B (1920x1080)	Kimono1	1.0	47.3	0.4	31.2	1.3	58.2	1.5	52.3	1.3	51.4
	ParkScene	0.9	45.2	1.0	33.7	1.2	52.6	1.9	60.2	1.2	61.2
	Cactus	1.0	42.1	3.2	40.6	2.8	56.8	2.6	53.4	1.7	55.3
	BasketballDrive	1.0	41.8	1.4	28.7	2.0	50.9	1.5	48.9	1.9	56.1
	BQTerrace	1.1	49.7	1.2	35.7	1.6	54.4	2.4	62.0	1.5	64.1
Class C (832x480)	BasketballDrill	2.1	39.8	5.2	34.1	1.9	45.2	1.3	42.2	1.8	48.1
	BQMall	1.6	40.0	2.9	28.8	2.2	48.6	2.6	52.7	1.8	56.2
	PartyScene	1.1	45.8	2.5	27.0	0.8	37.7	2.0	44.5	1.1	52.4
	RaceHorseC	1.8	38.5	2.0	17.7	2.2	33.9	2.0	31.4	1.7	44.7
Class D (416x240)	BasketballPass	1.8	28.9	2.2	16.7	1.5	33.6	2.1	36.1	1.2	46.4
	BQSquare	0.4	34.4	0.4	22.9	0.6	45.1	1.8	57.3	0.7	57.5
	Blowingbubbles	1.3	33.7	2.6	23.4	0.7	38.2	2.0	46.7	0.9	51.8
	RaceHorses	1.8	24.4	2.1	12.8	1.1	26.6	2.5	30.3	1.3	44.0
Class E (1280x720)	FourPeople	1.4	66.3	2.9	56.9	1.7	74.1	0.9	75.9	0.8	69.9
	Johnny	0.9	67.8	1.0	57.8	1.3	75.7	0.6	81.1	0.4	73.1
	KristenAndSara	1.3	62.5	2.7	51.8	1.2	73.1	0.9	76.8	0.8	70.2
	Average	1.2	45.1	2.2	32.6	1.4	49.6	1.9	52.7	1.3	56.3

TABLE X
PERFORMANCE COMPARISON OF PROPOSED METHOD WITH RECENT WORKS FOR LD B CONDITION

Resolution	sequences	ShenTCSVT [29]		ShenTMM [20]		AhnTCSVT [18]		ECU+CBF+ESD		Proposed	
		BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A (2560x1600)	PeopleOnstreet	0.4	32.5	2.2	19.9	1.5	28.0	1.3	25.9	0.7	45.0
	Traffic	1.0	54.7	2.0	41.3	2.0	41.7	1.5	56.5	1.3	59.9
Class B (1920x1080)	Kimono1	0.8	38.0	0.3	27.8	0.8	47.9	1.2	43.4	1.0	44.3
	ParkScene	0.8	40.5	0.9	28.9	1.1	48.3	1.5	51.2	1.3	56.5
	Cactus	0.5	43.5	3.1	36.4	2.2	48.0	1.8	45.5	1.3	49.6
	BasketballDrive	0.8	42.3	1.0	25.7	1.2	42.6	1.0	42.1	1.5	46.7
	BQTerrace	1.3	42.7	1.8	32.2	0.3	47.5	1.3	54.4	1.5	58.0
Class C (832x480)	BasketballDrill	1.9	44.0	1.3	38.5	2.0	37.0	1.0	35.2	1.4	43.9
	BQMall	2.0	43.5	2.2	25.3	1.2	40.9	1.5	43.4	1.8	51.8
	PartyScene	1.0	40.0	3.3	25.1	0.4	33.0	1.1	33.6	0.6	46.4
	RaceHorseC	1.0	34.1	1.1	14.0	1.1	26.1	1.0	25.9	1.2	41.4
Class D (416x240)	BasketballPass	1.5	35.4	1.6	14.8	1.2	27.2	1.1	30.0	0.9	44.1
	BQSquare	0.4	36.0	0.4	44.3	0.2	38.8	1.2	42.3	0.3	50.0
	Blowingbubbles	2.2	39.7	2.4	22.0	0.3	31.5	1.4	35.5	0.6	45.4
	RaceHorses	0.7	30.8	1.0	12.9	0.7	21.2	1.1	23.5	0.7	41.3
Class E (1280x720)	FourPeople	1.1	64.0	2.4	47.8	1.6	65.6	1.3	70.3	1.4	66.3
	Johnny	1.3	66.9	2.0	56.0	-0.3	73.9	1.3	77.2	1.1	70.3
	KristenAndSara	1.4	60.9	2.8	48.5	0.5	69.6	1.1	70.8	0.9	65.6
	Average	1.1	43.9	1.8	31.2	1.0	42.7	1.3	44.8	1.1	51.5

adaptive CU depth decision algorithm can early skip low CU depths for complex-motion videos and early determine high CU depth for slow motion videos, respectively.

C. Comparisons With the State-of-the-Art Methods

To make full comparisons, the state-of-the-art fast CU depth decision algorithms including ShenTCSVT [29], ShenTMM [20] and AhnTCSVT [18] are selected as benchmarks. Besides, the fast algorithm which combines ECU [14], CBF [15] and ESD [30] is also selected for comparisons because it is adopted by the original HM reference software. Comparisons are made under the RA and LD B conditions, respectively. Table IX and X reports the results of the proposed approach and four benchmarks. Please note that all these fast algorithms are compared with the original HM12.0 in terms of BDBR and encoding time saving, respectively. From Table IX, we know that compared with the existing fast algorithms, the proposed approach achieves more encoding time saving for fast motion videos such as *RaceHorses*. About 13%-31% encoding time is further reduced while maintaining negligible loss of RD performance. It mainly benefits from the fact that the proposed method can effectively skip unnecessary lower CU depths for fast motion videos. Moreover, compared with ShenTCSVT [29], ShenTMM [20], AhnTCSVT [18] and ECU+CBF+ESD [14], [15], [30], the proposed algorithm reduces more encoding time for about 11.2%, 23.7%, 6.7% and 3.6%, respectively. Meanwhile, it achieves similar BDBR increase with them. From Table X, the proposed CU depth decision approach also outperforms the existing methods in terms of total encoding time saving under LD B condition. Therefore, the proposed approach, which exploits OSMCD and encoding parameters, is efficient for videos with various motion and achieves desirable RD performance. In Table XI, the proposed method is further compared with the state-of-the-art fast CU depth decision

TABLE XI
PERFORMANCE COMPARISON OF PROPOSED METHOD WITH RECENT WORKS FOR LD B CONDITION

Resolution	sequences	ShenEVIP [27]		XiongTMM [26]		ZhangTIP [24]		Proposed	
		BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Class A (2560×1600)	PeopleOnstreet	5.0	26.5	11.0	32.0	1.2	44.4	0.7	45.0
	Traffic	4.9	52.4	6.4	46.6	2.0	56.1	1.3	59.9
Class B (1920×1080)	Kimono1	4.8	52.4	6.4	46.6	2.0	56.1	1.0	44.3
	ParkScene	4.5	49.5	5.3	42.3	1.5	47.8	1.3	56.5
	Cactus	5.2	49.0	11.9	52.0	1.8	48.4	1.3	49.6
	BasketballDrive	3.0	48.4	10.1	52.1	2.2	51.0	1.5	46.7
	BQTerrace	2.8	50.9	6.0	40.2	2.1	45.1	1.5	58.0
Class C (832×480)	BasketballDrill	4.6	39.5	7.1	38.9	1.7	47.6	1.4	43.9
	BQMall	5.3	41.6	6.0	37.1	1.6	46.0	1.8	51.8
	PartyScene	5.4	36.0	5.5	32.6	1.3	36.6	0.6	46.4
	Mobisode2	4.8	62.4	9.1	53.8	3.7	62.9	2.0	54.9
Class D (416×240)	BasketballPass	5.2	28.4	2.6	40.9	1.2	36.6	0.9	44.1
	BQSquare	4.6	40.4	3.2	29.5	0.8	30.7	0.3	50.0
	Blowingbubbles	5.0	34.5	4.6	26.7	0.9	28.8	0.6	45.4
	RaceHorses	5.5	22.5	4.0	30.4	1.2	33.4	0.7	41.3
Class E (1280×720)	FourPeople	5.4	64.8	11.7	58.4	2.8	66.4	1.4	66.3
	Johnny	4.4	70.3	8.9	61.1	2.5	70.9	1.1	70.3
	KristenAndSara	4.1	67.4	7.1	57.6	2.3	68.2	0.9	65.6
	Vidyo1	4.3	65.5	8.6	60.2	2.7	68.2	1.5	65.8
	Vidyo3	6.0	65.6	7.6	58.9	3.2	67.5	2.2	65.8
	Vidyo4	4.6	66.6	6.4	59.2	2.9	67.8	1.5	65.6
	Average	4.7	49.3	7.1	45.6	2.0	51.5	1.2	54.2

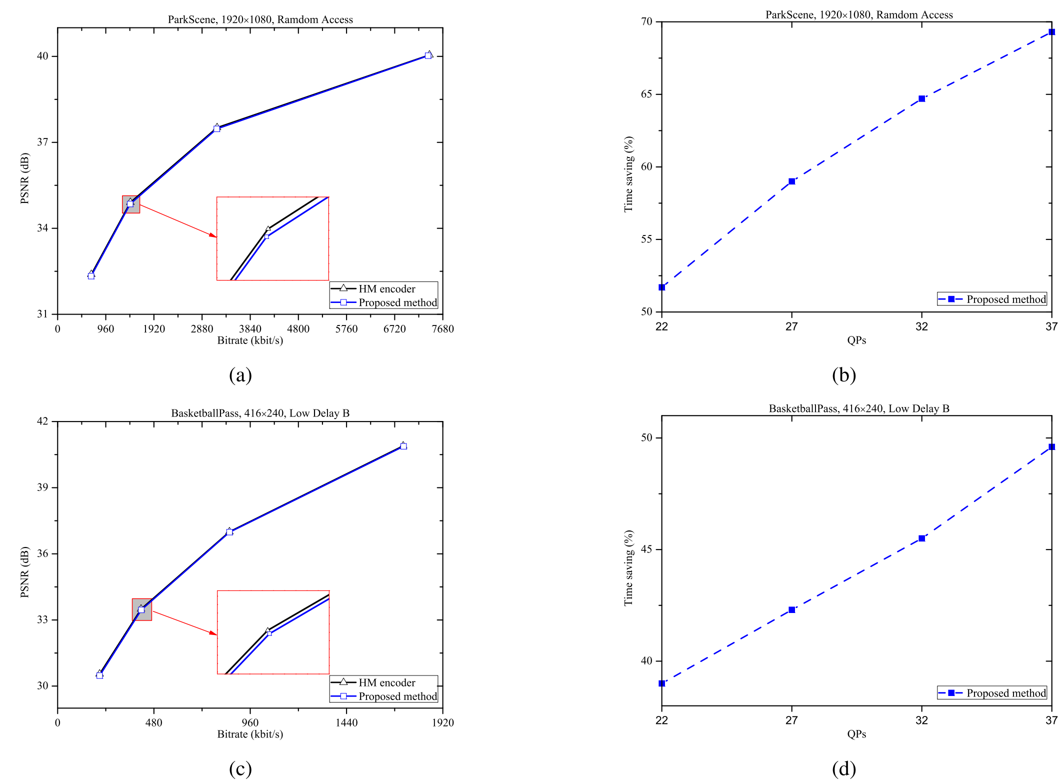


Fig. 5. Performance of *ParkScene* (1920×1080, 24Hz) and *BasketballPass* (416×240, 50Hz) under different QPs (22, 27, 32, 37). (a) RD curves of *ParkScene*. (b) Time saving of *ParkScene*. (c) RD curves of *BasketballPass*. (d) Time saving of *BasketballPass*.

methods [24], [26], [27]. Since ZhangTIP [24] only provides the results under LD B test condition, comparisons are made under the same configurations with [24]. From Table XI, we can also observe that the proposed method achieves more time saving while maintaining limited BDBR increase. Fig. 5 shows the results of the proposed method under different QPs (22, 27, 32, 37) for two typical video sequences including *ParkScene* and *BasketballPass*. It still achieves more encoding time saving as the QP increases, with almost no bitrate increase and no PSNR degradation.

D. Performance of the Proposed Method for Video Sequences With Scene Change

Table XII summarizes five video sequences with scene change, which are obtained by cascading two videos of

TABLE XII
CHARACTERISTICS OF THE SCENE CHANGE SEQUENCES

Sequences	Resolution	Frame rate	Encoding frames
BasketballPass_BlowingBubbles	416×240	50	960
PartyScene_BasketballDrill	832×480	50	960
Johnny_FourPeople	1280×720	60	1200
ParkScene_Kimono1	1920×1080	24	480
Traffic_PeopleOnStreet	2560×1600	30	300

TABLE XIII
PERFORMANCE OF PROPOSED METHOD FOR SCENES CHANGE SEQUENCES

Sequences	RA		LD B	
	BDBR (%)	TS (%)	BDBR (%)	TS (%)
BasketballPass_BlowingBubbles	1.2	47.5	0.7	43.8
PartyScene_BasketballDrill	1.4	48.7	1.1	44.1
Johnny_FourPeople	1.1	67.8	1.2	54.4
ParkScene_Kimono1	1.6	54.7	1.2	49.1
Traffic_PeopleOnStreet	1.6	52.0	1.0	49.0
Average	1.4	54.1	1.1	48.1

different types. For them, there is a scene change every 30 frames. Table XIII reports the experimental results. We observe that for video sequences with scene change, the proposed approach saves about 54% and 48% encoding time with only 1.4% and 1.1% BDBR increases under RA and LD B configurations, respectively. Thus, the proposed method has good robustness because it still efficiently reduces the total encoding time with negligible RD degradation.

V. CONCLUSION

In this paper, an adaptive CU depth decision approach is proposed to reduce the computational complexity of HEVC encoder by exploiting OSMCD and encoding parameters. Firstly, the OSMCD model is established based on the temporal correlation of CU depth to estimate the depth range for the current CU. Secondly, the encoding parameters including PU size and CBF are further exploited for fast CU depth decision. Experimental results show that the proposed approach saves 56.3% and 51.5% encoding time with only 1.3% and 1.1% BDBR increases under RA and LD B conditions, respectively. It outperforms the existing approaches. In future research, we will investigate the pipeline structure of HEVC encoder for its implementation on graphical processing unit (GPU) and other fast techniques for HEVC encoder [33], [34], multiview video coding [35], [36] and related security issues [37], [38].

REFERENCES

[1] G. J. Sullivan, J. 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.

[2] T. Wiegand, G. J. Sullivan, G. Bjontegaard, 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.

[3] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—Including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[4] C. Zhu, X. Lin, and L.-P. Chau, "Hexagon-based search pattern for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 5, pp. 349–355, May 2002.

[5] Z. Pan, J. Lei, Y. Zhang, X. Sun, and S. Kwong, "Fast motion estimation based on content property for low-complexity H.265/HEVC encoder," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 675–684, Sep. 2016.

[6] S. Cho and M. Kim, "Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1555–1564, Sep. 2013.

[7] B. Min and R. C. C. Cheung, "A fast CU size decision algorithm for the HEVC intra encoder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 5, pp. 892–896, May 2015.

[8] Y. Zhang *et al.*, "Low complexity HEVC INTRA coding for high-quality mobile video communication," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1492–1504, Dec. 2015.

[9] A. Öztekin and E. Erçelebi, "An early split and skip algorithm for fast intra CU selection in HEVC," *J. Real-Time Image Process.*, vol. 12, no. 2, pp. 273–283, Aug. 2016.

[10] L. Shen, Z. Zhang, and Z. Liu, "Effective CU size decision for HEVC intracoding," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4232–4241, Oct. 2014.

[11] H. Zhang and Z. Ma, "Fast intra mode decision for high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 660–668, Apr. 2014.

[12] S. J. Park, "CU encoding depth prediction, early CU splitting termination and fast mode decision for fast HEVC intra-coding," *Signal Process. Image Commun.*, vol. 42, pp. 79–89, Mar. 2016.

[13] L. Shen, Z. Zhang, and P. An, "Fast CU size decision and mode decision algorithm for HEVC intra coding," *IEEE Trans. Consum. Electron.*, vol. 59, no. 1, pp. 207–213, Feb. 2013.

[14] R.-H. Gweon, Y. Lee, and J. Lim, *Early Termination of CU Encoding to Reduce HEVC Complexity*, JCT-VC document JCTVC-F045, JCTVC, Turin, Italy, Jul. 2011.

[15] K. Choi, S. H. Park, and E. S. Jang, *Coding Tree Pruning Based CU Early Termination*, JCT-VC document JCTVC-F092, JCTVC, Turin, Italy, Jul. 2011.

[16] K. Goswami, B.-G. Kim, D.-S. Jun, S.-H. Jung, and J. S. Choi, "Early coding unit-splitting termination algorithm for high efficiency video coding (HEVC)," *EURASIP J. Image Video Process.*, vol. 36, no. 3, pp. 407–417, Jun. 2014.

[17] H. L. Tan, C. C. Ko, and S. Rahardja, "Fast coding quad-tree decisions using prediction residuals statistics for high efficiency video coding (HEVC)," *IEEE Trans. Broadcast.*, vol. 62, no. 1, pp. 128–133, Mar. 2016.

[18] S. Ahn, B. Lee, and M. Kim, "A novel fast CU encoding scheme based on spatiotemporal encoding parameters for HEVC inter coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 422–435, Mar. 2015.

[19] J. Xiong, H. Li, F. Meng, Q. Wu, and K. N. Ngan, "Fast HEVC inter CU decision based on latent SAD estimation," *IEEE Trans. Multimedia*, vol. 17, no. 12, pp. 2147–2159, Dec. 2015.

[20] L. Shen, Z. Liu, X. Zhang, W. Zhao, and Z. Zhang, "An effective CU size decision method for HEVC encoders," *IEEE Trans. Multimedia*, vol. 15, no. 2, pp. 465–470, Feb. 2013.

[21] Z. Pan, S. Kwong, Y. Zhang, J. Lei, and H. Yuan, "Fast coding tree unit depth decision for high efficiency video coding," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 3214–3218.

[22] J. Lee, S. Kim, K. Lim, and S. Lee, "A fast CU size decision algorithm for HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 411–421, Mar. 2015.

[23] 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.

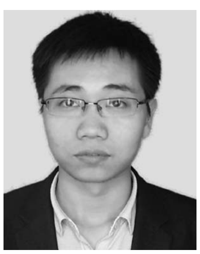
[24] Y. Zhang *et al.*, "Machine learning-based coding unit depth decisions for flexible complexity allocation in high efficiency video coding," *IEEE Trans. Image Process.*, vol. 24, no. 7, pp. 2225–2238, Jul. 2015.

[25] J. Xiong *et al.*, "MRF-based fast HEVC inter CU decision with the variance of absolute differences," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2141–2153, Dec. 2014.

[26] J. Xiong, H. Li, Q. Wu, and F. Meng, "A fast HEVC inter CU selection method based on pyramid motion divergence," *IEEE Trans. Multimedia*, vol. 16, no. 2, pp. 559–564, Feb. 2014.

[27] X. Shen and L. Yu, "CU splitting early termination based on weighted SVM," *EURASIP J. Image Video Process.*, vol. 2013, no. 4, pp. 1–11, Jan. 2013.

[28] 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.
[29] L. Shen, Z. Zhang, and Z. Liu, "Adaptive inter-mode decision for HEVC jointly utilizing inter-level and spatiotemporal correlations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 10, pp. 1709–1722, Oct. 2014.
[30] J. Yang, J. Kim, K. Won, and B. Jeon, *Early Skip Detection for HEVC*, JCT-VC document JCTVC-G543, JCTVC, Geneva, Switzerland, Nov. 2011.
[31] F. Bossen, *Common Test Conditions and Software Reference Configurations*, document Rec. JCTVC-L1100, Joint Collaborative Team Video Coding, Geneva, Switzerland, Jan. 2013.
[32] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," document VCEG-M33 ITU-T Q6/16, ITU, Austin, TX, USA, Apr. 2001.
[33] Z. Pan *et al.*, "Fast reference frame selection based on content similarity for low complexity HEVC encoder," *J. Vis. Commun. Image R.*, vol. 40, pp. 516–524, Oct. 2016.
[34] Y. Li, G. B. Yang, Y. P. Zhu, X. L. Ding, and X. M. Sun, "Unimodal stopping model based early SKIP mode decision for high efficiency video coding," *IEEE Trans. Multimedia*, to be published, doi: 10.1109/TMM.2017.2669863.
[35] Y. Li, G. B. Yang, N. Chen, Y. P. Zhu, and X. L. Ding, "Early DIRECT mode decision for MVC using MB mode homogeneity and RD cost correlation," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 700–708, Sep. 2016.
[36] Z. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 166–176, Jun. 2015.
[37] Z. Zhou, Y. Wang, Q. M. J. Wu, C.-N. Yang, and X. Sun, "Effective and efficient global context verification for image copy detection," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 1, pp. 48–63, Jan. 2017.
[38] Z. Li, Z. Zhang, S. Guo, and J. Wang, "Video inter-frame forgery identification based on the consistency of quotient of MSSIM," *Security Commun. Netw.*, vol. 9, no. 17, pp. 4548–4556, 2016.



Yue Li received the B.S. degree from Hengyang Normal University in 2010 and the M.S. degree from Central South University in 2013. He is currently pursuing the Ph.D. degree with Hunan University, China. His current research interests include video coding and image processing.

Gaobo Yang received the Ph.D. degree in communication and information system from Shanghai University in 2004. He is a Professor with Hunan University, China. He is a Principal Investigator of several projects such as Natural Science Foundation of China, Special Pro-phase Project on National Basic Research Program of China (973) and program for New Century Excellent Talents in university. His current research interests are in the area of image and video signal processing and digital media forensics.

Yapei Zhu received the B.S. degree from Hengyang Normal University, China, in 2010 and the M.S. degree from Ningbo University in 2013. She is currently a Lecturer with Hengyang Normal University. Her research interest concentrates on image/video compression.

Xiangling Ding received the B.S. and M.S. degrees from Hunan Normal University in 2003 and 2006, respectively. He is currently pursuing the Ph.D. degree with Hunan University, China. His current research interests include video information security and image processing.

Xingming Sun received the B.S. degree in mathematics from Hunan Normal University, Hunan, China, in 1984, the M.S. degree in computing science from the Dalian University of Science and Technology, Dalian, China, in 1988, and the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2001. He is currently a Professor with the School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include network and information security, digital watermarking, cloud computing security, and wireless network security.