

# Probabilistic Decision Based Block Partitioning for Future Video Coding

Zhao Wang<sup>✉</sup>, Shiqi Wang, *Member, IEEE*, Jian Zhang, *Member, IEEE*,  
Shanshe Wang, and Siwei Ma, *Member, IEEE*

**Abstract**—In the latest Joint Video Exploration Team development, the quadtree plus binary tree (QTBT) block partitioning structure has been proposed for future video coding. Compared to the traditional quadtree structure of High Efficiency Video coding (HEVC) standard, QTBT provides more flexible patterns for splitting the blocks, which results in dramatically increased combinations of block partitions and high computational complexity. In view of this, a confidence interval based early termination (CIET) scheme is proposed for QTBT to identify the unnecessary partition modes in the sense of rate-distortion (RD) optimization. In particular, a RD model is established to predict the RD cost of each partition pattern without the full encoding process. Subsequently, the mode decision problem is casted into a probabilistic framework to select the final partition based on the confidence interval decision strategy. Experimental results show that the proposed CIET algorithm can speed up QTBT block partitioning structure by reducing 54.7% encoding time with only 1.12% increase in terms of bit rate. Moreover, the proposed scheme performs consistently well for the high resolution sequences, of which the video coding efficiency is crucial in real applications.

**Index Terms**—Block partitioning, probabilistic decision, confidence interval, rate-distortion model, quadtree plus binary tree.

## I. INTRODUCTION

THE ever-increasing demand to support High Definition (HD) and Ultra-High Definition (UHD) videos is creating new requirements to develop video coding technologies beyond the High Efficiency Video Coding (HEVC)

Manuscript received March 22, 2017; revised September 9, 2017; accepted November 6, 2017. Date of publication November 29, 2017; date of current version December 27, 2017. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2015CB351800, in part by the National Natural Science Foundation of China under Grant 61632001 and Grant 61571017, in part by the Top-Notch Young Talents Program of China and Shenzhen Peacock Plan, City University of Hong Kong under Grant 7200539/CS. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gene Cheung. (*Corresponding author: Siwei Ma*)

Z. Wang and S. Wang are with the Institute of Digital Media, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: zhaowang@pku.edu.cn; sswang@pku.edu.cn).

S. Wang is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: shiqwang@cityu.edu.hk).

J. Zhang is with the Visual Computing Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia (e-mail: jian.zhang@kaust.edu.sa).

S. Ma is with the Institute of Digital Media, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China and also with the Cooperative Medianet Innovation Center, Peking University, Beijing 100871, China (e-mail: swma@pku.edu.cn).

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/TIP.2017.2778564

standard. To meet the requirements, Joint Video Exploration Team (JVET) jointly developed by Moving Picture Experts Group (MPEG) and Video Coding Experts Group (VCEG) has proposed a series of coding tools for enhanced video coding, including Overlapped Block Motion Compensation (OBMC), Signal Dependent Transform (SDT) and Affine Inter Prediction Mode, etc [1]–[3]. Among these novel techniques, Quadtree Plus Binary Tree (QTBT) [4], [5] block partitioning structure has received significant interests due to its remarkable coding performance improvement. In particular, in addition to the traditional quadtree block partitioning structure in HEVC [6], QTBT provides recursive binary tree partition to adapt to the diverse properties of the video content.

In the hybrid video coding framework, block partitioning structure has been recognized as the core infrastructure based on which sophisticated coding tools are supported. In HEVC, each frame is first divided into equal-sized Coding Tree Units (CTUs), which are composed of one luma Coding Tree Blocks (CTB) and two chroma CTBs sharing identical partition decisions. Moreover, each CTU can be further divided into smaller blocks called Coding Units (CUs), following a quadtree partition structure to adapt to the local properties of video content. In particular, the decision of inter- or intra-coding is made at CU level. Each CU can be further split into one, two or four Prediction Units (PUs) according to the PU splitting mode, as shown in Fig. 1. After obtaining the prediction residuals, CU can also be recursively partitioned into Transform Units (TUs), which serve as the basic units for transformation and quantization. The current block partitioning structure of HEVC is recognized as one of the representative changes that largely outperform H.264/AVC [7]. However, there are still some limitations based on which potential improvements can be attained,

- 1) CU, which is the basic granularity for intra- or inter-prediction, can only be square following the quadtree structure. This may greatly lack the coding flexibility.
- 2) The PU splitting only has a few fixed types, which restricts the potentials of the prediction ability.
- 3) The residuals can only be transformed following square shape, which limits the potential of transformation.
- 4) The same block partitioning is applied to luma and chroma components simultaneously. However, the properties of luma component may not be identical with the properties of chroma components.

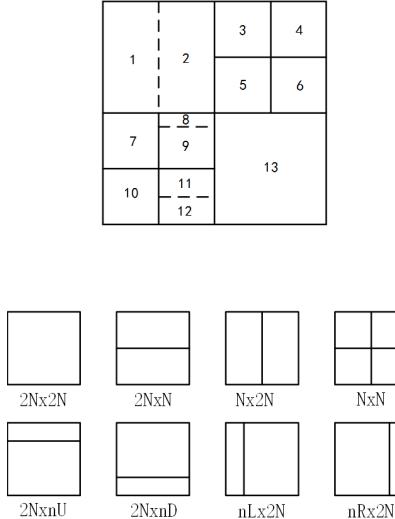


Fig. 1. HEVC quadtree CU structure and the PU splitting modes.

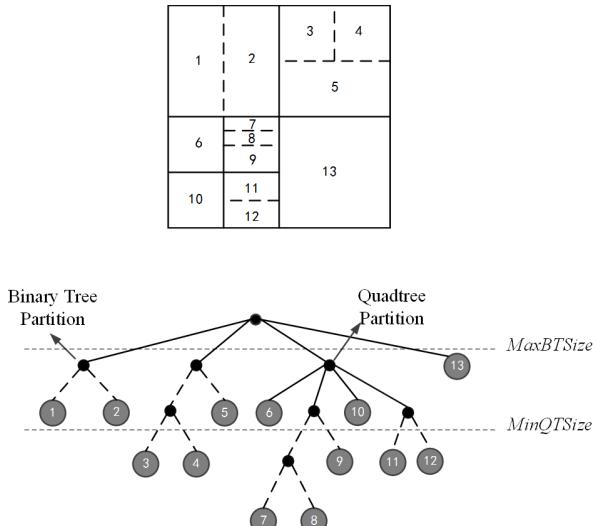


Fig. 2. QTBT block partitioning structure.

To address the above issues, a novel QTBT block partitioning structure has been proposed during the recent JVET development [4], [5]. In QTBT, a CTU is first partitioned by quadtree to obtain the smaller square blocks, which is similar to the quadtree partition in HEVC. In addition to quadtree partition, binary tree partition can be performed sequentially to split the quadtree leaf node into various rectangle patterns. In particular, the binary tree partition can be either symmetric horizontal or vertical splittings. The binary tree leaf node is treated as the CU and it will be directly used for prediction and transformation without any further partition. As such, the concepts of PU and TU are no longer necessary here, since they are always identical with CU. One example of QTBT is illustrated in Fig. 2, where *MinQTSIZE* and *MaxBTSIZE* are partition parameters to restrict the depths of quadtree and binary tree, respectively. In particular, *MinQTSIZE* represents the minimal allowed quadtree leaf node size, which implies that the quadtree partition must be terminated if the block size reaches *MinQTSIZE*. Correspondingly, *MaxBTSIZE*

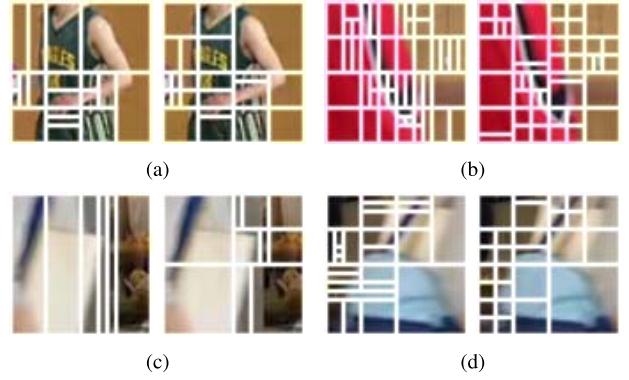


Fig. 3. Four pairs of CTUs coded by QTBT and HEVC respectively. In each pair: left: QTBT; right: HEVC.

specifies the maximal allowed binary tree root node size which indicates that the binary tree partition can only be applied for the blocks no larger than *MaxBTSIZE*. Moreover, a luma-chroma-separated block partitioning structure is adopted for I slice. Specifically, the QTBT structure of the chroma CTBs can be totally different from that of the luma CTB. However, with respect to P and B slices, the partitions of luma and chroma CTBs are still shared.

The main novelty of QTBT lies in the adoption of binary tree partition. In HEVC, though CU can be further split into rectangle PUs following one of the patterns in Fig. 1, these patterns can only be applied once-in-a-life-time for a given CU, which greatly limits the coding flexibility. For example, for the CU containing both homogeneous and complex content, driven by the complex area it should be elaborately structured with deep partition, which may result in *over-segmentation* for the adjacent smooth area [8]. In contrast to the PU patterns in HEVC, the binary tree partition in QTBT allows recursive splitting, such that the blocks can be flexibly partitioned into diverse shapes to adapt to the video content with fine granularity. Fig. 3 illustrates four pairs of CTUs partitioned by QTBT and HEVC, respectively. It can be obviously observed that binary tree partition appears frequently in QTBT, which further verifies its superior performance in terms of rate-distortion optimization.

Nevertheless, the performance gain of QTBT is brought at the cost of significantly increased encoding complexity, which mainly originates from the exponential growth of the combinations of partition patterns. This may hinder its practical implementations in real application scenarios. This paper aims at addressing this issue by proposing an effective partition decision scheme. In particular, a RD model based on motion divergence field (MDF) is established to estimate the RD cost of each partition mode. Based on the RD model, a confidence interval based early termination (CIET) algorithm is further proposed to remove the unnecessary partition iterations. The main contributions of this work are as follows,

- 1) A RD-model based on MDF is established to estimate the RD cost of each partition mode.
- 2) With the RD-model, we cast the mode decision problem into a probabilistic framework to develop a confidence interval based block partitioning decision scheme.

- 3) Regarding the coding performance and computational complexity, the proposed scheme achieves a better trade-off compared to JVET reference software as well as the state-of-the-art QTBT schemes, which is meaningful for the development of future video coding.

The remaining of this paper is organized as follows. The review of related works is presented in Section II. Section III develops the RD model based on MDF, and the CIET scheme is presented in Section IV. Extensive simulations and comparisons are shown in Section V and Section VI concludes this paper.

## II. RELATED WORKS

QTBT is found to be an effective block partitioning structure beyond the quadtree in HEVC, but has not been fully exploited from the perspective of complexity optimization. In the literature, numerous approaches have been proposed regarding the block partitioning decision based on H.264/AVC and HEVC. Generally speaking, they can be divided into five categories, including encoding information based, texture complexity based, neighbouring blocks based, RD cost estimation based, and machine learning based methods.

The first category speeds up the block partitioning decision by referring the intermediate parameters in the encoding process, such as the coded block flag (CBF), motion vector (MV), skip/merge mode, sample adaptive offset (SAO) parameters, etc. For example, Gweon *et al.* [9] proposed a CBF based early termination method by detecting zero CBFs to skip the remaining partitions of the current CU. Kim *et al.* [10] proposed to early terminate when the best prediction mode of the current CU is skip mode. Chiang and Chang also proposed an early termination method by applying zero-block detection using the differential MV [11]. In addition, Pan *et al.* developed a merge mode detection algorithm based on zero-block detection [12] and Kim *et al.* employed the differential MV and CBF to detect the skip mode [13]. In [14] and [15], the energy of the prediction residual was utilized to early terminate the CU splitting. In [16], SAO parameters of spatio-temporal CTUs were jointly used to estimate the CU encoding complexity, and MV, CBF and TU sizes were used to measure the motion complexity.

Methods of the second category exploit the correlations between the texture complexity of the video content and the block partitioning results. Generally speaking, the area rich with texture tends to be split into many small blocks to obtain well prediction, and the homogenous area may be roughly split. Kim *et al.* proposed a multi-stage sequential mode decision process that used spatial and transform domain features to filter out unlikely candidate modes [17]. Shen *et al.* [18] proposed an early determination of CU size decision with adaptive thresholds developed based on the texture homogeneity. For intra coding, Zhang *et al.* analyzed the relations between the texture characteristics of a block and its best coding mode, and developed an adaptive strategy for fast mode decision [19]. Luo *et al.* proposed a fast intra CU size decision scheme based on the keypoint detection [20]. By jointly utilizing the texture complexity and the neighbouring blocks information, Tian and Goto proposed a two-stage fast PU decision algorithm [21].

The schemes in the third category employ the depth information of the neighbouring blocks to predict the proper partitioning of the current block. The motivation is that there exist strong correlations among the current, spatial and temporal adjacent blocks. For example, Shen *et al.* proposed fast CU depth estimation and inter mode decision methods by utilizing the spatio-temporal correlations [22]–[24]. In order to reduce the depth range, Correa *et al.* also suggested to estimate the depth of CU using the depth of spatio-temporal adjacent CUs [25]. With the purpose of speeding up the rate-distortion optimization (RDO) in both the frame and the CU levels, Leng *et al.* employed the depth information of the neighbouring CTUs and the co-located CTU of the previous frame [26]. Recently, a Markov random field (MRF) based fast CU decision algorithm was proposed in [27] and [28]. The MRF can efficiently employ the neighbouring information in CU partitioning decision. These spatio-temporal based methods can perform well when the adjacent blocks belong to the same object. Nevertheless, around the boundary areas, the neighbouring CUs tend to be encoded with different sizes and modes, which may decrease the coding efficiency to some extent.

Methods of the fourth category aim to model the rate and distortion in a low-complexity function to substitute the time-consuming RDO process. Zhao *et al.* [29] proposed a RD estimation algorithm by modeling the transform coefficients with generalized Gaussian distribution. Analogously, a RD model based on a mixture of multiple Laplacian distributions was presented in [30]. Ma *et al.* in [31] proposed a novel RD model based on the quantization parameter (QP) difference among consecutive frames. For the intra mode, a mode selection method based on RD cost and sum of absolute difference (SAD) was proposed in [32]. In [33], the fast CU decision was implemented by comparing a derived threshold with the SAD cost difference between upper and sub CUs. However, these works can only skip the quantization, entropy coding and inverse quantization process in RDO, while the most time-consuming motion estimation (ME) should be still applied.

The fifth category is the machine learning based methods. By jointly utilizing the information of skip flags, merge flags, and RD costs of specific modes, Correa *et al.* trained fast decision trees for CU depth, PU mode and TU size, respectively [34], [35]. Shen *et al.* suggested a fast Bayesian theory based CU decision algorithm [36], where the sum of absolute transform differences (SATD), MV and RD costs were employed as the features. In [37], a three-output joint classifier was designed to control the risk of false prediction. In the machine learning based methods, almost all the information can be utilized to make the decision, such as the encoding parameters, intermediate values, spatio-temporal neighbouring blocks, texture complexities, and the performance of decision structure strongly relies on the selected features.

The aforementioned methods have been sophisticatedly developed for H.264/AVC or HEVC codecs and achieved significant encoding time savings. However, they cannot be directly used for speeding up QTBT due to the new design philosophy of the block partitioning structure and the

elimination of inter splitting modes. In particular, there are three further partition modes allowed other than only quadtree partition, which significantly increases the risk of partitioning decision. Moreover, the inter PU splitting modes in HEVC can produce valuable intermediate coding information in the encoding process, which can be employed for fast mode decision. However, such strategy may not be that efficient in the QTBT structure. In view of this, in [38], the QTBT partition is decided based on the adaptive adjustment of *MaxBTDepth* for each frame according to the temporal level. Moreover, in our previous work [39], a local-constraint QTBT scheme was proposed by dynamically deriving the partition parameters for each CTU. In this paper, to achieve a better trade-off between the coding performance and the encoding complexity, a confidence interval based early termination scheme is proposed for the QTBT block partitioning structure.

### III. RATE AND DISTORTION MODELS BASED ON MOTION DIVERGENCE FIELD

The optimal partition pattern in the hybrid video coding framework is optimized by minimizing the RD cost [40],

$$J = D + \lambda \cdot R \quad (1)$$

where  $J$  represents the RD cost to be minimized and  $\lambda$  denotes the Lagrangian multiplier.  $D$  and  $R$  are the reconstruction distortion and entropy coding bits, respectively. However, because the values of  $R$  and  $D$  are obtained after the overall encoding process, the RDO technique is extremely time-consuming. In this section, we explore efficient rate and distortion models to fast estimate the RD cost.

#### A. Analyses of Prediction Error Based on Motion Divergence Field

Motion compensation (MC) serves as the key component in removing the inter-frame redundancies in video coding. Therefore, we start from MC process to analyse the prediction error for RD modelling. In traditional video coding standards, each block utilizes block-level MV to obtain the corresponding prediction signals. In essence, motion activities in a local area may appear complicate and divergent, such that pixels in a block will have different motions, which can be further characterized by the motion vector field (MVF) [41]. From the perspective of MVF, the block based MV can be regarded as the approximation with the identical MV for all pixels, as shown in Fig. 4. Therefore, there may exist divergences between the actual pixel-level motions (denoted by  $\mathbf{m}_i$ ) and the approximated block-level MV (denoted by  $\mathbf{m}^*$ ),

$$\Delta m_i = \|\mathbf{m}_i - \mathbf{m}^*\|_2, \quad (2)$$

where  $\|\cdot\|$  denotes the  $\ell_2$  norm.

As such, the motion divergence field (MDF) can be generated by calculating the motion divergence  $\Delta m_i$  at the pixel level. Generally speaking, MDF with low energy corresponds to homogeneous MVF, such that accurate predictions are expected due to the consistent distributions of MVs. By contrast, MDF with high activities may bring about large prediction errors, which are closely relevant to the

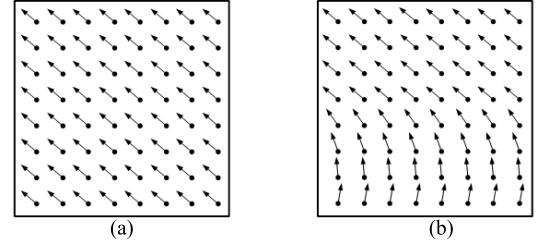


Fig. 4. Illustration of the motion divergence field. (a) Block-based MV. (b) Pixel-based MVF.

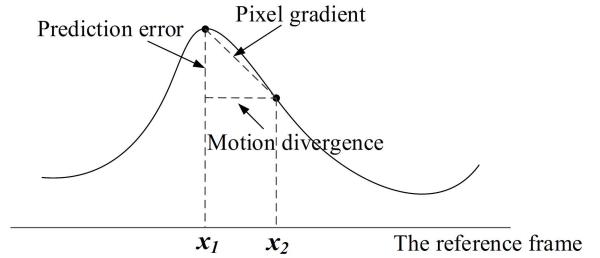


Fig. 5. Illustration of the prediction error.

reconstruction distortion and coding bits [42], [43]. As illustrated in Fig. 5, the curve represents the pixels in the reference frame, where  $x_1$  and  $x_2$  are the locations of the prediction pixels obtained by the pixel-level and block-level MVs, respectively. Assuming that  $x_1$  leads to the perfect match in motion estimation, it is obvious that larger divergence between  $x_1$  and  $x_2$  leads to higher prediction error. Therefore, we have

$$e_i \propto \Delta m_i, \quad (3)$$

where  $e_i$  denotes the prediction error.

Moreover, the gradient information which characterizes the complexity of local region is employed as another important feature to infer the prediction error. Due to the fact that gradient magnitude can successfully reflect the variations of local content, it has been widely applied in the complexity control algorithms [17], [19]. In particular, homogeneous region can easily obtain a good prediction, while complex region usually has large prediction error due to imperfect matching. Here, the prediction error is further modelled as,

$$e_i \propto \Delta m_i \cdot G_i, \quad (4)$$

where  $G_i$  denotes the gradient magnitude.

In this work, instead of pixel based MV searching, for simplification  $4 \times 4$  sub-block serves as the basic unit in describing the MDF. It is also worth mentioning that the MVs of  $4 \times 4$  units are obtained through motion estimation only on the nearest reference frame. Given the MV of each  $4 \times 4$  unit  $\mathbf{m}_i$ , the block-level MV  $\mathbf{m}^*$  is obtained by

$$\mathbf{m}^* = \left( \sum_{i=0}^n \mathbf{m}_i \right) / n, \quad (5)$$

where  $n$  is the number of  $4 \times 4$  units in the block.

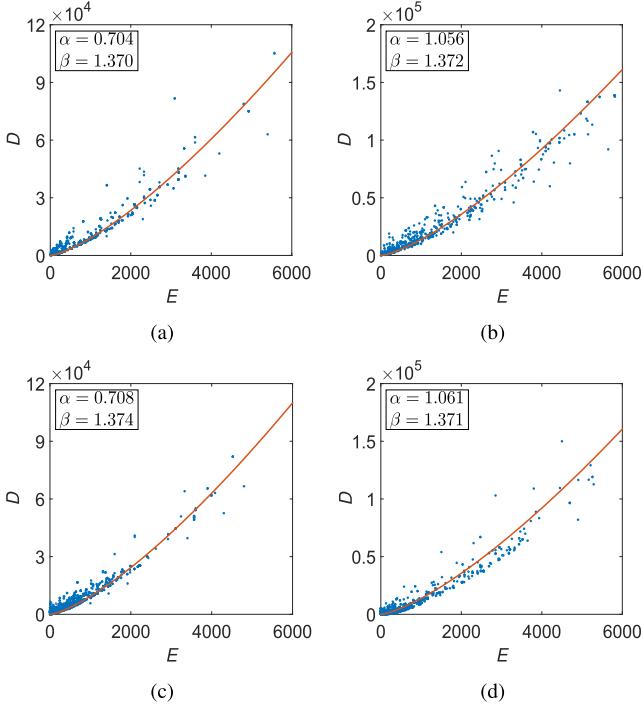


Fig. 6. Illustration of the relationship between between  $D$  and  $E$ . (a) *BasketballPass*, QP = 32. (b) *BasketballPass*, QP = 42. (c) *BQMall*, QP = 32. (d) *BQMall*, QP = 42.

With respect to the gradient magnitude of the pixels, it can be computed by

$$G_i = \sum_{x=0}^3 \sum_{y=0}^3 (|p_{x,y} - p_{x+1,y}| + |p_{x,y} - p_{x,y+1}|), \quad (6)$$

where  $p_{x,y}$  is the value of pixel within the current unit. Therefore, we can combine the motion and texture activities to characterize the energy of prediction errors,

$$E = \sum_{i=0}^n G_i \cdot \Delta m_i. \quad (7)$$

#### B. The Rate and Distortion Models Based on Motion Divergence Field

In hybrid video coding framework, prediction residuals are transformed, quantized and entropy coded to generate the final bit stream. Therefore, the energy of prediction residuals is closely relevant to the rate and distortion in video coding. To exactly explore such relationship, simulations are conducted on the sequences with different video contents and motion activities. The simulation results are illustrated in Fig. 6, where  $x$ -axis denotes  $E$  in (7), and  $y$ -axis is the corresponding reconstructed distortion  $D$ . From Fig. 6, it is observed that there exists a clear exponential relationship between them, which can be modelled as:

$$D = \alpha \cdot E^\beta, \quad (8)$$

where  $\alpha$  and  $\beta$  are the model parameters. In particular, the fitting value of  $\alpha$  varies when QP increases from 32 to 42, while  $\beta$  remains almost invariant. This is reasonable as the

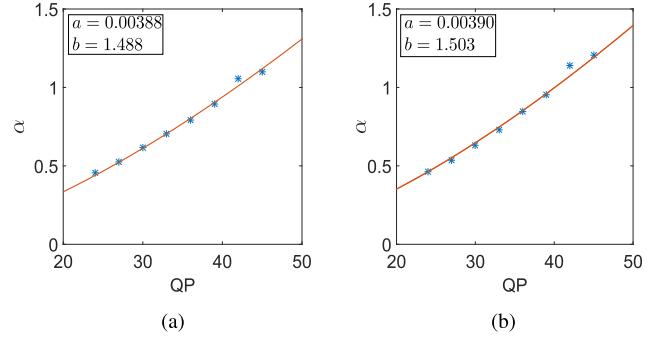


Fig. 7. Illustration of the relationship between  $\alpha$  and QP. (a) *BasketballPass*. (b) *BQMall*.

distortion increases monotonously with the QP value. Therefore, we further conduct an experiment to investigate the relationship between  $\alpha$  and QP. When QP spans in a wide range, the relationship between  $\alpha$  and QP is plotted in Fig. 7. This inspires us to model  $\alpha$  as follows

$$\alpha = a \cdot QP^b. \quad (9)$$

With the combination of (8) and (9), we can obtain the distortion model as

$$D = a \cdot QP^b \cdot E^\beta. \quad (10)$$

With respect to the encoding rate, it includes the entropy coding bits of residuals and the header bits of side information, which can be expressed as follows

$$R = R_r + R_h. \quad (11)$$

Here, the head bits  $R_h$  can be regarded as a constant and we focus on the residual bits  $R_r$ . Generally speaking, the block with high energy of prediction error requires more bits to be coded. In Fig. 8, the relationship between bit rate  $R_r$  and  $E$  is illustrated. One can discern that there is a linear relationship between  $R_r$  and  $E$ , implying that blocks with higher motion and texture activities consume more coding bits. As such, the bit rate  $R_r$  can be modelled as,

$$R_r = \gamma \cdot E. \quad (12)$$

Again, the parameter  $\gamma$  is closely relevant with the QP value, as larger QP leads to lower entropy coding bits and vice versa. This relationship is investigated in Fig. 9, which can be expressed as

$$\gamma = p \cdot e^{-q \cdot QP}. \quad (13)$$

By incorporating (13) into (12), the encoding rate of the residuals can be modelled as

$$R_r = p \cdot e^{-q \cdot QP} \cdot E. \quad (14)$$

For the model parameters  $a$ ,  $b$ ,  $\beta$ ,  $p$ , and  $q$  in (10)&(14), a practical way of obtaining the values is using offline training. The training set and fitting values are presented in Table I. It is observed that the parameters differ slightly for various sequences, which can be treated as constants and independent of the video content. As such, the parameters  $a$ ,  $b$ ,  $\beta$ ,  $p$ , and  $q$  are set to be 0.00388, 1.489, 1.372, 0.4394 and 0.1255 in this work.

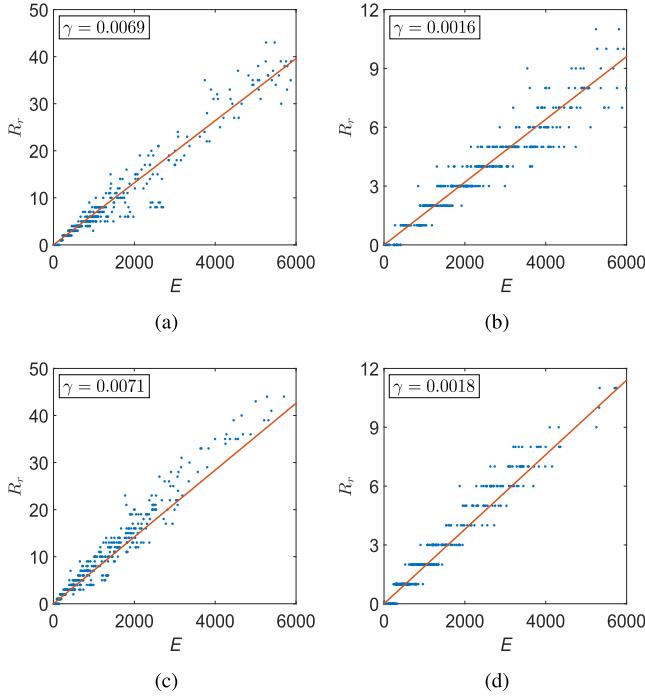


Fig. 8. Illustration of the relationship between  $R_r$  and  $E$ . (a) *BasketballPass*, QP = 32. (b) *BasketballPass*, QP = 42. (c) *BQMall*, QP = 32. (d) *BQMall*, QP = 42.

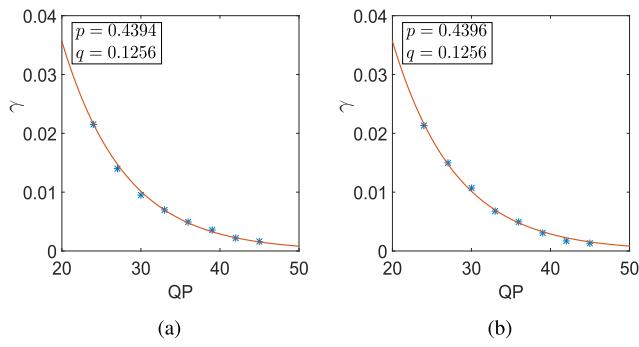


Fig. 9. Illustration of the relationship between  $\gamma$  and QP. (a) *BasketballPass*. (b) *BQMall*.

TABLE I

ESTIMATION OF PARAMETERS  $a$ ,  $b$ ,  $\beta$ ,  $p$ , AND  $q$

Sequences	$a$	$b$	$\beta$	$p$	$q$
<i>BasketballPass</i>	0.00388	1.488	1.371	0.4394	0.1256
<i>BQSquare</i>	0.00394	1.485	1.377	0.4398	0.1252
<i>BasketballDrill</i>	0.00389	1.492	1.370	0.4395	0.1255
<i>BQMall</i>	0.00390	1.490	1.372	0.4396	0.1256
<i>PartyScene</i>	0.00395	1.486	1.374	0.4396	0.1254
<i>FourPeople</i>	0.00383	1.492	1.369	0.4391	0.1258
<i>Johnny</i>	0.00385	1.491	1.367	0.4392	0.1258
<i>Kimono</i>	0.00387	1.490	1.375	0.4394	0.1254
<i>BasketballDrive</i>	0.00388	1.489	1.371	0.4393	0.1255
<i>Cactus</i>	0.00385	1.491	1.373	0.4395	0.1253
<b>Average</b>	<b>0.00388</b>	<b>1.489</b>	<b>1.372</b>	<b>0.4394</b>	<b>0.1255</b>

#### IV. CONFIDENCE INTERVAL BASED EARLY TERMINATION

The proposed RD model enables us to efficiently estimate the RD cost of each partition mode in a low complexity way. However, instead of binary classification on quadtree splitting,

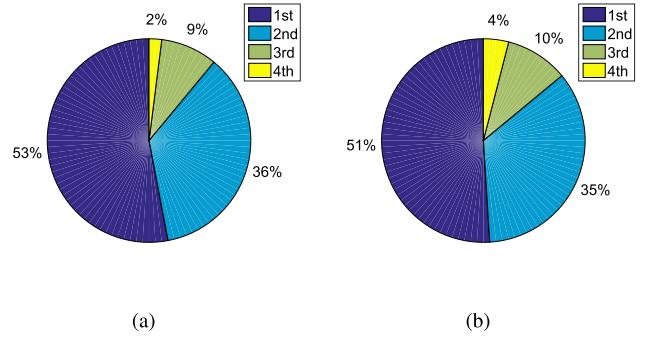


Fig. 10. Distribution of the selected modes with full RDO in terms of the rankings in order of estimated RD cost (from low to high). (a) *BasketballPass*. (b) *BQMall*.

the design philosophy of QTBT provides four splitting modes and only one is desired. To further improve the partition accuracy and achieve a good trade-off between the coding performance and the computational complexity, we cast the partition decision problem into a probabilistic framework and propose a novel early termination algorithm based on the proposed RD model.

In HEVC, the binary-classification problem on CU splitting can be formulated as follows,

$$p = \begin{cases} Y, & \hat{J}_{\text{split}} \leq \hat{J}_{\text{unsplit}} \\ N, & \hat{J}_{\text{split}} > \hat{J}_{\text{unsplit}}. \end{cases} \quad (15)$$

where  $p \in \{Y, N\}$  denotes the output of the predictor, and  $\hat{J}_{\text{split}}$  and  $\hat{J}_{\text{unsplit}}$  denote the estimated RD cost of splitting and un-splitting, respectively. In particular, if the estimated RD cost of splitting is smaller, the CU will be further split into the next depth. Otherwise, the CU splitting will be terminated. Compared to HEVC, QTBT is more complicated due to the four partition candidates: un-splitting, quadtree splitting, horizontal binary tree splitting and vertical binary tree splitting. For the multi-classification problem in QTBT, the optimal decision can be straightforwardly achieved according to

$$p = \arg \min_c \hat{J}_{RD}(c) \quad \{c \in S\}, \quad (16)$$

where  $c$  denotes each partition mode and  $S$  represents the candidate set. However, due to the variety of the video contents, encoding environments and limited available information, deciding the best partition mode purely based on the RD cost estimation may cause inaccurate predictions, such that a probabilistic decision framework is proposed. Specifically, by comparing the estimated RD costs, we first remove the unnecessary partition modes that fall outside of the confidence interval. Subsequently, the preserved partition modes are further examined by computing the exact RD costs to finally select the optimal mode.

Fig. 10 illustrates the distribution of the selected modes with full RDO in terms of the rankings in order of estimated RD cost. For example, the first component in *BasketballPass* indicates that 53% of the actual optimal modes searched by full RDO are identical with the predicted modes from the estimated RD costs. For other 36% blocks, the partition modes with the second small estimated RD costs are the optimal modes

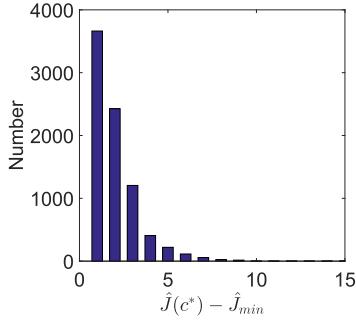


Fig. 11. Illustration of the  $\hat{J}(c^*) - \hat{J}_{min}$  distribution for wrongly predicted blocks.

in full RDO, and there are still 9% and 2% blocks whose full RDO selected modes rank 3rd and 4th according to the estimated RD costs, respectively. Similar simulation results can be observed for the sequence *BQMall* in Fig. 10-(b) as well. Let  $c^*$  denote the optimal partition mode with full RDO and  $c_i$  ( $i = 1, 2, 3, 4$ ) denote the partition modes ranked from low to high in terms of their estimated RD costs, we will have

$$\hat{J}(c_1) < \hat{J}(c_2) < \hat{J}(c_3) < \hat{J}(c_4). \quad (17)$$

According to Fig. 10, the probability that  $c^* = c_i$  follows,

$$P(c^* = c_1) > P(c^* = c_2) > P(c^* = c_3) > P(c^* = c_4). \quad (18)$$

Based on (17)&(18), it is concluded that the partition mode with smaller estimated RD cost has higher probability to be the full RDO optimal mode and vice versa.

In the following, we focus on the blocks whose optimal modes are not identical with the modes predicted by the estimated RD cost. Let  $\hat{J}_{min}$  denote the minimal value of the estimated RD cost. The distribution of the difference between  $\hat{J}(c^*)$  and  $\hat{J}_{min}$  is plotted in Fig. 11. One can discern that though these blocks are wrongly predicted, the probability of  $\hat{J}(c^*) - \hat{J}_{min}$  converges in a small range. This observation inspires us to propose the confidence interval based early termination scheme. In particular, given a confidence level  $\xi$  ( $0 \leq \xi \leq 1$ ), a threshold  $x_0$  can be obtained satisfying

$$\int_0^{x_0} Pd(x)dx = \xi, \quad (19)$$

where  $Pd(x)$  is the probability density function (pdf) of  $\hat{J}(c^*) - \hat{J}_{min}$ . This implies that the prediction accuracy can be ensured to be  $\xi$  if we preserve all the partition modes in the following subset,

$$\{R_p : \hat{J}(c_i) - \hat{J}_{min} < x_0\}. \quad (20)$$

As such,  $(\hat{J}_{min}, \hat{J}_{min} + x_0)$  is regarded as the confidence interval with confidence level  $\xi$  for the early termination problem. In this manner, the decision scheme can ensure a high prediction accuracy and meanwhile remove the other unnecessary partition modes falling outside of the confidence interval.

To determine the confidence interval, we first explore the pdf of  $\hat{J}(c^*) - \hat{J}_{min}$ . Considering the practical coding circumstances, the blocks are classified into two categories depending on whether they can be further partitioned by the quadtree

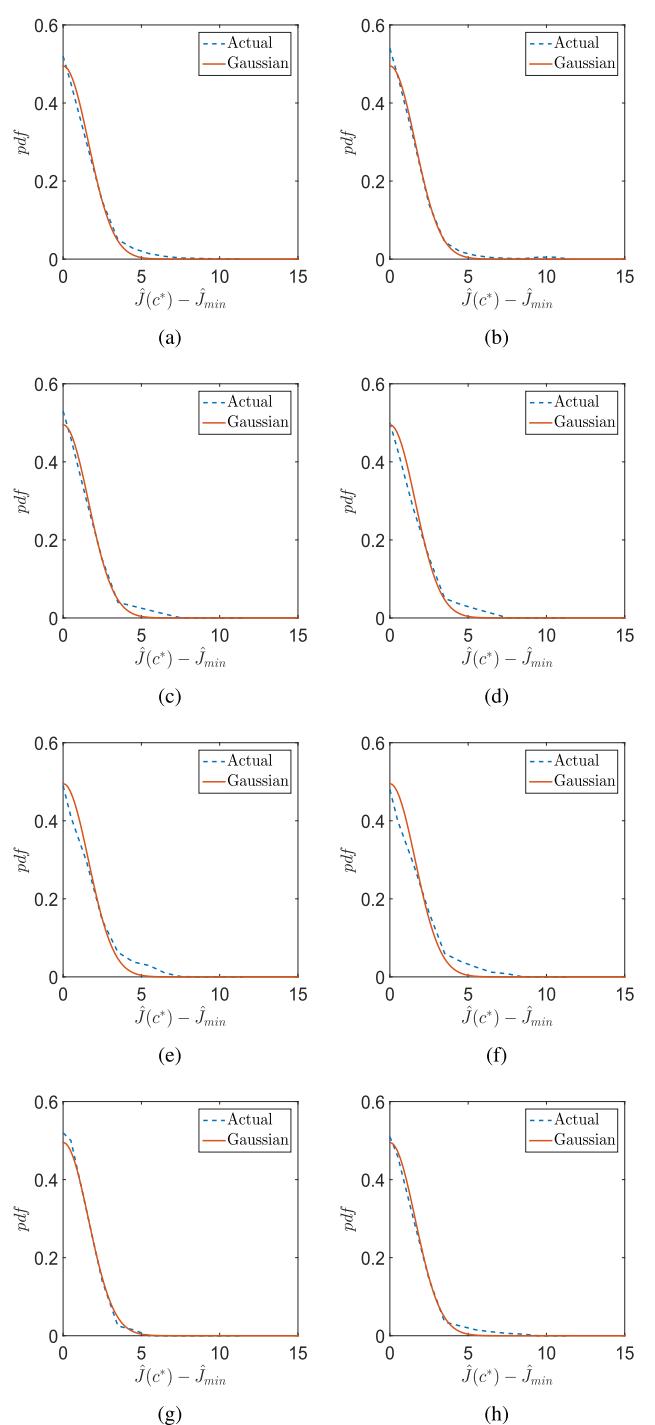


Fig. 12. Pdf of  $\hat{J}(c^*) - \hat{J}_{min}$  for the blocks with category A. (a) *BasketballPass*, QP=32. (b) *BasketballPass*, QP=42. (c) *BQMall*, QP=32. (d) *BQMall*, QP=42. (e) *Kimono*, QP=32. (f) *Kimono*, QP=42. (g) *Cactus*, QP=32. (h) *Cactus*, QP=42.

splitting. For the blocks in the first category A, both quadtree splitting and binary tree splitting are allowed. However, for the other blocks which reach the *MinQTSize* or have been partitioned by the binary tree, quadtree splitting is forbidden and only binary tree splitting is allowed. Such blocks are categorized into category B. The reason for differentiating them lies in the fact that different candidate sets have an effect on the decision scheme, which will be illustrated in the following simulations.

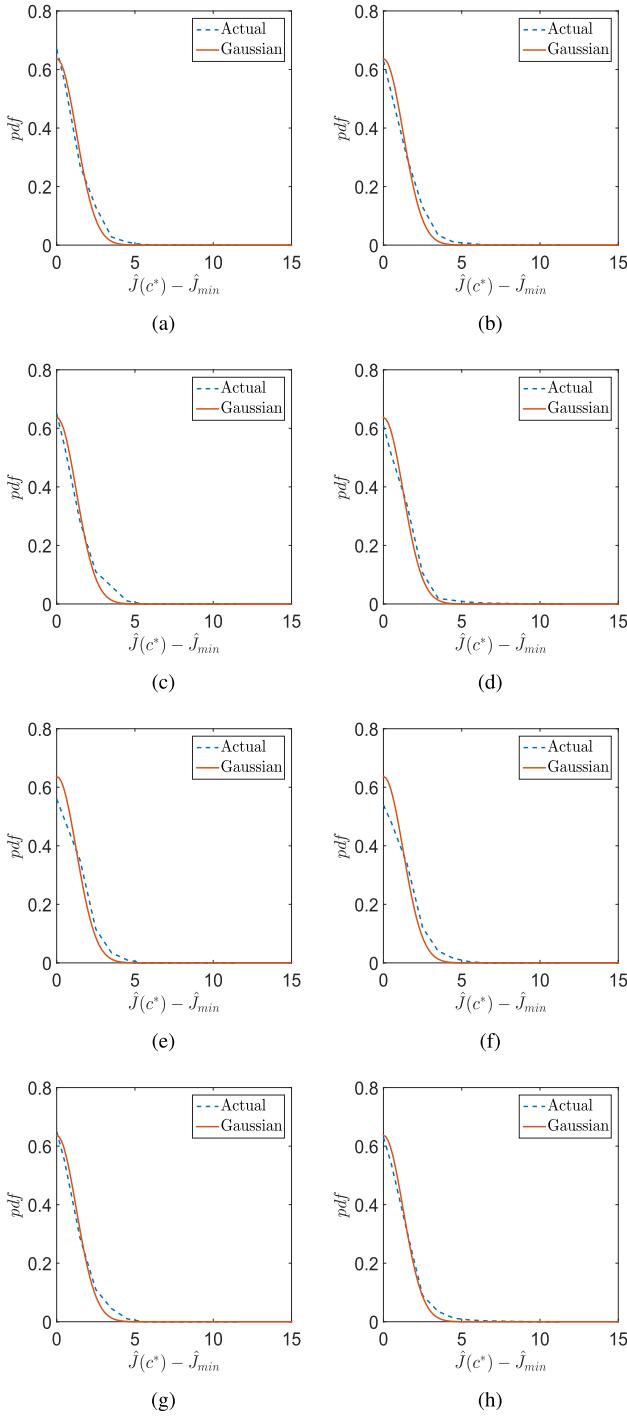


Fig. 13. Pdf of  $\hat{J}(c^*) - \hat{J}_{min}$  for the blocks with category B. (a) *BasketballPass*, QP=32. (b) *BasketballPass*, QP=42. (c) *BQMall*, QP=32. (d) *BQMall*, QP=42. (e) *Kimono*, QP=32. (f) *Kimono*, QP=42. (g) *Cactus*, QP=32. (h) *Cactus*, QP=42.

The pdf of  $\hat{J}(c^*) - \hat{J}_{min}$  for category A and B blocks are plotted in Figs. 12 and 13, respectively. Note that here the RD cost is computed on the pixel level to eliminate the influence of block size. From Figs. 12&13, it is obviously observed that the pdfs decreases with the increase of  $\hat{J}(c^*) - \hat{J}_{min}$ , which confirms the above conclusion that the partition mode with smaller estimated RD cost has higher probability to be the optimal mode and vice versa. It is also found that the

TABLE II  
ESTIMATED PARAMETERS  $\sigma$  IN EQ. (21)

Sequence	Category A	Category B
<i>BasketballPass</i>	1.612	1.253
<i>BQSquare</i>	1.540	1.347
<i>BasketballDrill</i>	1.687	1.371
<i>BQMall</i>	1.577	1.293
<i>PartyScene</i>	1.788	1.412
<i>FourPeople</i>	1.602	1.308
<i>Johnny</i>	1.632	1.288
<i>Kimono</i>	1.738	1.467
<i>BasketballDrive</i>	1.589	1.254
<i>Cactus</i>	1.522	1.282
Average	1.628	1.327

distribution of  $\hat{J}(c^*) - \hat{J}_{min}$  can be modelled with Gaussian-like distribution given by

$$Pd(\hat{J}(c^*) - \hat{J}_{min}) \approx \frac{2}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}. \quad (21)$$

The estimations of the modelling parameter  $\sigma$  are shown in Table II, and it is clearly observed that parameter  $\sigma$  can be treated as a constant independent of the video content. As such, it is chosen to be 1.628 and 1.327 for categories A and B, respectively. Moreover, it is interesting to find that  $\sigma$  for category B is smaller than that of category A, which coincides with the distributions in Figs. 12&13, as the distribution in Fig. 13 is sharper than that in Fig. 12. This implies that better prediction performance can be achieved for the blocks which can only be partitioned by the binary tree compared to both binary and quadtrees. The main reason lies in the fact that less candidates make the predictor more accurate to select the optimal mode, especially when the RD costs between horizontal and vertical binary tree splittings are apparently distinguished.

## V. EXPERIMENTAL RESULTS

In this section, we firstly determine the confidence levels by investigating the relationship between the confidence level and RD performance, as well as the relationship between the confidence level and encoding complexity. Then, the performance and analyses of the proposed CIET algorithm implemented on HM-13.0-QTBT [44] is detailed. Furthermore, we integrate the proposed scheme into the latest reference software JEM-7.0 [45] released by JVET and evaluate the performance under JVET common test conditions (CTC) [46]. Finally, we compare the proposed method with state-of-the-art low complexity QTBT algorithms to further demonstrate the advantages of our method.

### A. Investigations on the Confidence Level

In the proposed scheme, the confidence level plays a crucial role in determining the trade-off between the RD performance and coding complexity. Apparently, higher confidence level ensures better prediction accuracy while leading to lower complexity saving. To select the appropriate confidence levels, simulations are conducted for blocks in categories A and B, respectively. In particular, the confidence level ranges from 90% to 100%, and the encoding time reduction( $\Delta ET$ ) and

TABLE III  
PERFORMANCE OF THE PROPOSED CIET ALGORITHM ON HM-13.0-QTBT

Class	Sequence	Random Access			Lowdelay-B			Lowdelay-P					
		Y	U	V	$\Delta ET$	Y	U	V	$\Delta ET$	Y	U	V	
A	<i>PeopleOnStreet</i>	+1.3%	+1.6%	+1.9%	62.5%	+1.3%	+1.5%	+1.4%	62.9%	+1.1%	+1.0%	+1.2%	56.9%
	<i>Traffic</i>	+0.9%	+0.7%	+1.1%	48.3%	+1.2%	+1.4%	+1.5%	54.5%	+1.0%	+1.2%	+1.1%	48.2%
B	<i>Kimono</i>	+1.2%	+0.9%	+1.4%	57.3%	+1.3%	+1.4%	+1.3%	56.2%	+1.2%	+1.1%	+1.3%	54.0%
	<i>ParkScene</i>	+0.9%	+1.2%	+0.8%	47.2%	+1.3%	+1.1%	+1.2%	59.7%	+1.0%	+1.2%	+1.0%	51.4%
C	<i>Cactus</i>	+1.3%	+1.1%	+1.6%	52.4%	+1.5%	+1.7%	+1.7%	49.2%	+1.3%	+1.5%	+0.9%	54.5%
	<i>BasketballDrive</i>	+1.2%	+1.3%	+1.4%	51.3%	+1.3%	+1.2%	+1.0%	52.9%	+1.3%	+1.3%	+1.4%	62.1%
D	<i>BQTerrace</i>	+0.7%	+1.0%	+0.7%	41.2%	+0.8%	+0.7%	+1.2%	44.3%	+0.7%	+0.9%	+0.8%	40.2%
	<i>BasketballDrill</i>	+1.3%	+0.9%	+1.4%	67.1%	+1.6%	+1.8%	+1.3%	67.0%	+1.6%	+1.4%	+1.3%	64.2%
E	<i>BQMall</i>	+1.1%	+1.0%	+0.7%	51.8%	+1.4%	+1.2%	+1.2%	54.8%	+1.3%	+1.1%	+1.5%	48.4%
	<i>PartyScene</i>	+1.6%	+1.7%	+1.4%	65.8%	+1.8%	+1.7%	+2.0%	68.3%	+1.3%	+1.9%	+1.4%	59.6%
F	<i>RaceHorsesC</i>	+1.5%	+1.6%	+1.4%	63.2%	+1.3%	+1.1%	+1.2%	49.4%	+1.4%	+1.2%	+1.0%	51.8%
	<i>BasketballPass</i>	+0.8%	+0.7%	+0.8%	38.7%	+0.9%	+1.0%	+1.5%	43.2%	+0.9%	+1.1%	+1.3%	44.1%
G	<i>BQSquare</i>	+1.1%	+1.4%	+1.3%	54.9%	+1.0%	+1.2%	+1.2%	55.3%	+1.0%	+1.3%	+1.3%	54.7%
	<i>BlowingBubbles</i>	+1.0%	+1.2%	+1.3%	53.2%	+1.3%	+1.4%	+1.1%	59.2%	+1.4%	+1.4%	+1.3%	54.5%
H	<i>RaceHorses</i>	+1.4%	+1.6%	+1.8%	62.2%	+1.5%	+1.4%	+1.3%	67.2%	+1.2%	+1.4%	+1.3%	57.9%
	<i>FourPeople</i>	+1.3%	+1.4%	+1.2%	52.4%	+0.8%	+0.7%	+0.7%	49.5%	+1.2%	+1.5%	+0.9%	44.3%
I	<i>Johnny</i>	+1.2%	+1.6%	+1.9%	67.5%	+1.3%	+1.5%	+1.7%	63.0%	+1.0%	+1.0%	+1.2%	56.8%
	<i>KristenAndSara</i>	+0.9%	+0.7%	+1.1%	48.8%	+1.2%	+1.4%	+1.5%	64.2%	+0.7%	+0.8%	+1.1%	38.9%
<b>Overall</b>		<b>+1.12%</b>	<b>+1.22%</b>	<b>+1.29%</b>	<b>54.7%</b>	<b>+1.27%</b>	<b>+1.30%</b>	<b>+1.34%</b>	<b>56.7%</b>	<b>+1.16%</b>	<b>+1.24%</b>	<b>+1.18%</b>	<b>52.4%</b>

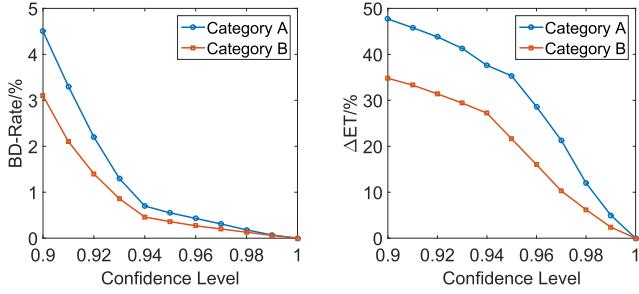


Fig. 14. Encoding time reductions and RD performance variations under different confidence levels.

RD performance are presented in Fig. 14, where Bjontegaard delta rate (BD-Rate) [47] is adopted to evaluate the RD performance variation. From Fig. 14(a)&(b), it can be obviously found that significant RD performance loss will be introduced when the confidence level is lower than 94%. This implies that we may suffer from more false predictions in these circumstances. For the confidence levels between 94% and 100%, there is a clear linear relationship between BD-Rate and the confidence level. Regarding the encoding time savings, it is observed that the slope of the curve from 90% to 95% is much lower than that from 95% to 100% for category A. Moreover, with respect to the blocks in category B, which can only be partitioned by the binary tree, it is found that 94% is the singularity in terms of time saving. As such, to achieve a good trade-off between RD performance and coding complexity, 95% and 94% are selected as the confidence levels for blocks belonging to categories A and B, respectively. For the selected confidence levels, the threshold  $x_0$  can be computed by (21), which are 3.2 and 2.5 for category A and category B, respectively.

### B. Performance Evaluation of the Proposed Method

This subsection evaluates the RD performance and time saving of the proposed CIET scheme. We firstly integrate the proposed algorithm on the platform HM-13.0-QTBT [44],

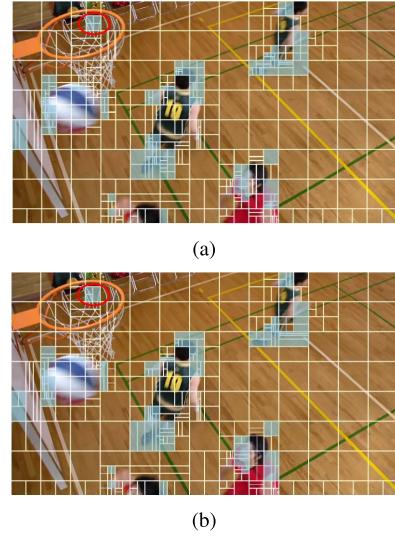


Fig. 15. Comparisons of the block partitions between anchor (HM-13.0-QTBT) and the proposed scheme. (a) Anchor. (b) Proposed.

and conduct simulations conforming to the HEVC CTC [48]. Table III shows the performance of the proposed algorithm, where Y, U and V represent the BD-Rate of luma and chroma components respectively. It is observed that approximately 54.7%, 56.7% and 52.4% time savings can be achieved for Random Access (RA), Lowdelay-B (LDB) and Lowdelay-P (LDP) configurations, respectively. In particular, the coding time reductions are more significant for high-activity sequences such as *PeopleOnStreet*, *BasketballDrill*, *PartyScene* and *RaceHorses*. This is mainly because the areas rich with motion activities have high energy in terms of MDF, such that the proposed scheme can efficiently differentiate the RD costs of different partitions. For the homogenous areas, since the motion divergence and texture complexity may approach zero, it is difficult for the proposed model to distinguish different partition patterns.

TABLE IV  
PERFORMANCE OF THE PROPOSED CIET ALGORITHM ON JEM-7.0

Class	Sequence	Random Access				Lowdelay-B				Lowdelay-P			
		Y	U	V	$\Delta ET$	Y	U	V	$\Delta ET$	Y	U	V	$\Delta ET$
A1	Tango2	+1.3%	+1.5%	+1.5%	50.4%	+1.3%	+1.6%	+1.6%	51.7%	+1.2%	+1.1%	+1.1%	48.4%
	Drums100	+1.1%	+1.2%	+1.1%	57.2%	+1.2%	+1.3%	+1.2%	57.8%	+1.5%	+1.4%	+1.3%	51.4%
	Campfire	+1.7%	+1.7%	+1.8%	46.7%	+1.8%	+1.8%	+1.6%	50.3%	+1.5%	+1.7%	+1.6%	53.7%
	ToddlerFountain2	+1.2%	+1.3%	+1.4%	48.4%	+1.4%	+1.4%	+1.3%	52.9%	+1.3%	+1.0%	+1.2%	50.7%
A2	CatRobot	+1.3%	+1.2%	+1.3%	42.9%	+1.5%	+1.6%	+1.5%	47.2%	+1.4%	+1.1%	+1.3%	50.1%
	TrafficFlow	+0.7%	+1.2%	+1.2%	60.2%	+0.9%	+1.1%	+1.0%	56.4%	+1.0%	+0.7%	+0.4%	55.5%
	DaylightRoad2	+0.8%	+1.1%	+1.0%	52.8%	+1.1%	+1.2%	+1.3%	55.3%	+1.0%	+1.2%	+1.3%	52.7%
	Rollercoaster2	+1.4%	+1.5%	+1.7%	50.3%	+1.5%	+1.6%	+1.7%	49.4%	+1.6%	+1.6%	+1.6%	47.9%
B	Kimono	+1.4%	+1.1%	+1.5%	54.7%	+1.4%	+1.2%	+1.3%	54.1%	+1.4%	+1.6%	+1.7%	52.0%
	ParkScene	+1.2%	+1.1%	+0.9%	45.7%	+1.3%	+1.3%	+1.3%	56.9%	+1.3%	+1.5%	+1.4%	47.4%
	Cactus	+1.5%	+1.2%	+1.4%	48.2%	+1.7%	+2.0%	+1.9%	47.4%	+1.2%	+1.3%	+1.0%	52.8%
	BasketballDrive	+1.0%	+1.1%	+1.2%	42.3%	+1.5%	+1.5%	+1.3%	49.6%	+1.4%	+1.4%	+1.5%	60.3%
C	BQTerrace	+0.9%	+1.1%	+0.9%	37.1%	+0.9%	+0.8%	+1.2%	44.1%	+0.9%	+1.1%	+1.2%	39.6%
	BasketballDrill	+1.5%	+1.3%	+1.5%	62.3%	+1.5%	+1.4%	+1.1%	60.3%	+1.4%	+1.0%	+0.6%	61.6%
	BQMall	+0.9%	+1.0%	+0.8%	47.7%	+1.5%	+1.4%	+1.5%	55.1%	+1.4%	+1.3%	+1.2%	42.5%
	PartyScene	+1.7%	+1.7%	+1.6%	62.3%	+1.9%	+1.7%	+1.9%	63.7%	+1.5%	+1.7%	+1.5%	55.3%
D	RaceHorsesC	+1.6%	+1.7%	+1.7%	60.5%	+1.8%	+1.7%	+1.6%	45.7%	+1.2%	+1.5%	+1.3%	50.9%
	BasketballPass	+1.0%	+1.0%	+1.0%	36.9%	+1.1%	+1.2%	+1.3%	41.7%	+1.0%	+1.1%	+1.2%	40.5%
	BQSquare	+1.3%	+1.5%	+1.6%	51.8%	+1.5%	+1.5%	+1.3%	53.2%	+1.3%	+1.7%	+1.9%	53.3%
	BlowingBubbles	+1.2%	+1.4%	+1.3%	50.3%	+1.2%	+1.3%	+1.0%	56.4%	+1.5%	+1.5%	+1.4%	52.1%
E	RaceHorses	+1.3%	+1.7%	+1.6%	56.5%	+1.5%	+1.2%	+1.4%	62.8%	+1.3%	+1.3%	+1.4%	53.8%
	FourPeople	+1.3%	+1.5%	+1.3%	47.7%	+0.9%	+0.9%	+0.8%	45.7%	+1.0%	+1.3%	+1.1%	42.3%
	Johnny	+1.1%	+1.3%	+1.4%	55.4%	+1.4%	+1.6%	+1.8%	60.2%	+1.3%	+1.2%	+1.7%	52.5%
	KristenAndSara	+1.0%	+0.9%	+1.0%	45.6%	+1.1%	+1.3%	+1.4%	57.0%	+0.9%	+1.2%	+1.2%	38.0%
<b>Overall</b>		<b>+1.23%</b>	<b>+1.28%</b>	<b>+1.31%</b>	<b>50.6%</b>	<b>+1.37%</b>	<b>+1.40%</b>	<b>+1.39%</b>	<b>53.1%</b>	<b>+1.27%</b>	<b>+1.31%</b>	<b>+1.30%</b>	<b>50.2%</b>

TABLE V  
COMPARISONS WITH THE STATE-OF-THE-ART METHODS ON HM-13.0-QTBT UNDER RA CONFIGURATION

Class	Sequence	Yamamoto's [38]				Wang <i>et al.</i> 's [39]				CIET			
		Y	U	V	$\Delta ET$	Y	U	V	$\Delta ET$	Y	U	V	$\Delta ET$
A	PeopleOnStreet	+0.5%	+0.4%	+0.3%	12.5%	+0.5%	+0.6%	+0.6%	22.2%	+1.3%	+1.6%	+1.9%	62.5%
	Traffic	+0.4%	+0.3%	+0.3%	18.3%	+0.7%	+1.0%	+0.4%	29.1%	+0.9%	+0.7%	+1.1%	48.3%
	Kimono	+0.5%	+0.4%	+0.5%	16.8%	+0.5%	+0.5%	+0.7%	38.3%	+1.2%	+0.9%	+1.4%	57.3%
	ParkScene	+0.4%	+0.1%	+0.2%	17.8%	+0.6%	+0.8%	+0.7%	27.6%	+0.9%	+1.2%	+0.8%	47.2%
B	Cactus	+0.5%	+0.3%	+0.2%	18.4%	+0.6%	+0.9%	+0.7%	25.4%	+1.3%	+1.1%	+1.6%	52.4%
	BasketballDrive	+0.7%	+0.7%	+0.5%	17.5%	+1.2%	+1.5%	+1.2%	37.9%	+1.2%	+1.3%	+1.4%	51.3%
	BQTerrace	+0.4%	-0.1%	+0.1%	17.6%	+0.3%	+0.4%	+0.6%	16.7%	+0.7%	+1.0%	+0.7%	41.2%
	BasketballDrill	+0.5%	+0.5%	+0.6%	22.2%	+0.4%	+0.9%	+0.5%	19.4%	+1.3%	+0.9%	+1.4%	67.1%
C	BQMall	+0.5%	+0.4%	+0.2%	19.4%	+0.5%	+0.8%	+0.7%	17.8%	+1.1%	+1.0%	+0.7%	51.8%
	PartyScene	+0.4%	+0.2%	+0.2%	18.6%	+0.1%	+0.4%	+0.3%	4.8%	+1.6%	+1.7%	+1.4%	65.8%
	RaceHorsesC	+0.7%	+0.5%	+0.8%	21.5%	+0.3%	+0.8%	+0.5%	16.7%	+1.5%	+1.6%	+1.4%	63.2%
	BasketballPass	+0.5%	+0.5%	+0.6%	17.7%	+0.6%	+0.5%	+0.9%	21.2%	+0.8%	+0.7%	+0.8%	38.7%
D	BQSquare	+0.3%	-0.6%	-0.5%	17.5%	+0.3%	+0.4%	+0.1%	4.6%	+1.1%	+1.4%	+1.3%	54.9%
	BlowingBubbles	+0.2%	-0.2%	-0.1%	15.4%	+0.2%	+0.4%	+0.3%	8.4%	+1.0%	+1.2%	+1.3%	53.2%
	RaceHorses	+0.6%	+0.4%	+0.6%	18.2%	+0.2%	+0.3%	+0.6%	9.0%	+1.4%	+1.6%	+1.8%	62.2%
	FourPeople	+0.8%	+0.7%	+1.0%	22.4%	+0.9%	+1.2%	+1.0%	28.8%	+1.3%	+1.4%	+1.2%	52.4%
E	Johnny	+0.7%	+0.6%	+0.5%	21.5%	+1.1%	+1.1%	+1.3%	31.7%	+1.2%	+1.6%	+1.9%	67.5%
	KristenAndSara	+0.8%	+0.7%	+0.7%	24.2%	+0.8%	+1.3%	+1.1%	31.6%	+0.9%	+0.7%	+1.1%	48.8%
<b>Overall</b>		<b>+0.52%</b>	<b>+0.32%</b>	<b>+0.37%</b>	<b>18.7%</b>	<b>+0.54%</b>	<b>+0.77%</b>	<b>+0.68%</b>	<b>21.8%</b>	<b>+1.12%</b>	<b>+1.22%</b>	<b>+1.29%</b>	<b>54.7%</b>

Moreover, tolerable RD performance loss is observed, and on average 1.12%, 1.27% and 1.16% BD-Rate increases for RA, LDB and LDP configurations are introduced. Fig. 15 compares the block partitions compressed by anchor and the proposed scheme respectively, where the shadow areas represent the blocks partitioned by different modes. It is observed that only marginal parts of the blocks are partitioned differently. Moreover, even though these blocks are partitioned by different modes, their final splitting patterns are still similar. For example, as shown in Fig. 15, the block denoted by red ellipse applies the quadtree partition in full RDO. Regarding the proposed scheme, the block is first partitioned by the

vertical binary tree partition, and then the left sub-block is further partitioned by the horizontal binary tree partition. Therefore, only the right sub-block brings variations in the coding performance.

We also integrate the proposed algorithm into the latest JVET reference software JEM-7.0 [45], and conduct simulations conforming to the JVET common test conditions defined in [46]. According to the JVET CTC, the test sequences include HEVC test sequences (*Class B-E*, 8-bit) and the new added UHD sequences (*Class A1-A2*, 10-bit). Simulation results are shown in Table IV. It is observed that approximately 50.6%, 53.1% and 50.2% time savings with tolerable 1.23%,

1.37% and 1.27% BD-Rate increases can be achieved for RA, LDB and LDP configurations, respectively. Moreover, it can also be observed that the proposed algorithm performs consistently well for the high resolution, high bit-depth sequences, which illustrates the robustness of our method.

To further verify the superior performance of CIET, we compare it with the state-of-the-art methods, including Yamamoto and Ikai [38] and our previous Wang *et al.* [39] methods. To ensure the fairness, we implement all algorithms on the HM-13.0-QTBT platform and the comparison results are shown in Table V. Compared with the two state-of-the-art methods, it is observed that the proposed CIET algorithm achieves significant time saving while maintaining an acceptable coding loss, which promotes QTBT to be applied in the real-time applications.

## VI. CONCLUSIONS

We propose a confidence interval based early termination scheme for the QTBT block partitioning structure with the aim of achieving a good trade-off between the coding performance and the encoding complexity. In particular, a RD model based on motion divergence field is established to predict the RD cost of each partition mode without full encoding process. By exploring the relationship between the optimal partition mode and its estimated RD cost, we cast the mode decision problem into a probabilistic framework to eliminate unnecessary partition iterations. Extensive experimental results verify that the proposed scheme is capable of speeding up QTBT block partitioning structure by reducing 54.7% encoding time on average with only 1.12% increase in terms of bit rate. Moreover, the proposed scheme performs consistently well on the latest reference software JEM-7.0 for the high resolution sequences, of which the video coding efficiency is crucial in real applications.

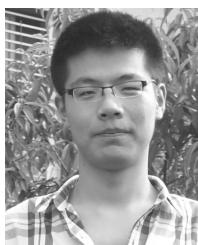
## ACKNOWLEDGEMENT

The authors would like to thank the Associate Editor Prof. Gene Cheung and anonymous reviewers for their valuable comments and suggestions which help to improve the technical presentation of this paper. The authors would also like to thank Dr. Tiesong Zhao for his insightful comments.

## REFERENCES

- [1] J. Chen, X. Li, F. Zou, M. Karczewicz, W.-J. Chien, and T. Hsieh, *Performance Evaluation of JEM 1 Tools by Qualcomm*, document JVET-B0045, Joint Video Exploration Team (JVET), Feb. 2016.
- [2] E. Alshina, A. Alshin, K. Choi, and M. Park, *Performance of JEM 1 Tools Analysis*, Joint Video Exploration Team (JVET), document JVET-B0022, Feb. 2016.
- [3] H. Zhang, H. Chen, X. Ma, and H. Yang, *Performance Analysis of Affine Inter Prediction in JEM1.0*, document JVET-B0037, Joint Video Exploration Team (JVET), Feb. 2016.
- [4] J. An, Y.-W. Chen, K. Zhang, H. Huang, Y.-W. Huang, and S. Lei, *Block Partitioning Structure for Next Generation Video Coding*, document m37524 and ITU-T SG16 COM16-C966, MPEG, Oct. 2015.
- [5] J. An, H. Huang, K. Zhang, Y.-W. Huang, and S. Lei, *Quadtree Plus Binary Tree Structure Integration With JEM Tools*, document JVET-B0023, Joint Video Exploration Team (JVET), Feb. 2016.
- [6] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.
- [7] 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.
- [8] V. Sze, M. Budagavi, and G. J. Sullivan, “High efficiency video coding (HEVC),” *Algorithms Architectures (Integrated Circuits and Systems)*. New York, NY, USA: Springer, 2014, pp. 1–375.
- [9] R. Gweon, Y. Lee, and J. Lim, *Early Termination of CU Encoding to Reduce HEVC Complexity*, document JCTVC-F045, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), 2011.
- [10] J. Kim, S. Jeong, S. Cho, and J. S. Choi, “Adaptive coding unit early termination algorithm for HEVC,” in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2012, pp. 261–262.
- [11] P.-T. Chiang and T. S. Chang, “Fast zero block detection and early CU termination for HEVC video coding,” in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 1640–1643.
- [12] Z. Pan, S. Kwong, M.-T. Sun, and J. Lei, “Early MERGE mode decision based on motion estimation and hierarchical depth correlation for HEVC,” *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 405–412, Jun. 2014.
- [13] J. Kim, J. Yang, K. Won, and B. Jeon, “Early determination of mode decision for HEVC,” in *Proc. Picture Coding Symp. (PCS)*, 2012, pp. 449–452.
- [14] S. Ma, S. Wang, S. Wang, L. Zhao, Q. Yu, and W. Gao, “Low complexity rate distortion optimization for HEVC,” in *Proc. Data Compress. Conf. (DCC)*, Mar. 2013, pp. 73–82.
- [15] S. Wang *et al.*, “Low complexity encoder optimization for HEVC,” *J. Vis. Commun. Image Represent.*, vol. 35, pp. 120–131, Feb. 2016.
- [16] S. Ahn, M. Kim, and S. Park, “Fast decision of CU partitioning based on SAO parameter, motion and PU/TU split information for HEVC,” in *Proc. Picture Coding Symp. (PCS)*, Dec. 2013, pp. 113–116.
- [17] C. Kim, H.-H. Shih, and C.-C. Kuo, “Feature-based intra-prediction mode decision for H. 264,” in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 2, 2004, pp. 769–772.
- [18] 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.
- [19] M. Zhang, C. Zhao, and J. Xu, “An adaptive fast intra mode decision in HEVC,” in *Proc. 19th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2012, pp. 221–224.
- [20] F. Luo, S. Wang, S. Ma, N. Zhang, Y. Zhou, and W. Gao, “Fast intra coding unit size decision for HEVC with GPU based keypoint detection,” in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2017, pp. 1–4.
- [21] G. Tian and S. Goto, “Content adaptive prediction unit size decision algorithm for HEVC intra coding,” in *Proc. Picture Coding Symp. (PCS)*, May 2012, pp. 405–408.
- [22] 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.
- [23] 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.
- [24] 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.
- [25] G. Correa, P. Assuncao, L. Agostini, and L. A. Da Silva Cruz, “Coding tree depth estimation for complexity reduction of HEVC,” in *Proc. Data Compress. Conf. (DCC)*, 2013, pp. 43–52.
- [26] J. Leng, L. Sun, T. Ikenaga, and S. Sakaida, “Content based hierarchical fast coding unit decision algorithm for HEVC,” in *Proc. Int. Conf. Multimedia Signal Process. (CMSPI)*, vol. 1, May 2011, pp. 56–59.
- [27] J. Xiong, H. Li, F. Meng, S. Zhu, Q. Wu, and B. Zeng, “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.
- [28] J. Xiong, H. Li, F. Meng, B. Zeng, S. Zhu, and Q. Wu, “Fast and efficient inter CU decision for high efficiency video coding,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 3715–3719.
- [29] X. Zhao, J. Sun, S. Ma, and W. Gao, “Novel statistical modeling, analysis and implementation of rate-distortion estimation for H. 264/AVC coders,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 5, pp. 647–660, May 2010.
- [30] B. Lee and M. Kim, “Modeling rates and distortions based on a mixture of Laplacian distributions for inter-predicted residues in quadtree coding of HEVC,” *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 571–574, Oct. 2011.

- [31] S. Ma, J. Si, and S. Wang, "A study on the rate distortion modeling for high efficiency video coding," in *Proc. Int. Conf. Image Process.*, Sep. 2012, pp. 181–184.
- [32] J. Tariq, S. Kwong, and H. Yuan, "HEVC intra mode selection based on rate distortion (RD) cost and sum of absolute difference (SAD)," *J. Vis. Commun. Image Represent.*, vol. 35, pp. 112–119, Sep. 2016.
- [33] 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.
- [34] G. Corrêa, 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.
- [35] G. Correia, P. Assuncao, L. Agostini, and L. A. da Silva Cruz, "A method for early-splitting of HEVC inter blocks based on decision trees," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2014, pp. 276–280.
- [36] X. Shen, L. Yu, and J. Chen, "Fast coding unit size selection for HEVC based on Bayesian decision rule," in *Proc. Picture Coding Symp. (PCS)*, May 2012, pp. 453–456.
- [37] Y. Zhang, S. Kwong, X. Wang, H. Yuan, Z. Pan, and L. Xu, "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.
- [38] Y. Yamamoto and T. Ikai, *AHG5: Fast QTBT Encoding Configuration*, document JVET-D0095, Oct. 2016.
- [39] Z. Wang, S. Wang, J. Zhang, and S. Ma, "Local-constrained quadtree plus binary tree block partition structure for enhanced video coding," in *Proc. Vis. Commun. Image Process. (VCIP)*, Nov. 2016, pp. 1–4.
- [40] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.
- [41] Q. Lu, X. Fang, and C. Wang, "Occlusion-adaptive bidirectional motion compensation for pixel-based motion vector field," *Electron. Lett.*, vol. 50, no. 7, pp. 514–516, Mar. 2014.
- [42] 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.
- [43] L. Shen, Z. Liu, Z. Zhang, and X. Shi, "Fast inter mode decision using spatial property of motion field," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1208–1214, Oct. 2008.
- [44] *JVET Software Repository*. [Online]. Available: [https://jvet.hhi.fraunhofer.de/svn/svn\\_HMJEMSoftware/branches/HM-13.0-QTBT/](https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/branches/HM-13.0-QTBT/)
- [45] *JVET Software Repository*. [Online]. Available: [https://jvet.hhi.fraunhofer.de/svn/svn\\_HMJEMSoftware/branches/HM-16.6-JEM-7.0-dev/](https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/branches/HM-16.6-JEM-7.0-dev/)
- [46] K. Suehring and X. Li, *Common Test Conditions and Software Reference Configurations*, document JVET-G1010, Joint Video Exploration Team (JVET), Jul. 2017.
- [47] *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33 ITU-T, Apr. 2001.
- [48] F. Bossen, *Common HM Test Conditions and Software Reference Configurations*, document JCTVC-F900, Joint Collaborative Team on Video Coding (JCT-VC), 2011.



**Zhao Wang** received the B.S. degree in software engineering from Fudan University, China in 2014. He is currently pursuing the Ph.D. degree in computer science with Peking University, Beijing, China. His research interests include video/image compression and processing, computer vision, and quality assessment.



**Shiqi Wang** (M'15) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008 and the Ph.D. degree in computer application technology from the Peking University in 2014. From 2014 to 2016, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada. From 2016 to 2017, he was a Research Fellow with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He has authored over 30 technical proposals to ISO/MPEG, ITU-T and AVS standards. His research interests include image/video compression, analysis and quality assessment.



**Jian Zhang** (M'14) received the B.Sc. degree from the Department of Mathematics, Harbin Institute of Technology ( HIT ), Harbin, China, in 2007, and the M.Eng. and Ph.D. degrees from the School of Computer Science and Technology, HIT, in 2009 and 2014, respectively. From 2014 to 2016, he was a Post-Doctoral Fellow with the Institute of Digital Media, Peking University, Beijing, China. He is currently a Post-Doctoral Fellow with the Visual Computing Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. His research interests include image/video compression and restoration, compressive sensing, and deep learning. He was a recipient of the Best Paper Award and Best Student Paper Award at the IEEE International Conference on Visual Communication and Image Processing, in 2011 and 2015, respectively.



**Shanshe Wang** received the B.S. degree from the Department of Mathematics, Heilongjiang University, Harbin, China, in 2004, the M.S. degree in computer software and theory from Northeast Petroleum University, Daqing, China, in 2010, and the Ph.D. degree in computer science from the Harbin Institute of Technology. He currently holds a post-doctorate position with Peking University. His current research interests include video compression, image and video quality assessment.



**Siwei Ma** (S'03–M'12) received the B.S. degree from Shandong Normal University, Jinan, China, in 1999, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005. From 2005 to 2007, he held a post-doctorate position with the University of Southern California, Los Angeles, USA. Then, he joined the Institute of Digital Media, School of Electronic Engineering and Computer Science, Peking University, Beijing, where he is currently a Professor. He has authored over 100 technical articles in refereed journals and proceedings in the areas of image and video coding, video processing, video streaming, and transmission.