Hybrid All Zero Soft Quantized Block Detection for HEVC

Jing Cui, Ruiqin Xiong, *Senior Member, IEEE*, Xinfeng Zhang, Shiqi Wang, Shanshe Wang, Siwei Ma, *Senior Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*— Transform and quantization account for a considerable amount of computation time in video encoding process. However, there are a large number of discrete cosine transform coefficients which are finally quantized into zeros. In essence, blocks with all zero quantized coefficients do not transmit any information, but still occupy substantial unnecessary computational resources. As such, detecting all-zero block (AZB) before transform and quantization has been recognized to be an efficient approach to speed up the encoding process. Instead of considering the hard-decision quantization (HDQ) only, in this paper, we incorporate the properties of soft-decision quantization into the AZB detection. In particular, we categorize the AZB blocks into genuine AZBs (G-AZB) and pseudo AZBs (P-AZBs) to distinguish their origins. For G-AZBs directly generated from HDQ, the sum of absolute transformed difference-based approach is adopted for early termination. Regarding the classification of P-AZBs which are generated in the sense of rate-distortion optimization, the rate-distortion models established based on transform coefficients together with the adaptive searching of the maximum transform coefficient are jointly employed for the discrimination. Experimental results show that our algorithm can achieve up to 24.16% transform and quantization time-savings with less than 0.06% RD performance loss. The total encoder time saving is about 5.18% on average with the maximum value up to 9.12%. Moreover, the detection accuracy of larger TU sizes, such as 16 × 16 and 32 × 32 can reach to 95% on average.**

***Index Terms*— DCT, all zero block (AZB) detection, soft-decision quantization, rate-distortion modeling.**

I. Introduction

THE state-of-the-art video coding standard High Efficiency Video Coding (HEVC) [1] jointly developed by the ISO/IEC MPEG and ITU-T, has achieved significant coding efficiency improvement compared with H.264/AVC [2]. The superior coding performance originates from the advanced coding tools adopted in HEVC, including quad-tree partition, extended-size discrete cosine transform (DCT) and additional discrete sine transform (DST) [3], etc. DCT is a vital but time-consuming module in HEVC, due to the computational burden of transform and the exhaustive rate-distortion optimization (RDO) process. However, in the case that one transform unit (TU) including non-zero residual data is identified as zero block after forward transform and quantization, there will be no subsequent entropy coding for RDO. In practice, there are a multitude of TUs to be quantized into all zero coefficients, especially for the small size TUs such as 4×4 and 8×8 in the low bit rate coding scenario. Thus, the all zero block (AZB) early determination ahead of transform tends to be useful for saving encoding time caused by redundant transform and quantization.

In literature, there have been a number of all zero block detection technologies [4]–[27] proposed for H.264/AVC and HEVC, aiming to accurately detect the AZB early with low complexity. Similar to early determination of CU splitting algorithms which are implemented with very different ways, such as [28] and [29], various AZB detection methods are designed to save encoding time in RDO process as well. Xuan *et al.* [4] derived the upper bound of the sum of the absolute difference (SAD), and determined the block as AZB if the SAD of the coding block is below the upper bound. The threshold value of the upper bound is theoretically derived based on the relationship between SAD and DCT coefficients. The similar idea based on SAD can also be found in [5]–[16]. In [5], the quantization parameter (QP) has been introduced to assist the threshold design. The detection method proposed in [17] refined the SAD threshold condition depending on the position of DCT coefficients to improve the detection accuracy. In [18], more sufficient and specific conditions are derived for three types of transforms, including integer 4×4 DCT for all the 4×4 blocks, Hadamard transform for 4×4 Luma DC coefficients and intra 16×16 blocks, and Hadamard transform for 2×2 Chroma DC coefficients.

Another method to improve detection accuracy with the consideration of frequency characteristics has been proposed in [19], where the sum of the squares is used to check AC energy of the residual data. However, the sum of squares should be calculated with multiplications which increase the computation burden in AZB detection in turn. Wang *et al.* [20], [21] further improved the AZB detection in [18] by utilizing a hybrid model for zero quantized

Manuscript received July 31, 2017; revised January 12, 2018 and April 12, 2018; accepted April 26, 2018. Date of publication May 16, 2018; date of current version July 9, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61632001, in part by the National Basic Research Program of China (973 Program) under Grant 2015CB351800, and in part by the Top-Notch Young Talents Program of China. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Catarina Brites. *(Corresponding author: Siwei Ma.)*

J. Cui, R. Xiong, S. Wang, S. Ma, and W. Gao are with the School of Electronic Engineering and Computer Science, Institute of Digital Media, Peking University, Beijing 100871, China, and also with the National Engineering Laboratory for Video Technology, Peking University, Beijing 100871, China (e-mail: jingcui106@pku.edu.cn; rqxiong@pku.edu.cn; sswang@pku.edu.cn; swma@pku.edu.cn; wgao@pku.edu.cn).

X. Zhang is with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: xinfengz@usc.edu).

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

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.2018.2837351

DCT coefficients to further reduce redundant computations. However, the AZB detection conditions derived by above methods cannot be applied straightforwardly when the Hadmamard transform is enabled in H.264/AVC coding, such that the sum of absolute transformed different (SATD) is used instead of SAD in [22]. In addition, the AZB detection in [23] utilized the Hadamard coefficients to evaluate the zero blocks and there is hardly any extra computation cost since the Hadamard coefficients can already be obtained in the prediction stage. However, the detection accuracy may suffer from the difference between Hadmard and DCT transform.

Regarding HEVC, larger transform block sizes are employed compared with H.264/AVC, such as 16×16 and 32×32, which comprise more complex and diverse content characteristics. This brings new challenges to AZB detection as well, since the AZB detection methods in H.264/AVC are not appropriate to be straightforwardly applied in HEVC. In view of this, several AZB early detection approaches have been designed for HEVC structure. In [24], the genuine zero block (GZB), which represents the blocks identified as AZB right after hard-decision quantization (HDQ), is detected by extending the method in [23]. For larger TU sizes, i.e., 16×16 and 32×32, the AZB is examined through the DC coefficients of each 8×8 Hadmard transformed sub-block, and these DC coefficients are transformed by 2×2 and 4×4 Hadamard transform again. Moreover, Lee *et al.* [24] proposed a method to determine the pseudo zero blocks (PZBs), which denote the blocks quantized as non-zero blocks are finally forced to be AZBs by RDO. Although the AZB detection rate has been increased by combining multiple 8×8 Hadmard transform matrices to detect AZB for 16×16 and 32×32, it decreases the effectiveness for small TUs, and has the limitations in practice due to the empirical values introduced in the RD cost calculation. Wang *et al.* [25] provided an efficient AZB detection method for HEVC with sufficient conditions. However, this method only targets at the 4×4 block size. Recently, state-of-the-art AZB detection methods were proposed in [26] and [27]. In [26], the upper and lower bounds of SAD and one SATD threshold were used to detect GZB. For PZB, a fast rate-distortion cost estimation scheme was proposed in order to improve the detection rate. However, several empirical values were introduced in the GZB detection and the proposed distortion and rate estimation models in RDO for PZB determination tend to be sensitive to the initial values. The AZB detection proposed in [27] introduced the idea that the DCT coefficients can be approximated by multiplying the sparse matrix with the Walsh Hadamard transform (WHT) matrix. Based on the SATD and WHT coefficients, the AZB detection works without performing actual DCT and quantization. However, the detection rates for the 16×16 and 32×32 block size are not satisfactory since the approximated coefficients cannot well match the DCT coefficients.

For soft-decision quantization (SDQ), the final quantized coefficients are not only dependent on how the coefficients are quantized, but also rely on how these coefficients are entropy coded. The quantized coefficient becomes the free parameter to be optimized with the RDO, which provides more flexibilities in the AZB detection process. Therefore, it is difficult to directly infer the AZB conditions from the block features such as residual energy or SATD. Recently, Yin *et al.* [30] proposed an detection method based on more accurate zero-quantized deadzone offset model in Rate Distortion Optimized Quantization (RDOQ). Our previous work [31] also presented a detection method based on RDOQ. However, Although the characteristics of AZB after RDOQ were taken into account, research on the relationship between AZB and RD cost is still required. In this paper, we make further efforts to solve the AZB detection problem by jointly considering the HDQ and SDQ. Specifically, the conditions of AZB for HDQ are firstly derived, followed by the strategies for SDQ to further identify the blocks that are required to be quantized to AZB in the sense of RDO. In particular, in the identification of AZB with SDQ, the maximum transform coefficient amplitude is detected within the low frequency part of TU, and the rate and distortion estimation models are established based on the transform coefficients and SATD of blocks to further detect the AZB in the RDO framework.

The remainder of this paper is organized as follows. In Section II, the relative works on AZB detection are reviewed. The proposed AZB detection scheme is introduced in Section III, and the experimental results are provided in Section IV. Finally, Section V concludes this paper.

II. Relative Works

Generally speaking, many block level features have been adopted as criterions to determine the AZB, such as residual energy, SAD and SATD of the given residual block. In particular, for the residual data $z(x, y)$ in the given $N \times N$ transform block, the transform coefficient $Z(u, v)$ can be described as follows based on the integer DCT transforms,

$$Z(u,v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} z(x,y) \cdot A(x,u) \cdot A(y,v) \tag{1}$$

where,

$$A(m,n) = \left\lceil \sqrt{\frac{2}{N}} \cdot a(m) \cdot \cos\left(\frac{2n+1}{2N} \cdot m\pi\right) \right\rceil \tag{2}$$

$$m, n = 0, 1, \ldots, N-1$$

and

$$a(m) = \begin{cases} 1/\sqrt{2} & m = 0 \\ 1 & m = 1, \ldots, N-1 \end{cases} \tag{3}$$

In the Dead Zone plus Uniform Threshold Quantization (DZ+UTQ), the transform coefficients $Z(u, v)$ are quantized as $L(u, v)$,

$$L(u,v) = sign(Z(u,v)) \left(|Z(u,v)| \cdot M_{QP/6} + offset \right) \gg Q_{bits} \tag{4}$$

where $M_{QP/6}$ is the multiplication factor equal to $2^{Q_{bits}} / Q_{step}$, and Q_{step} denotes the quantization step which is associated

with the quantization parameter (QP). In practical implementation, the DZ+UTQ can be formulated by Eq.(4) with the $M_{QP/6}$, *offset*, and Q_{bits} defined as:

$$M_{QP/6} = \{26214, 23302, 20560, 18396, 16384, 14564\},$$

$$offset = \begin{cases} 171 \ll (Q_{bits} - 9), & for\ I\ slice \\ 85 \ll (Q_{bits} - 9), & for\ P/B\ slice, \end{cases}$$

$$Q_{bits} = 29 + \mathrm{QP}/6 - bitDepth - \log_2 N, \tag{5}$$

where $bitDepth$ is the bit depth of input image pixel and N denotes the TU width. If the block is detected as AZB, it implies all the quantized coefficients $L(u, v)$ should satisfy:

$$|L(u, v)| < 1. \tag{6}$$

Then we have

$$|Z(u, v)| < Th(u, v), Th(u, v) = \frac{2^{Q_{bits}} - offset}{M_{QP/6}}, \tag{7}$$

where $Th(u, v)$ denotes the threshold of AZB. In the previous work, SAD in the residual domain is frequently used to detect the AZB. In particular, the detection condition is defined as:

$$SAD < \Gamma\left(Th(u, v)\right), \tag{8}$$

where $\Gamma(\cdot)$ denotes the function based on $Th(u, v)$ and it has been derived in various fashions [20], [25], [32]. However, the SAD value implies the residual difference in the residual domain, which cannot directly reflect the AZB characteristics. As such, in order to approximate the DCT coefficients in frequency domain, the Hadamard transform is applied to substitute the integer DCT transform such that the SATD from the Hadamard transform naturally becomes an alternative feature for AZB detection. However, although the SATD-based detection is able to improve the detection accuracy to some extent, previous works [24], [26] indicate that it may suffer from the undesirable detection rate for larger TU sizes due to more diverse content characteristics within one TU.

III. The Proposed AZB Detection

In this section, we propose the AZB detection scheme for all sizes of TUs based on both mathematical derivations and empirical analyses. In HEVC reference software, the SDQ strategy-Rate Distortion Optimized Quantization (RDOQ) is performed by default after transform. In contrast with DZ+UTQ, RDOQ is able to determine the optimal quantized transform coefficients from the perspective of coding performance, which may result in some transform blocks which are non-zero blocks after DZ+UTQ actually quantized to AZB by RDOQ. Therefore, in this work, the AZBs are categorized into genuine all zero block (G-AZB) and pseudo all zero block (P-AZB) according to different quantization strategies, as shown in Fig. 1. In particular, the G-AZB denotes the block quantized to all zero coefficients after DZ+UTQ. Otherwise, if the non-zero coefficients in the transform block are further quantized to all zero coefficients by RDOQ, such block is

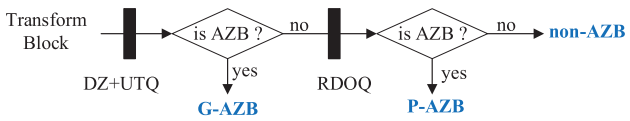


Fig. 1. The workflow of G-AZB and P-AZB detection.

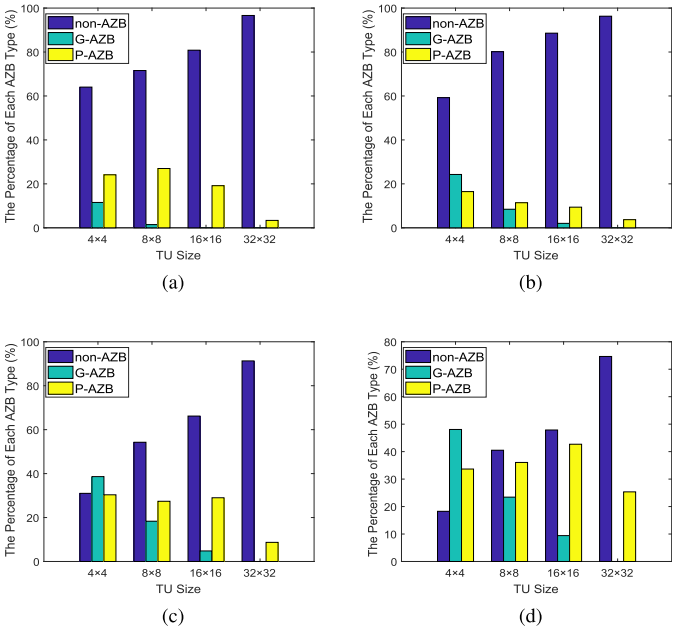


Fig. 2. AZB distributions for each TU layer in different QPs for sequence *BasketballDrive* under Random Access (RA). (a) QP = 12. (b) QP = 22. (c) QP = 32. (d) QP = 42.

regarded as P-AZB. The AZB distributions of each type for different TU layers are illustrated in Fig. 2. Obviously, we can see that the percentages of AZB blocks including G-AZB and P-AZB types increase monotonously with QP. Moreover, for the larger TUs such as 32×32 TU, there is hardly any G-AZB even for high QP cases. As such, in 32×32 TU, almost all the AZBs are P-AZBs, which implies there is slight possibility that the 32×32 TU is finally quantized to AZB without RDOQ. In other words, RDOQ contributes significantly to determine AZB blocks in this scenario. This motivates us to deal with G-AZB and P-AZB sequentially. In addition, as the residual coefficients of intra block tend to be larger even under the large QP values, here we only focus on the TUs from inter prediction.

A. Transform Specification

DCT has been proved to be an efficient transform method used in video/image compression, but the computational complexity is much higher than Hadamard transform. To reduce the computational cost of DCT, for the 8×8 and 4×4 transform blocks, the DCT transform cores are able to be simulated with Walsh-ordered Hadamard transform core and the sparse matrix A [27] with:

$$DCT_N = \frac{1}{\sqrt{N}} A_N H_{w,N}. \tag{9}$$

Specifically, sparse matrix A is defined as:

$$A_8 = \begin{bmatrix} 64 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 58 & 0 & 24 & 0 & -5 & 0 & 12 \\ 0 & 0 & 59 & 0 & 0 & 0 & 25 & 0 \\ 0 & -20 & 0 & 49 & 0 & 33 & 0 & 14 \\ 0 & 0 & 0 & 0 & 64 & 0 & 0 & 0 \\ 0 & 14 & 0 & -33 & 0 & 49 & 0 & 20 \\ 0 & 0 & -25 & 0 & 0 & 0 & 59 & 0 \\ 0 & -12 & 0 & -5 & 0 & -24 & 0 & 58 \end{bmatrix} \tag{10}$$

$$A_4 = \begin{bmatrix} 16 & 0 & 0 & 0 \\ 0 & 15 & 0 & 6 \\ 0 & 0 & 16 & 0 \\ 0 & -6 & 0 & 15 \end{bmatrix} \tag{11}$$

Considering $H_{w,N}$ is built based upon 1 and -1, only additions and subtractions are required in the computation. In addition, as there are a number of zeros in A_N, the calculation burden increase is marginal. For larger TU sizes, i.e., 16×16 and 32×32, DCT transform is used. Since different transforms are utilized for various TU sizes, the generalized sum of absolute transformed difference $SATD_T$ is introduced in the following sections as the substitution for SATD where the subscript T denotes the transform core.

B. G-AZB Detection

For G-AZB, SAD and $SATD_T$ carry important information that can be used to as the summary for the TU characteristics. Here, we adopt a mathematical derivation based method, and we want to emphasize that our scheme is advantageous over the previous works such as [26] in several ways by utilizing more strict threshold. Firstly, we use more efficient transform by DCT-like matrix for 4×4 and 8×8 TU instead of using Hadamard transform. Secondly, more strict conditions are introduced in the mathematical derivation of the threshold.

In theory, all transform coefficients in the given TU should satisfy the constraint in Eq.(7) if it is detected as G-AZB. Then we have:

$$\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}|Z(u,v)| < \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}Th(u,v), \tag{12}$$

which can be further derived as follows,

$$SATD_T < \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\left(\frac{2^{Q_{bits}}-offset}{M_{QP/6}}\right) = N^2\left(\frac{2^{Q_{bits}}-offset}{M_{QP/6}}\right). \tag{13}$$

Therefore, one threshold $\Gamma^1_{SATD_T}$ can be obtained as,

$$\Gamma^1_{SATD_T} = N^2\left(\frac{2^{Q_{bits}}-offset}{M_{QP/6}}\right). \tag{14}$$

For the prediction residuals, the Laplacian distribution is used to model the residual data distribution, and then the corresponding relationship between residuals and transform coefficients can be derived. The Laplacian distribution is formulated as follows

$$p(x) = \frac{1}{2b}e^{-\frac{|x|}{b}}, \tag{15}$$

where σ is the standard deviation of residual data x and b is the scale parameter. The expected value of $|x|$ can be deduced as:

$$E[|x|] = \int_{-\infty}^{+\infty}|x|\cdot\frac{1}{2b}e^{-\frac{|x|}{b}}d_x = b, \tag{16}$$

with,

$$b = \frac{\sigma}{\sqrt{2}}.$$

Since

$$SAD = \sum_{u=0}^{N-1}\sum_{v=0}^{N-1}|x|, \tag{17}$$

and

$$E[|x|] \approx SAD/N^2, \tag{18}$$

we can have the following relationship,

$$\sigma \approx \sqrt{2}\cdot\frac{SAD}{N^2} \tag{19}$$

The variance of transform coefficients in $(u,v)^{th}$ position can be calculated according to [33], [34] as follows,

$$\sigma^2_{dct}(u,v) = \frac{\sigma^2}{N^2}\cdot uv, \tag{20}$$

with,

$$uv = \left[AH_wCA^TH_w^T\right]_{u,u}\left[AH_wCA^TH_w^T\right]_{v,v},$$

where $[\cdot]_{u,u}$ denotes the element in $(u,u)^{th}$ position and C is the relevance metric,

$$C = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{bmatrix}. \tag{21}$$

Here, ρ is the correlation coefficient. In particular, larger ρ value indicates that the input signal shares more homogenous visual content. Generally, ρ is set to be 0.6 following published works [20], [22], [35], [36]. As such, the standard derivation in $(u,v)^{th}$ position can be obtained by combining Eq.(19) and Eq.(20):

$$\sigma_{dct}(u,v) = \frac{\sqrt{2}}{N^3}\cdot SAD\cdot\sqrt{uv}. \tag{22}$$

It is has been proved that there is a linear relationship between SAD and $SATD_T$ [22], and then we have:

$$SAD = \kappa\cdot SATD_T, \tag{23}$$

where κ is defined as:

$$\kappa = \frac{1}{2N^2}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}\sqrt{uv}. \tag{24}$$

Thus, by substituting SAD in Eq.(22) with Eq.(23), $\sigma_{dct}(u,v)$ can be represented as:

$$\sigma_{dct}(u,v) = \frac{\sqrt{2}\kappa}{N^3} \cdot SATD_T \cdot \sqrt{uv}. \tag{25}$$

The $\sigma_{dct}(u,v)$ denotes the standard deviation in the band (u,v) and it can be used to characterize the transform coefficients based on the hypothesis that transformed coefficients obey Laplacian distribution [37]. According to the theory of probability, the probability of absolute of transform coefficient which is less than $3\sigma_{dct}(u,v)$ approximately accounts for 99.7%. Therefore, it is reasonable to have,

$$Z(u,v) \leq 3\sigma_{dct}(u,v) \tag{26}$$

and

$$\max_{u,v\in[0,N-1]} (\sigma_{dct}(u,v)) < Th(u,v)/3. \tag{27}$$

By combining Eq.(25) and Eq.(27), we can have,

$$SATD_T < \frac{N^3 \cdot Th(u,v)}{3\sqrt{2}\kappa \max\limits_{u,v\in[0,N-1]} \left(\sqrt{uv}\right)}. \tag{28}$$

Therefore, the other threshold $\Gamma^2_{SATD_T}$ can be derived as,

$$\Gamma^2_{SATD_T} = \frac{N^3}{3\sqrt{2}\kappa \max\limits_{u,v\in[0,N-1]} \left(\sqrt{uv}\right)} \cdot \frac{2^{Q_{bits}} - offset}{M_{QP/6}}. \tag{29}$$

Finally, the SAD and $SATD_T$ based G-AZB detection threshold can be summarized as:

$$SATD_T < \min_{u,v\in[0,N-1]} \left\{ \Gamma^1_{SATD_T}, \Gamma^2_{SATD_T} \right\}. \tag{30}$$

C. P-AZB Detection

P-AZB detection is applied to detect the blocks which are quantized to AZBs via RDOQ. In other words, when only DZ+UTQ is applied, these blocks will not be quantized into AZBs. Therefore, P-AZB detection is performed right after G-AZB detection when the conditions in G-AZB detection are not satisfied.

Since accurate P-AZB detection method contributes significantly for the larger TU blocks, here we further investigate the scenario of P-AZB detection for larger TU sizes. In particular, the coefficient distributions for the larger TU, i.e., 16×16, 32×32, are explored in Fig. 3 with absolute quantized coefficient values being 1, 2 or larger than 2 for AZBs and non-AZBs. For the AZBs, the maximum coefficient amplitude tends to be smaller than 2 under different QPs. As such, the coefficient magnitude is an efficient clue to determine non-AZBs. Thus, the AZB detection can be performed based on searching transform coefficient amplitudes in the given TU. However, for large TUs, to avoid time-consuming operation traversing all coefficients, we propose a new method to detect the maximum transform coefficients based on the frequency band.

Before performing RDOQ, the temporal level termed as $l(u,v)$ is pre-calculated based on DZ+UTQ irrespective of frame type. In RDOQ, there are several candidate levels for each $l(u,v)$, and the optimal one among these candidates is determined by RDO. Table I lists the available candidates for each $l(u,v)$. In particular, for the case of $l(u,v)$ being 0, the final quantized coefficient will be definitely zero even without RDOQ. However, only $l(u,v)$ being 1 or 2 is able to be adjusted to 0, which means if the given TU is AZB after RDOQ, the prerequisite is that the $l(u,v)$ of all coefficients need to be within 2. Fig. 4 to Fig. 7 illustrate the $l(u,v)$ distributions with different QPs for each TU layers in these P-AZB TUs. We can observe that all the non-zero $l(u,v)$ are not larger than 2. The statistical results suggest that only all the $l(u,v)$ are within the range [0, 2], the TU can be encoded as P-AZB. This motivates us to search for the largest coefficient in the TU for the identification of P-AZB.

To reduce the coefficients searching burden, we divide each TU into two parts: low frequency part and high frequency part. In essence, the boundary for low and high frequency parts balances the searching complexity and detection accuracy. For high bit rate coding, more non-zero $l(u,v)$ are generated, such that our detection area should be larger than that in the

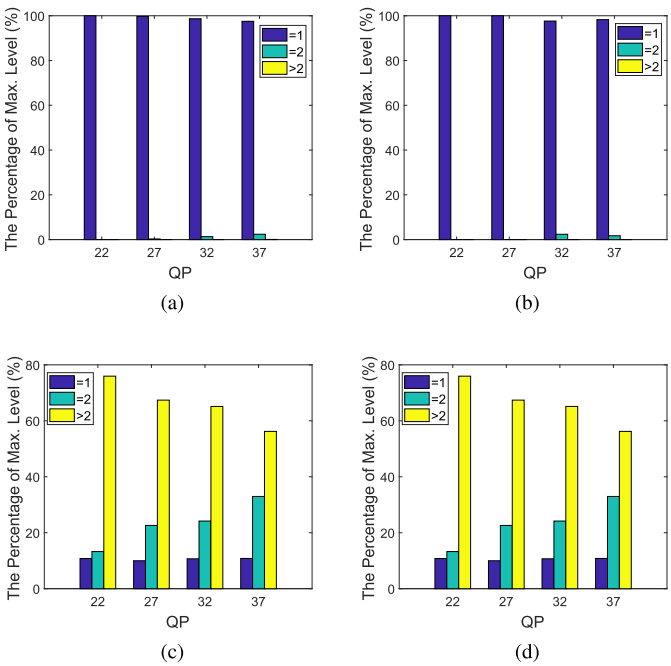
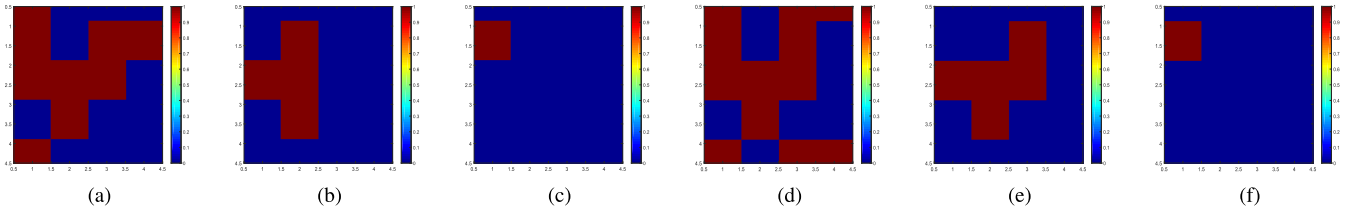


Fig. 3. Coefficients distributions in AZB and non-AZB of 16×16 and 32×32 TUs for *BasketballDrive* under RA. (a) AZB 16×16. (b) AZB 32×32. (c) non-AZB 16×16. (d) non-AZB 32×32.

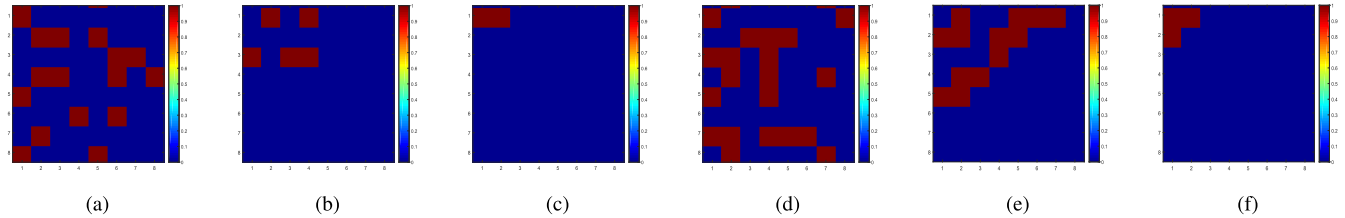
TABLE I
THE CANDIDATES OF $l(u,v)$ IN RDOQ

$l(u,v)$	*Candidates*
0	0
1	0, 1
2	0, 1, 2
3	2, 3
...	...
N	N-1, N



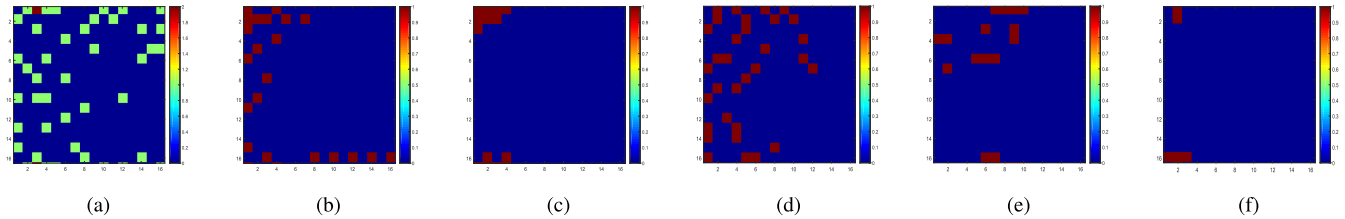
(a) (b) (c) (d) (e) (f)

Fig. 4. The examples of $l(u,v)$ distribution map in the 4×4 P-AZB. (a) *BasketballPass* QP = 22. (b) *BasketballPass* QP = 32. (c) *BasketballPass* QP = 42. (d) *Cactus* QP = 22. (e) *Cactus* QP = 32. (f) *Cactus* QP = 42.



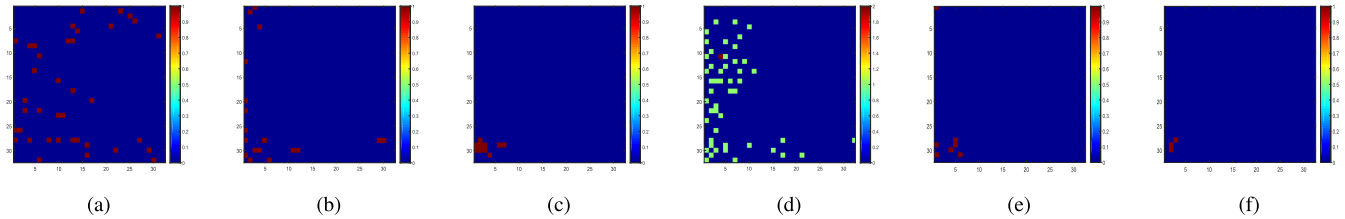
(a) (b) (c) (d) (e) (f)

Fig. 5. The examples of $l(u,v)$ distribution map in the 8×8 P-AZB. (a) *BasketballPass* QP = 22. (b) *BasketballPass* QP = 32. (c) *BasketballPass* QP = 42. (d) *Cactus* QP = 22. (e) *Cactus* QP = 32. (f) *Cactus* QP = 42.



(a) (b) (c) (d) (e) (f)

Fig. 6. The examples of $l(u,v)$ distribution map in the 16×16 P-AZB. (a) *BasketballPass* QP = 22. (b) *BasketballPass* QP = 32. (c) *BasketballPass* QP = 42. (d) *Cactus* QP = 22. (e) *Cactus* QP = 32. (f) *Cactus* QP = 42.



(a) (b) (c) (d) (e) (f)

Fig. 7. The examples of $l(u,v)$ distribution map in the 32×32 P-AZB. (a) *BasketballPass* QP = 22. (b) *BasketballPass* QP = 32. (c) *BasketballPass* QP = 42. (d) *Cactus* QP = 22. (e) *Cactus* QP = 32. (f) *Cactus* QP = 42.

low bit rate scenario. Therefore, the boundary should be adaptive to the QP values. Here, we proposed a deterministic scheme for the identification of the block for the low frequency part, which locates in the top left corner of TU, as shown in Fig. 8,

$$B_{LF} = \max_{\substack{QP\in(0,\ldots,maxQP)\\ N\in[4,8,16,32]}} \left\{ round\left((1-\frac{QP}{\max QP})\cdot N \right), 1 \right\}, \quad (31)$$

where B_{LF} is the block size of low frequency part. $maxQP$ denotes the maximum QP setting in HEVC and QP is the quantization parameter for current TU.

Considering the fact that only the $l(u,v)$ of 1 or 2 in TU can be further quantized to zero, and moreover large amplitude coefficients always concentrate in the low frequency part, we only focus on the low frequency part and guarantee the maximum absolute $l(u,v)$ within it to be smaller than 2. To reduce computational cost of DZ+UTQ, we construct our detection threshold based on the transform coefficients instead

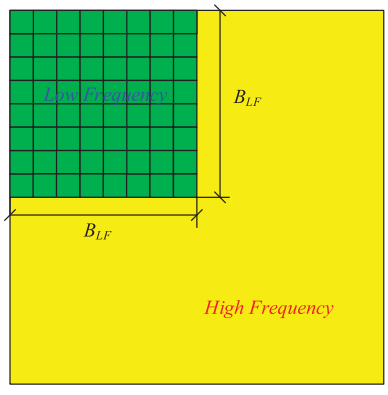


Fig. 8. Illustration of the low and high frequency partitions.

of $l(u,v)$, and thus we have,

$$\max_{\substack{u,v\in[0,B_{LF}]\\ \varepsilon\in[0,0.5)}} \{Z(u,v)\} \geq \frac{(MUL+\varepsilon)\cdot 2^{Q_{bits}} - offset}{M_{QP/6}}, \quad (32)$$

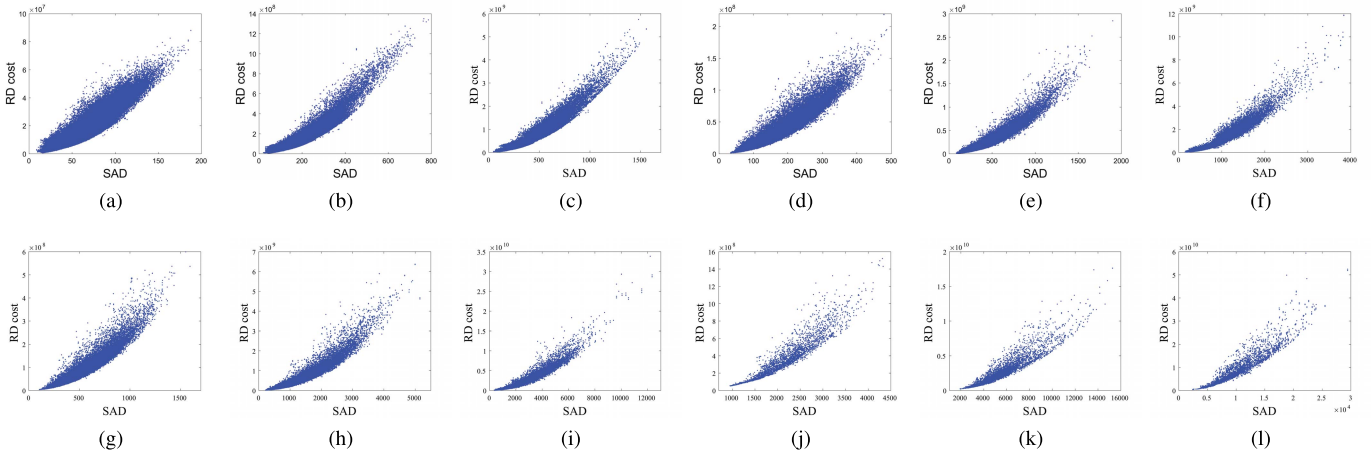


Fig. 9. The relationship between RD Cost and SAD in 16×16 and 32×32 AZB TUs for *BasketballPass*. (a) 4×4 QP $= 22$. (b) 4×4 QP $= 32$. (c) 4×4 QP $= 42$. (d) 8×8 QP $= 22$. (e) 8×8 QP $= 32$. (f) 8×8 QP $= 42$. (g) 16×16 QP $= 22$. (h) 16×16 QP $= 32$. (i) 16×16 QP $= 42$. (j) 32×32 QP $= 22$. (k) 32×32 QP $= 32$. (l) 32×32 QP $= 42$.

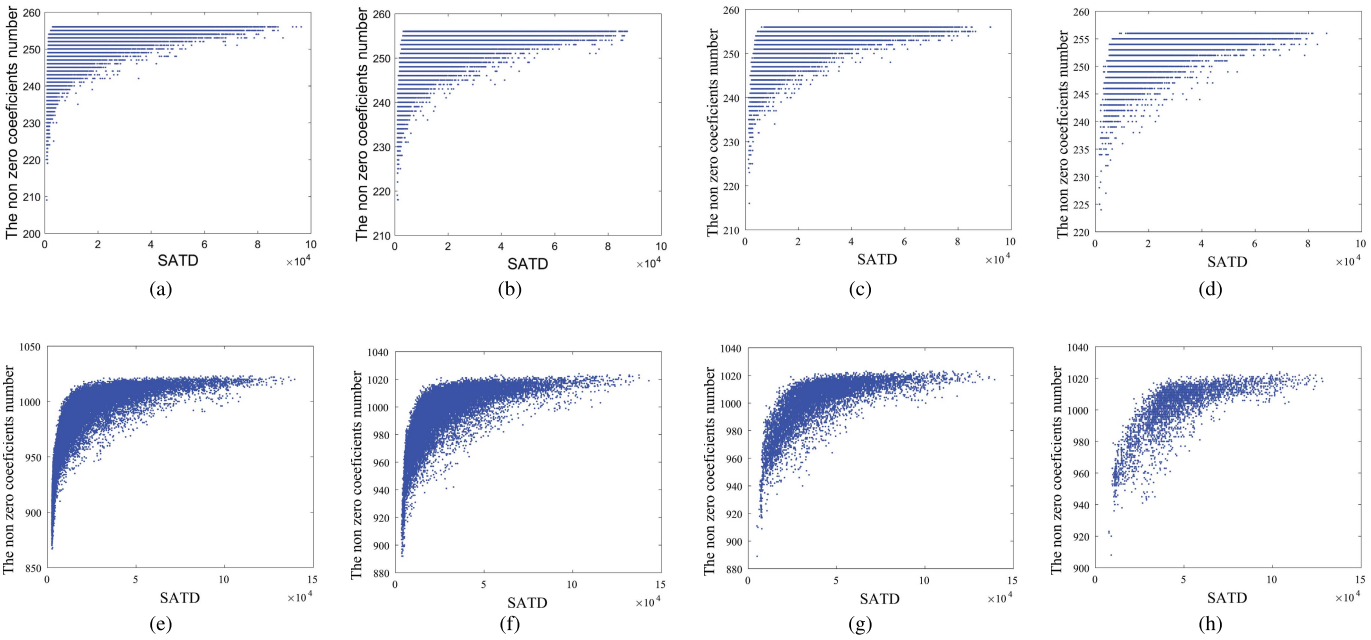


Fig. 10. The relationship between the number of non-zero transform coefficients *nonZeroNum* and $SATD_T$ in 16×16 and 32×32 AZB TU for *BasketballPass*. (a) 16×16 QP $= 12$. (b) 16×16 QP $= 22$. (c) 16×16 QP $= 32$. (d) 16×16 QP $= 42$. (e) 32×32 QP $= 12$. (f) 32×32 QP $= 22$. (g) 32×32 QP $= 32$. (h) 32×32 QP $= 42$.

where MUL denotes the maximum $l(u, v)$ and ε is the compensating factor. In this method, we set MUL and ε as 2 and 0.2, respectively.

In addition, to further detect the AZBs which are generated by the RDOQ, we design the RDO-based threshold by comparing the RD costs of a TU when it is detected as AZB and non-AZB. In particular, the RD costs can be calculated by:

$$\begin{aligned} RDCost_{AZB} &= SSD_{AZB} + \lambda \times R_{AZB}, \\ RDCost_{non\text{-}AZB} &= SSD_{non\text{-}AZB} + \lambda \times R_{non\text{-}AZB}, \end{aligned} \tag{33}$$

where SSD_* and R_* denote the distortion and entropy bits, respectively. In particular, the R_{AZB} can be approximated to 1 since there is no coefficients needed to encode and only one bit flag is required for signal, as described in [1]. Therefore, we can have,

$$RDCost_{AZB} = SSD_{AZB} + \lambda. \tag{34}$$

Here, the SSD_{AZB} is calculated by:

$$SSD_{AZB} = ||\boldsymbol{r} - \boldsymbol{rr}|| = ||\boldsymbol{r}||^2, \tag{35}$$

where $\boldsymbol{r}$ and $\boldsymbol{rr}$ are the original and reconstructed residual data vector, respectively. For TUs including only zero coefficients after RDOQ, the $\boldsymbol{rr}$ is zero vector as well. As such, the $RDCost_{AZB}$ in Eq.(34) can be written as,

$$RDCost_{AZB} = ||\boldsymbol{r}||^2 + \lambda. \tag{36}$$

TABLE II
THE VALUES OF Th_c INDEXED BY QP AND TU SIZE

TU Size	QP 17	QP 22	QP 27	QP 32	QP 37	QP 42
4×4	0.9	0.88	0.89	0.71	0.67	0.64
8×8	4.26	4.12	4.05	3.92	3.71	3.68
16×16	18.53	18.34	18.06	18.79	18.29	18.11
32×32	78.21	77.72	77.78	78.85	73.48	74.02

TABLE III
THE VALUES OF α INDEXED BY QP AND TU SIZE

TU Size	QP 17	QP 22	QP 27	QP 32	QP 37	QP 42
4×4	0.3629	0.3072	0.2981	0.2346	0.2658	0.2433
8×8	0.3528	0.3315	0.2694	0.3255	0.3670	0.3721
16×16	0.3923	0.3820	0.4039	0.3257	0.2905	0.2766
32×32	0.3877	0.3560	0.4062	0.3104	0.3774	0.3457

TABLE IV
THE VALUES OF β INDEXED BY QP AND TU SIZE

TU Size	QP 17	QP 22	QP 27	QP 32	QP 37	QP 42
4×4	3.2789	2.9643	3.0172	3.2001	1.9866	2.1309
8×8	3.8817	4.3108	3.8922	3.1236	3.7435	3.4003
16×16	3.8506	3.7204	3.0726	3.1145	3.2756	3.2731
32×32	2.7496	2.1109	3.5112	3.1469	3.7692	2.7932

Moreover, as the residual data for inter prediction will be mostly 1 and 0, we can have,

$$RDCost_{AZB} = {r_0}^2 + \cdots + {r_{N^2-1}}^2 + \lambda \approx (|r_0| + \cdots + |r_{N^2-1}|) \times (|r_0| + \cdots + |r_{N^2-1}|) + \lambda = SAD^2 + \lambda. \quad (37)$$

The relationship between RD cost and SAD is shown in Fig. 9, where we can see that approximation of the RD cost with Eq.(37) is practical. To unify the RD cost at pixel level, we modify Eq.(37) by a constant c, which is a function of TU size N,

$$RDCost_{AZB} = (SAD^2 + \lambda)/c. \quad (38)$$

Here, c is set to be 17, 37, 148 and 580 which are the same as in [24] and [26] for 4×4, 8×8, 16×16 and 32×32 TUs, respectively. Combining Eq.(38) with Eq.(23), we can have:

$$RDCost_{AZB} = (\kappa^2 \cdot {SATD_T}^2 + \lambda)/c, \quad (39)$$

where the κ is obtained from the G-AZB detection.

For the RD cost of non-AZB, both $SSD_{non\text{-}AZB}$ and $R_{non\text{-}AZB}$ are derived. Assuming that the distortion $SSD_{non\text{-}AZB}$ can be obtained in the transform domain by original and reconstructed transform coefficients, we can have

$$SSD_{non\text{-}AZB} = s^2 \cdot Q_{step}^2 \cdot \sum_{i=0}^{N^2-1} \left(C_i - \bar{C}_i\right)^2, \quad (40)$$

TABLE V
THE EXAMPLE VALUES OF Th_{SATD_T} AND SLOPE φ INDEXED BY QP AND TU SIZE

TU Size		QP 22	QP 27	QP 32	QP 37
Th_{SATD_T}	16×16	3248	4603	6102	8108
	32×32	15450	17480	24090	28840
φ	16×16	0.01792	0.0126	0.00863	0.00658
	32×32	0.01046	0.00847	0.00682	0.00506

TABLE VI
DETECTION ACCURACY EVALUATION CRITERIONS

Notation	Full Name	Definition
TP	True Positive	AZB which is detected as AZB
TN	True Negative	non-AZB which is detected as non-AZB
FP	False Postive	non-AZB which is detected as AZB
FN	False Negative	AZB which is detected as non-AZB
TPR	True Positive Rate	TP/(TP + FN)
FNR	False Negative Rate	FN/(TP + FN)
FPR	False Positive Rate	FP/(FP + TN)
DA	Detection Accuracy	1 - (FP+FN)/(TP + TN + FP + FN)

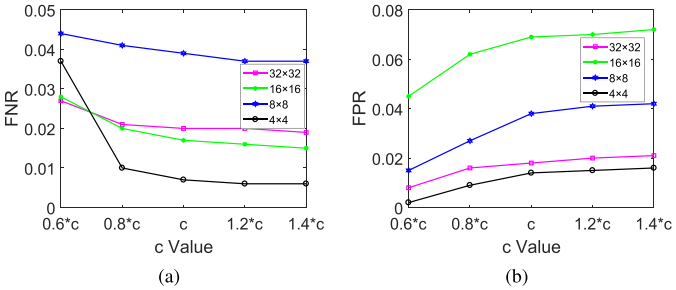


Fig. 11. Sensitivity of c regarding on the average Detection Rate of *Traffic* and *ParkScene* under RA. (a) *FNR*. (b) *FPR*.

where s is the scaling factor and C_i and $\bar{C}_i$ are denoted as the i-th quantized and reconstructed quantized coefficient matrices.

The quantized error $|C_i - \bar{C}_i|$ for i-th quantized coefficient tends to be small after RDOQ, such that we set a threshold value Th_c, which covers about 95.5% of $|C_i - \bar{C}_i|$ by learning from video sequences *BasketballPss, Cactus* with 100 frames,

$$|C_i - \bar{C}_i|^2 < Th_c. \quad (41)$$

The QP-based threshold value Th_c for each TU size can be summarized in Table II. By combining Eq.(40) and Eq.(41), we can have,

$$SSD_{non\text{-}AZB} = s^2 \cdot {Q_{step}}^2 \times \sum Th_c \quad (42)$$

For $R_{non\text{-}AZB}$, a rate estimation model is employed based on the transform coefficients as that in [38] and [39],

$$R_{non\text{-}AZB} = \alpha \cdot SATD_T + \beta \cdot nonZeroNum + \gamma, \quad (43)$$

with,

$$nonZeroNum = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} I\left(Z(u, v)\right)$$

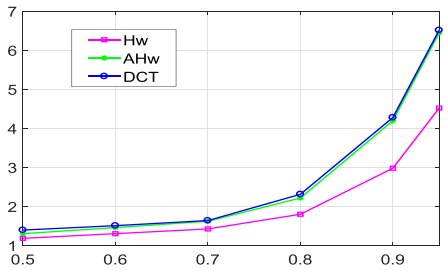


Fig. 12. The coding gain G_{TC} comparison for *BasketballPass* with QP 32 under RA.

and

$$I\left(Z(u,v)\right) = \begin{cases} 1, & if\ Z(u,v) = 0 \\ 0, & otherwise, \end{cases}$$

where the parameters α and β are obtained based on offline training, as shown in Table III and Table IV. The training scheme and method have been proposed in our previous work in [38]. γ is set as 0 since the TU will be skipped once all the transform coefficients are zero. Moreover, the relationship between *nonZeroNum* and $SATD_T$ can be well fitted by piecewise linear function based on the statistical results in Fig. 10, and Th_{SATD_T} is used to classify two parts of linear curve. φ is the slope between *nonZeroNum* and $SATD_T$. Thus the $R_{non\text{-}AZB}$ is modified as:

$$R_{non\text{-}AZB} = \alpha \cdot SATD_T + \beta \cdot \begin{cases} N^2, if\ SATD_T > Th_{SATD_T} \\ \varphi \times SATD_T, \quad else. \end{cases} \tag{44}$$

In particular, for 4×4 and 8×8 TU, the Th_{SATD_T} is approaching to zero, i.e., $nonZeroNum = N^2$. Then we have:

$$R_{non\text{-}AZB} = \alpha \cdot SATD_T + \beta \cdot N^2 \tag{45}$$

If the current TU is P-AZB, we can have the following relationship,

$$RDCost_{AZB} < RDCost_{non\text{-}AZB}. \tag{46}$$

Combining Eq.(33), Eq.(39), Eq.(42), and Eq.(44) with the case $SATD_T > Th_{SATD_T}$, we can formulate Eq.(46) as:

$$(\kappa^2 \cdot {SATD_T}^2 + \lambda)/c < s^2 \cdot Q_{step}^2 \times \sum Th_c + \lambda \cdot \left(\alpha \cdot SATD_T + \beta \cdot N^2\right). \tag{47}$$

The larger solution of this quadratic equation is:

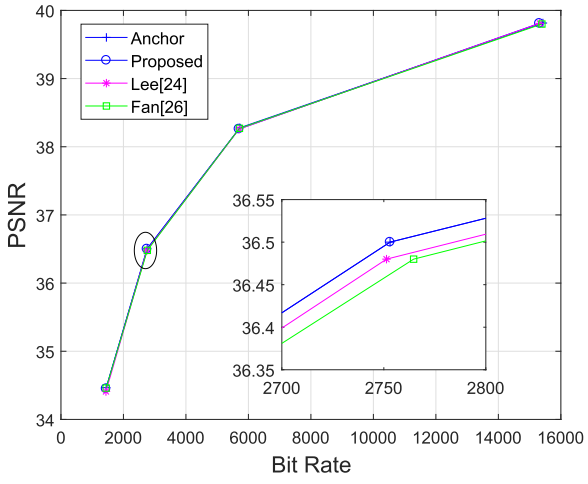
$${SATD_T}^{right} = \frac{\alpha \cdot \lambda \cdot c}{2\kappa^2} + \frac{\sqrt{\alpha^2 \cdot \lambda^2 \cdot c^2 - 4AC}}{2\kappa^2} \tag{48}$$

with,

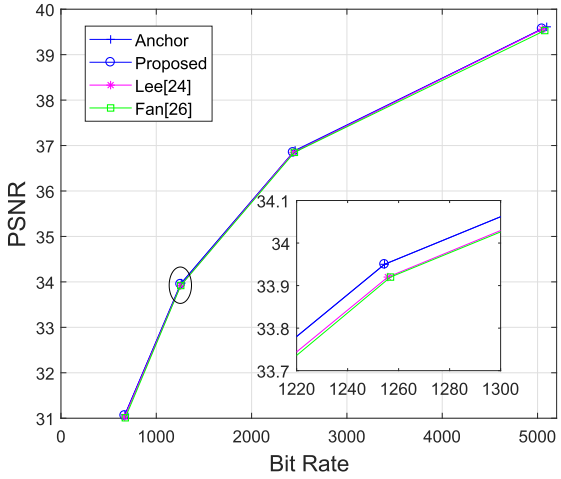
$$AC = \kappa^2 \cdot \left(\lambda - \beta \cdot \lambda \cdot c \cdot N^2 - s^2 \cdot Q_{step}^2 \cdot \sum Th_c\right).$$

Thus, we can get the left and right boundary of P-AZB as:

$$Th_{SATD_T} < SATD_T < \left\lfloor {SATD_T}^{right} + \frac{1}{2} \right\rfloor. \tag{49}$$



(a)



(b)

Fig. 13. The RD curves comparison for *BasketballDrive* and *BQMall*. (a) *BasketballDrive LD*. (b) *BQMall RA*.

Similarly, for the case $SATD_T \leq Th_{SATD_T}$, the smaller solution of this quadratic equation is:

$${SATD_T}^{left} = \frac{\lambda \cdot c \cdot (\alpha + \beta \cdot \varphi) - SQRT}{2\kappa^2} \tag{50}$$

with,

$$SQRT = sqrt\left(\begin{matrix} \lambda^2 \cdot c^2 \cdot (\alpha + \beta \cdot \varphi)^2 + 4\kappa^2 \cdot \\ (s^2 \cdot Q_{step}^2 \cdot c \cdot Th_c - \lambda) \end{matrix}\right). \tag{51}$$

Thus, the other limitation zone with left and right boundary can be described as:

$$\left\lfloor {SATD_T}^{left} + \frac{1}{2} \right\rfloor < SATD_T \leq Th_{SATD_T}. \tag{52}$$

Therefore, the final P-AZB detection can be transferred to detect whether the $SATD_T$ is located in the range formulated in Eq.(49) and Eq.(52).

D. AZB Detection Algorithm

The proposed AZB detection scheme with G-AZB and P-AZB detection can be summarized in Algorithm 1:

TABLE VII
PERFORMANCE EVALUATION OF THE PROPOSED AZB DETECTION METHOD

Sequence		RA					LD				
		G-AZB		G-AZB + P-AZB		TS	G-AZB		G-AZB + P-AZB		TS
		BDBR	TSq	BDBR	TSq		BDBR	TSq	BDBR	TSq	
A class	Traffic	0.08%	4.26%	-0.04%	25.67%	4.31%	0.06%	2.01%	-0.39%	22.53%	4.73%
	PeopleOnStreet	0.07%	7.60%	-0.16%	23.92%	4.01%	0.03%	3.74%	0.08%	19.81%	4.24%
B class	Kimono	0.04%	3.59%	0.12%	18.06%	3.74%	0.01%	1.51%	0.03%	22.79%	4.17%
	ParkScene	0.02%	2.11%	-0.06%	18.74%	3.82%	0.04%	7.23%	0.00%	38.21%	9.12%
	Cactus	0.11%	9.76%	0.13%	22.43%	3.66%	0.01%	3.03%	-0.01%	22.74%	4.64%
	BasketballDrive	0.04%	5.98%	0.08%	17.51%	3.46%	0.03%	4.21%	0.09%	26.93%	6.22%
	BQTerrace	0.02%	8.72%	0.09%	17.36%	3.17%	0.00%	0.76%	-0.13%	25.94%	6.28%
C class	BasketballDrill	0.02%	4.66%	0.07%	24.11%	4.84%	0.05%	6.22%	0.11%	31.76%	7.67%
	BQMall	0.01%	2.08%	0.03%	23.92%	4.30%	0.00%	1.49%	-0.01%	22.08%	4.94%
	PartyScene	0.01%	7.69%	0.13%	20.88%	4.95%	0.00%	2.98%	0.07%	28.72%	7.13%
	RaceHorses	0.00%	1.56%	0.04%	20.10%	3.07%	0.03%	5.84%	0.05%	15.01%	2.36%
D class	BasketballPass	0.05%	8.97%	0.24%	25.06%	4.71%	0.00%	0.79%	0.05%	19.65%	4.14%
	BQSquare	0.00%	7.62%	0.08%	37.11%	6.58%	0.03%	5.14%	0.14%	15.84%	7.05%
	BlowingBubbles	0.02%	2.99%	0.25%	21.25%	4.31%	0.00%	4.23%	0.11%	17.31%	4.27%
	RaceHorses	0.02%	4.06%	0.02%	21.73%	4.18%	0.06%	6.08%	0.07%	21.66%	4.98%
E class	FourPeople	0.03%	2.84%	-0.06%	15.82%	3.05%	0.04%	5.96%	0.06%	29.90%	5.25%
	Johnny	0.02%	0.76%	0.04%	26.91%	3.01%	0.03%	7.34%	0.10%	20.58%	3.33%
	KristenAndSara	0.03%	4.27%	-0.08%	19.77%	2.97%	0.05%	5.80%	-0.03%	23.07%	3.49%
F class	BasketballDrillText	0.11%	6.96%	0.13%	28.78%	4.45%	0.03%	3.73%	0.27%	31.60%	6.20%
	SlideEditing	0.13%	5.53%	-0.03%	25.66%	4.23%	0.09%	9.46%	0.15%	23.96%	3.91%
	SlideShow	0.13%	7.46%	0.21%	22.94%	3.45%	0.14%	10.57%	0.48%	27.33%	4.75%
Avg.		**0.05%**	**5.21%**	**0.06%**	**22.75%**	**4.01%**	**0.03%**	**4.67%**	**0.06%**	**24.16%**	**5.18%**

Algorithm 1 The Proposed AZB Detection Algorithm

Input: The residual data of the current TU.
Output: The current TU is $\mathbf{AZB}$ or Not.
if $\mathbf{SATD_T}$ satisfies $\mathbf{Eq.(30)}$ **then**
  $\mathbf{isAZB}$ equals to 1.
else if $\max_{QP\in(0,maxQP)}\{Z(u,v)\}$ satisfies $\mathbf{Eq.(32)}$ **then**
  $\mathbf{isAZB}$ equals to 0.
else if $\mathbf{SATD_T}$ satisfies $\mathbf{Eq.(49)}$ or $\mathbf{Eq.(52)}$ **then**
  $\mathbf{isAZB}$ equals to 1.
end if

IV. EXPERIMENTAL RESULTS

To validate the accuracy and efficiency of the scheme, we implement it in HEVC reference software HM 16.9. The performance of the proposed algorithm is evaluated based on test sequences in HEVC common test conditions [40] (Class A - Class F) under Random Access (RA) and Low Delay B (LD) main profile configurations with QPs being 22, 27, 32 and 37. In this work, before showing the algorithm performance, we evaluate the model parameters c and ρ. And then the R-D performance and computational complexity of the proposed scheme are presented, followed by the comparisons with state-of-the-art methods in terms of R-D performance, computational complexity and detection accuracy.

A. Model Parameters Evaluation

In this section, we firstly evaluate the parameters sensitivity of c in Eq.(38) and Eq. (39), and the correlation coefficient ρ in Eq.(21). The parameter c values for different TU sizes are derived from off-line training process by fitting the relationship between the RD cost and SAD for practical compressed video sequences, which is introduced in [24] and [26], and we use the same c values with them. The sensitivity of c regarding on the accuracy of AZB detection method has also been verified as shown in Fig.11. Table VI lists the explanations of the criterions which are used to describe the detection accuracy. When the c value becomes larger, there is a slight decrease on FNR and increase on FPR, respectively, since c value has an effect on the $RDCost_{AZB}$. Larger c will reduce the $RDCost_{AZB}$ and make more non-AZB blocks be detected as AZB resulting in the FPR increase. Similarly, when c value is smaller, there are many AZB blocks will be misclassified into non-AZB resulting in FNR increase.

The parameter ρ denotes the correlation coefficient of adjacent two pixels. To explore the influence of the parameter ρ on the coding performance, we have tested the transform coding gain G_{TC} for different transform kernels with different ρ. The transform coding gain G_{TC} is defined in [41] using the following formulations:

$$G_{TC} = \frac{\frac{1}{N}\sum_{i=1}^{N}\sigma_{y_{ii}}^2}{\left(\prod_{i=1}^{N}\sigma_{y_{ii}}^2\right)^{1/N}} \tag{53}$$

with,

$$\sigma_{y_{ii}}^2 = TCT^T,$$

where T means the transform matrix and C is the covariance matrix of input signal. Fig.12 shows the transform coding gain for three transform kernels at different parameter values of ρ, from which we can see that the transform core using

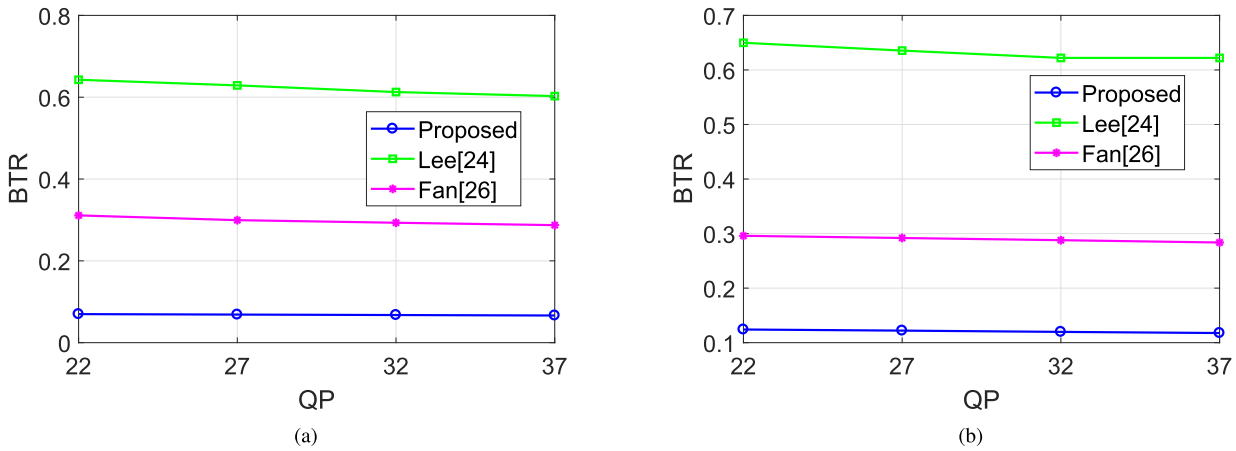


Fig. 14. The *BTR* under different QPs for *BasketballDrive* and *BQSquare*. (a) *BasketballDrive RA*. (b) *BQSquare LD*.

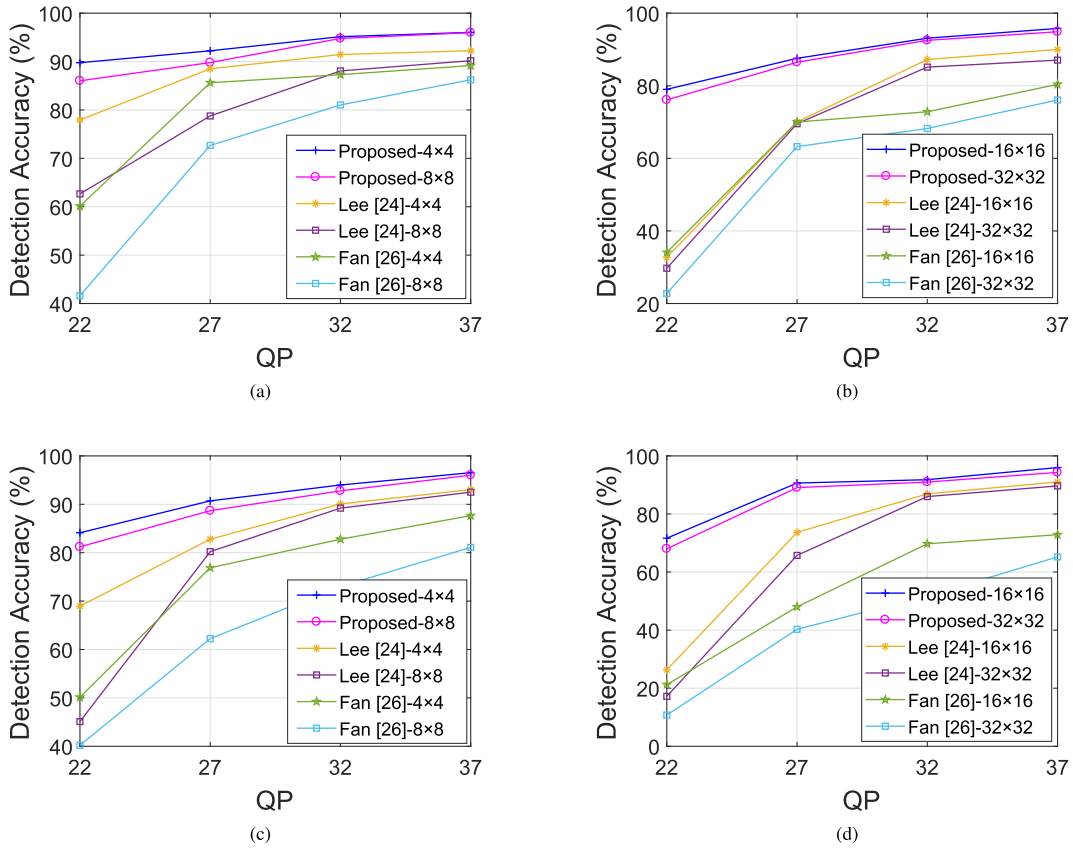


Fig. 15. The Detection Accuracy curves of proposed, Lee *et al.*'s and Fan *et al.*'s method for each TU size under RA configuration. (a) 4×4 and 8×8 (*Traffic*). (b) 16×16 and 32×32 (*Traffic*). (c) 4×4 and 8×8 (*BasketballPass*). (d) 16×16 and 32×32 (*BasketballPass*).

sparse matrix times Hadamard matrix can obtain a similar performance as DCT when ρ is great than 0.6.

B. RD Performance and Time Saving Evaluation

Table VII shows the experimental results of the proposed scheme, where TS_q denotes the time saving of transform and quantization including the corresponding inverse transform and de-quantization, T_{Anchor_q} and $T_{\mathrm{Pr}oposed\_q}$ represent time saving of anchor and proposed algorithm, respectively. Similarly, TS is the total encoding time saving with T_{Anchor} and $T_{\mathrm{Pr}oposed}$ being the encoding time of anchor and proposed method, respectively. In particular, TS_q and TS can be calculated by:

$$TS_q = \frac{T_{Anchor\_q} - T_{\mathrm{Pr}oposed\_q}}{T_{Anchor\_q}} \times 100\%. \tag{54}$$

$$TS = \frac{T_{Anchor} - T_{\mathrm{Pr}oposed}}{T_{Anchor}} \times 100\% \tag{55}$$

From Table VII, we can see that if only G-AZB detection method is used, only 4.67% to 5.21% time saving with about

TABLE VIII
THE CODING PERFORMANCE COMPARISONS OF THE STATE-OF-THE-ART METHODS (LEE *et al.*'S [24] AND FAN *et al.*'S [26])

Sequence		RA				LD			
		Lee *et al.*'s		Fan *et al.*'s		Lee *et al.*'s		Fan *et al.*'s	
		BDBR(%)	TSq(%)	BDBR(%)	TSq(%)	BDBR(%)	TSq(%)	BDBR(%)	TSq(%)
A class	Traffic	0.01	36.01	0.57	8.72	-0.21	30.21	-0.30	6.21
	PeopleOnStreet	0.04	44.56	0.02	33.06	0.65	32.16	0.46	30.69
B class	Kimono	0.46	32.23	0.70	-14.24	0.19	42.30	0.31	-5.78
	ParkScene	0.12	10.13	0.14	-5.71	0.57	55.68	0.25	46.88
	Cactus	0.24	35.76	0.72	33.24	0.06	34.16	0.15	40.27
	BasketballDrive	0.77	23.93	0.34	14.29	0.47	46.92	0.54	42.69
	BQTerrace	0.36	25.76	0.69	15.76	0.25	42.73	0.07	45.74
C class	BasketballDrill	0.13	28.91	0.31	20.70	0.15	39.72	0.24	36.20
	BQMall	0.58	30.88	1.08	34.36	0.35	26.64	0.21	32.08
	PartyScene	0.79	32.24	0.18	12.96	0.45	46.57	0.17	3.67
	RaceHorses	0.14	34.11	0.12	17.36	0.21	26.48	0.18	12.93
D class	BasketballPass	0.21	37.42	1.19	30.69	0.21	21.09	0.04	2.44
	BQSquare	0.97	63.28	0.10	38.41	0.75	19.31	0.31	6.98
	BlowingBubbles	0.75	51.22	0.63	34.00	0.31	36.76	0.28	32.77
	RaceHorses	-0.15	41.26	-0.08	0.76	0.15	39.11	0.21	7.13
E class	FourPeople	0.05	23.79	0.23	30.21	0.10	46.09	0.20	30.09
	Johnny	0.04	40.24	0.13	49.26	0.19	26.81	0.18	44.30
	KristenAndSara	-0.08	32.48	0.18	35.29	0.13	26.74	0.26	27.79
F class	BasketballDrillText	0.76	30.29	0.27	7.60	0.59	35.79	0.53	13.38
	SlideEditing	0.04	33.78	0.05	40.29	0.72	25.01	0.74	44.74
	SlideShow	2.75	30.97	-0.30	17.66	1.24	35.94	0.64	20.59
Avg.		**0.43**	**34.25**	**0.35**	**21.65**	**0.36**	**35.06**	**0.27**	**24.85**

0.03%∼0.05% BD-rate degradation is achieved, which is very limited to improve the coding efficiency, since the number of G-AZB accounts for a very small proportion, especially for larger TU size, i.e., 32×32, as shown in Fig. 2 (a)-(d). In addition, the proposed G-AZB detection method aims to find out all AZBs without RD consideration. If the G-AZB detection method fails to classify current block as AZB, this block will be checked by the following P-AZB detection method as well. Therefore, the P-AZB detection method is able to save up to 24.16% transform and quantization time with less than 0.06% BD-BR increase. The overall time saving TS are 4.01% and 5.18% for RA and LD configuration, respectively.

To further validate our scheme, we illustrate the performances of the state-of-the-art AZB detection methods proposed by Lee *et al.* [24] and Fan *et al.* [26], and the experimental results are shown in Table VIII. We can see that Lee *et al.*'s method achieves 34.25% run-time saving while leading to 0.43% RD performance degradation. For Fan *et al.*'s method, it obtains 21.65% - 24.85% time saving on average with up to 0.35% BDBR loss. In Fig.13, the RD curves of *BasketballDrive* and *BQMall* have been illustrated for the proposed, Lee *et al.*'s and Fan *et al.*'s methods. From curves in Fig.13, we can see that the proposed method has achieved a comparative RD performance with anchor (HM16.9), while the results of Lee *et al.*'s and Fan *et al.*'s method are worse than the anchor. Furthermore, to evaluate the trade-off between RD performance and time saving, we introduce a metric *BTR* (BDBR and Time saving Ratio) in Eq.(56) to verify the overall performance of AZB detection methods.

$$BTR = \frac{BDBR}{1 + TS_q} \times 100\%. \tag{56}$$

The constant 1 is utilized to make $1+TS_q$ positive. Obviously, the smaller the *BTR* is, the better the trade-off is. In Fig.14, our method shows the smallest *BTR* under each QP compared with Lee *et al.*'s and Fan *et al.*'s methods for the sequences *BasketballDrive* and *BQSquare*. Note that the BDBR is the result of four QPs (22, 27, 32 and 37), while TS_q in Fig.14 is for each QP. In addition, more time saving is achieved in the low bit-rate scenario compared with that of high bit-rate case for all three methods since more all zero blocks and non-zero blocks with only a few coefficients are generated when the QP is larger.

C. Detection Accuracy Evaluation

In this section, we verify the performance in terms of the detection rate and detection accuracy. The detection rates comparisons in terms of FNR and FPR are illustrated in Table IX, in which the average values of all sequences under QP values 22, 27, 32, and 37 for each class and each TU type. From the comparison results, we can see that the detection accuracy of the proposed method is up to 95.5% on average when detecting AZBs for 16×16 and 32×32 TUs. For the non-AZBs, only 5.8% to 6.5% non-AZBs have been misclassified by our method into AZBs. By contrast, the detection error for non-AZBs in [24] is only 5% on average, but that of AZBs reaches up to 34.1%. Moreover, the detection errors both for AZBs and non-AZBs in [26] are around 15%. To compare the detection accuracy in detail, the overall detection accuracies (DA) of different TU sizes for different QPs under RA configuration are illustrated in Fig.15. We can see that the overall detection accuracy is higher as QP becomes larger, because larger QP results in detecting less AZBs. Our method achieves the stable and higher overall detection accuracy compared

TABLE IX

THE DETECTION ACCURACY COMPARISONS WITH THE STATE-OF-THE-ART METHODS (LEE *et al.*'S [24] AND FAN *et al.*'S [26])

Sequence	TU Type	Lee *et al.*'s				Fan *et al.*'s				Proposed			
		RA		LD		RA		LD		RA		LD	
		FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR	FNR	FPR
A class	4×4	0.146	0.107	0.116	0.108	0.064	0.075	0.033	0.092	0.011	0.061	0.109	0.071
	8×8	0.093	0.008	0.327	0.027	0.059	0.145	0.130	0.085	0.045	0.043	0.077	0.124
	16×16	0.170	0.062	0.241	0.043	0.321	0.336	0.277	0.217	0.066	0.031	0.070	0.061
	32×32	0.409	0.052	0.583	0.003	0.293	0.278	0.252	0.326	0.010	0.040	0.041	0.050
B class	4×4	0.210	0.074	0.286	0.046	0.062	0.098	0.080	0.061	0.094	0.105	0.100	0.022
	8×8	0.228	0.057	0.503	0.095	0.085	0.060	0.150	0.114	0.094	0.062	0.066	0.096
	16×16	0.560	0.076	0.277	0.067	0.204	0.268	0.276	0.255	0.041	0.081	0.067	0.054
	32×32	0.379	0.043	0.309	0.004	0.317	0.292	0.366	0.430	0.049	0.038	0.048	0.070
C class	4×4	0.196	0.022	0.178	0.082	0.015	0.116	0.128	0.016	0.106	0.095	0.113	0.091
	8×8	0.208	0.020	0.262	0.023	0.035	0.043	0.267	0.010	0.030	0.063	0.031	0.018
	16×16	0.221	0.097	0.434	0.103	0.264	0.005	0.084	0.143	0.033	0.049	0.037	0.045
	32×32	0.480	0.031	0.791	0.011	0.455	0.312	0.130	0.232	0.052	0.077	0.011	0.039
D class	4×4	0.157	0.056	0.242	0.118	0.022	0.022	0.029	0.143	0.173	0.134	0.132	0.046
	8×8	0.152	0.104	0.138	0.017	0.055	0.028	0.050	0.078	0.056	0.112	0.099	0.123
	16×16	0.409	0.038	0.382	0.101	0.079	0.103	0.387	0.272	0.036	0.038	0.014	0.065
	32×32	0.426	0.032	0.427	0.040	0.196	0.115	0.141	0.231	0.044	0.074	0.022	0.088
E class	4×4	0.128	0.046	0.100	0.063	0.116	0.108	0.072	0.039	0.096	0.096	0.096	0.020
	8×8	0.167	0.070	0.125	0.006	0.127	0.097	0.139	0.011	0.066	0.050	0.203	0.020
	16×16	0.425	0.004	0.331	0.015	0.255	0.104	0.153	0.081	0.083	0.077	0.021	0.097
	32×32	0.519	0.006	0.360	0.038	0.283	0.172	0.271	0.142	0.002	0.041	0.080	0.030
F class	4×4	0.145	0.099	0.216	0.028	0.008	0.250	0.070	0.096	0.109	0.180	0.080	0.204
	8×8	0.071	0.058	0.327	0.007	0.030	0.027	0.020	0.033	0.017	0.045	0.032	0.027
	16×16	0.524	0.061	0.755	0.024	0.111	0.175	0.162	0.104	0.038	0.067	0.099	0.103
	32×32	0.850	0.082	0.483	0.032	0.156	0.176	0.273	0.227	0.091	0.087	0.123	0.077
Avg.	4×4 8×8	**0.158**	**0.060**	**0.235**	**0.052**	**0.056**	**0.089**	**0.097**	**0.065**	**0.075**	**0.087**	**0.095**	**0.072**
	16×16 32×32	**0.448**	**0.049**	**0.448**	**0.040**	**0.245**	**0.195**	**0.231**	**0.222**	**0.045**	**0.058**	**0.053**	**0.065**
	All	**0.303**	**0.054**	**0.341**	**0.046**	**0.150**	**0.142**	**0.164**	**0.143**	**0.060**	**0.073**	**0.074**	**0.068**

with Lee *et al.*'s and Fan *et al.*'s methods for different TU sizes. In Fig.15(a) and (c), Lee *et al.*'s method for 4×4 is quite comparable with our method, but for the larger TU sizes, e.g., 16×16 and 32×32, our detection algorithm is very superior to Lee *et al.*'s and Fan *et al.*'s methods as shown in Fig.15(b) and (d).

V. CONCLUSION

In this paper, an efficient AZB detection method with the consideration of SDQ is proposed, and the scenarios that may lead to zero coefficient blocks are adequately covered by the G-AZB followed by P-AZB detection. In particular, the P-AZB detection follows the design philosophy of RDOQ process, where maximum transform coefficient amplitude and RDO with adaptive rate-distortion estimations are employed. The experimental results show that the proposed detection method is able to efficiently reduce the transform and quantization time in the RDO procedure with ignorable R-D performance degradation. Moreover, the detection accuracy of proposed method is quite competitive compared with the state-of-the-art methods, especially for larger TU sizes, i.e., 16×16 and 32×32.

REFERENCES

[1] 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.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[3] 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.

[4] Z. Xuan, Y. Zhenghua, and Y. Songyu, "Method for detecting all-zero DCT coefficients ahead of discrete cosine transformation and quantisation," *Electron. Lett.*, vol. 34, no. 19, pp. 1839–1840, Sep. 1998.

[5] G. Song and X. Jiang, "All-zero block detection algorithm based on quantitative parameters," in *Proc. 3rd Int. Conf. Syst. Sci., Eng. Design Manuf. Informatization*, vol. 2, Oct. 2012, pp. 296–299.

[6] J.-F. Yang, S.-C. Chang, and C.-Y. Chen, "Computation reduction for motion search in low rate video coders," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 948–951, Oct. 2002.

[7] D. Wu, K. P. Lim, T. K. Chiew, J. Y. Tham, and K. H. Goh, "An adaptive thresholding technique for the detection of all-zeros blocks in H. 264," in *Proc. IEEE Int. Conf. Image Process.*, vol. 5, Sep./Oct. 2007, pp. V-329–V-332.

[8] G. Zhong, L. Lu, and N. Jiang, "Fast mode decision based on all-zero block in H.264/AVC," in *Proc. 2nd Int. Conf. Digit. Manuf. Automat.*, Aug. 2011, pp. 535–538.

[9] Z. Xin and Z. Wei, "Fast inter prediction block mode decision approach for H.264/AVC based on all-zero blocks detection," in *Proc. 8th Conf. Ind. Electron. Appl. (ICIEA)*, Jun. 2013, pp. 896–899.

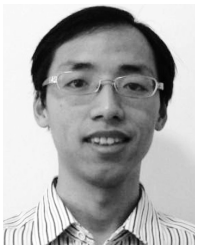
[10] H. Tang and H. Shi, "Fast mode decision algorithm for H.264/AVC based on all-zero blocks predetermination," in *Proc. Int. Conf. Measuring Technol. Mechatronics Autom.*, vol. 2, Apr. 2009, pp. 780–783.

[11] C. Zhu *et al.*, "Multi-level low-complexity coefficient discarding scheme for video encoder," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2014, pp. 5–8.

[12] 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.

[13] H. Wang, H. Du, and J. Wu, "Predicting zero coefficients for high efficiency video coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2014, pp. 1–6.
[14] X. Ji, S. Kwong, D. Zhao, H. Wang, C.-C. J. Kuo, and Q. Dai, "Early determination of zero-quantized 8×8 DCT Coefficients," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 12, pp. 1755–1765, Dec. 2009.
[15] H. Wang, S. Kwong, and C. W. Kok, "Analytical model of zero quantized DCT coefficients for video encoder optimization," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 801–804.
[16] H. Wang, S. Kwong, and C.-W. Kok, "Efficient predictive model of zero quantized DCT coefficients for fast video encoding," *Image Vis. Comput.*, vol. 25, no. 6, pp. 922–933, 2007, doi: 10.1016/j.imavis.2006.07.007.
[17] Y. H. Moon, G. Y. Kim, and J. H. Kim, "An improved early detection algorithm for all-zero blocks in H.264 video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 1053–1057, Aug. 2005.
[18] H. Wang, S. Kwong, and C.-W. Kok, "Efficient prediction algorithm of integer DCT coefficients for H.264/AVC optimization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 4, pp. 547–552, Apr. 2006.
[19] Z. Xie, Y. Liu, J. Liu, and T. Yang, "A general method for detecting all-zero blocks prior to DCT and quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 2, pp. 237–241, Feb. 2007.
[20] H. Wang and S. Kwong, "Hybrid model to detect zero quantized DCT coefficients in H.264," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 728–735, Jun. 2007.
[21] H. Wang, S. Kwong, and C.-W. Kok, "Effectively detecting all-zero DCT blocks for H.264 optimization," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 1329–1332.
[22] H. Wang and S. Kwong, "Prediction of zero quantized DCT coefficients in H.264/AVC using Hadamard transformed information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 4, pp. 510–515, Apr. 2008.
[23] Z. Liu, L. Li, Y. Song, S. Li, S. Goto, and T. Ikenaga, "Motion feature and Hadamard coefficient-based fast multiple reference frame motion estimation for H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 5, pp. 620–632, May 2008.
[24] K. Lee, H.-J. Lee, J. Kim, and Y. Choi, "A novel algorithm for zero block detection in high efficiency video coding," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 6, pp. 1124–1134, Dec. 2013.
[25] H. Wang *et al.*, "Early detection of all-zero 4×4 blocks in high efficiency video coding," *J. Vis. Commun. Image Represent.*, vol. 25, no. 7, pp. 1784–1790, 2014.
[26] H. Fan, R. Wang, L. Ding, X. Xie, H. Jia, and W. Gao, "Hybrid zero block detection for high efficiency video coding," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 537–543, Mar. 2016.
[27] B. Lee, J. Jung, and M. Kim, "An all-zero block detection scheme for low-complexity HEVC encoders," *IEEE Trans. Multimedia*, vol. 18, no. 7, pp. 1257–1268, Jul. 2016.
[28] Q. Yu, X. Zhang, S. Wang, and S. Ma, "Early termination of coding unit splitting for HEVC," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2012, pp. 1–4.
[29] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan. (2017). "Reducing complexity of HEVC: A deep learning approach." [Online]. Available: https://arxiv.org/abs/1710.01218
[30] H. Yin, H. Cai, and H. Lu, "A new all-zero block detection algorithm for high efficiency video coding," in *Proc. Data Compress. Conf. (DCC)*, Apr. 2017, p. 470.
[31] J. Cui, R. Xiong, F. Luo, S. Wang, and S. Ma, "An adaptive and low-complexity all-zero block detection for HEVC encoder," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 2, May 2017, pp. 1–4.
[32] H. Kibeya, F. Belghith, M. A. B. Ayed, and N. Masmoudi, "Fast coding unit selection and motion estimation algorithm based on early detection of zero block quantified transform coefficients for high-efficiency video coding standard," *IET Image Process.*, vol. 10, no. 5, pp. 371–380, May 2016.
[33] S. Annadurai, *Fundamentals of Digital Image Processing*. New York, NY, USA: Pearson, 2007.
[34] X. Zhang *et al.*, "Low-rank-based nonlocal adaptive loop filter for high-efficiency video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2177–2188, Oct. 2016.
[35] H. Wang, S. Kwong, and C.-W. Kok, "Fast video coding based on Gaussian model of DCT coefficients," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, p. 4.
[36] I.-M. Pao and M.-T. Sun, "Modeling DCT coefficients for fast video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 608–616, Jun. 1999.
[37] X. Zhang, R. Xiong, W. Lin, S. Ma, J. Liu, and W. Gao, "Video compression artifact reduction via spatio-temporal multi-hypothesis prediction," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6048–6061, Dec. 2015.
[38] J. Cui, S. Wang, S. Wang, X. Zhang, S. Ma, and W. Gao, "Hybrid Laplace distribution-based low complexity rate-distortion optimized quantization," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3802–3816, Aug. 2017.
[39] T.-Y. Huang and H. H. Chen, "Efficient quantization based on rate–distortion optimization for video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 6, pp. 1099–1106, Jun. 2016.
[40] F. Bossen, *Common HM Test Conditions and Software Reference Configurations*, document JCTVC-F900, Joint Collaborative Team Video Coding, 2011.
[41] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

Jing Cui received the B.S. degree from the Department of Mechanical and Electronic Engineering, Beijing Institute of Technology, in 2012, and the M.S. degree in electrical computer engineering from Seoul National University in 2016. She is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Peking University. Her research interests include data compression and image/video coding.



Ruiqin Xiong (M'08–SM'17) received the B.S. degree in computer science from the University of Science and Technology of China in 2001 and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, in 2007.

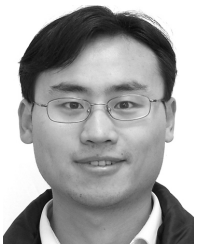
He was a Research Intern with Microsoft Research Asia from 2002 to 2007 and a Senior Research Associate with the University of New South Wales, Australia, from 2007 to 2009. He joined the School of Electronic Engineering and Computer Science, Institute of Digital Media, Peking University, in 2010, where he is currently a Professor.

He has published over 110 technical papers in referred international journals and conferences. His research interests include statistical image modeling, deep learning, image and video processing, compression, and communications. He received the Best Student Paper Award at SPIE Conference on Visual Communications and Image Processing 2005 and the Best Paper Award at the IEEE Visual Communications and Image Processing 2011. He was also a co-recipient of the Best Student Paper Award at the IEEE Visual Communications and Image Processing 2017.

Xinfeng Zhang received the B.S. degree in computer science from the Hebei University of Technology, Tianjin, China, in 2007, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2014. He is currently a Research Fellow with the University of Southern California, Los Angeles. His research interests include image and video processing and image and video compression.

Shiqi Wang 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 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 with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore, as a Research Fellow. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He has proposed over 30 technical proposals to ISO/MPEG, ITU-T, and AVS standards. His research interests include video compression, image/video quality assessment, and image/video search and analysis.

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-doctoral position with Peking University. His current research interests include video compression and image and video quality assessment.



Siwei Ma (M'03–SM'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. He held a post-doctoral position with the University of Southern California, Los Angeles, CA, USA, from 2005 to 2007. He joined the School of Electronics Engineering and Computer Science, Institute of Digital Media, Peking University, Beijing, where he is currently a Professor. He has authored over 200 technical articles in refereed journals and proceedings in image and video coding, video processing, video streaming, and transmission. He is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY and the *Journal of Visual Communication and Image Representation*.

Wen Gao (M'92–SM'05–F'09) received the Ph.D. degree in electronics engineering from The University of Tokyo, Japan, in 1991. He was a Professor of computer science with the Harbin Institute of Technology from 1991 to 1995, and a Professor with the Institute of Computing Technology, Chinese Academy of Sciences, from 1996 to 2006. He is currently a Professor of computer science with Peking University, Beijing, China. He has authored extensively, including five books and over 600 technical articles in refereed journals and conference proceedings in the areas of image processing, video coding and communication, pattern recognition, multimedia information retrieval, multimodal interface, and bioinformatics. He chaired a number of prestigious international conferences on multimedia and video signal processing, such as the IEEE ISCAS, ICME, and the ACM Multimedia, and also served on the advisory and technical committees of numerous professional organizations. He served or serves on the editorial board for several journals, such as the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT, the *Journal of Image Communications* (EURASIP), and the *Journal of Visual Communication and Image Representation*.