A Probability-Based Zero-Block Early Termination Algorithm for QSHVC

Dayong Wang, Xin Lu, Yu Sun, Qianmin Wang, Weisheng Li, *Member, IEEE,*
Frederic Dufaux, *Fellow, IEEE,*Ce Zhu, *Fellow, IEEE*

***Abstract*—To seamlessly adapt to time-varying network bandwidths, the Quality Scalable High-Efficiency Video Coding (QSHVC) is developed. However, its coding process is overly complex, and this seriously limits its wide applications in real-time environments. Therefore, it is of great significance to study fast coding algorithms for QSHVC. In this paper, we propose a novel probability-based zero-block early termination algorithm for QSHVC. First, we observed that the generated residual coefficients follow the Laplace distribution if a CU is accurately predicted. According to this observation, we derive the sum of squared differences based All-Zero Block (AZB) decision condition. Second, we develop the Hadamard Transform (HT)-based zero-valued quantized coefficient decision condition to obtain zero-valued quantized coefficients and the corresponding Partial-Zero Block (PZB). Third, the probability of each coding mode and coding depth being chosen as the best ones are combined with both AZBs and PZBs to derive the probability-based early termination condition. The experimental results show that the proposed algorithm can improve the average coding speed by 80.6% with a 0.31% decrease in BDBR.**

***Index Terms*—SHVC, probability, AII-Zero Block, Partial-Zero Block.**

I. Introduction

THE popularization of smartphones underpins the increasing demand for video applications. The use of video conferencing has seen a significant increase during the Covid-19 pandemic crisis. Heterogeneous networks with time-varying channel bandwidths are used in visual communication, and bandwidth fluctuates even working under the same network environment. This requires video streams to be scalable to a variety of network bandwidths. Quality Scalable High-Efficiency Video Coding (QSHVC) is developed to effectively address this need [1].

D. Wang is with Chongqing Key Laboratory on Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China(e-mail: wangdayong@cqupt.edu.cn).

X. Lu is with the School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, United Kingdom (e-mail: xin.lu@dmu.ac.uk).

Y. Sun is with the Department of Computer Science, University of Central Arkansas, Conway, AR 72035 USA (e-mail: yusun@uca.edu).

Q. Wang is with School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China(e-mail: 851979847@qq.com).

W. Li is with Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: liws@cqupt.edu.cn).

F. Dufaux is with the Laboratoire des Signaux et Systèmes,Université Paris-Saclay, CNRS, CentraleSupélec, 91192 Gif-sur-Yvette, France (e-mail: frederic.dufaux@l2s.centralesupelec.fr).

C. Zhu is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: eczhu@uestc.edu.cn).

QSHVC is one of the scalable versions of the High Efficiency Video Coding (HEVC) standard. To achieve scalable quality, SHVC uses a Base Layer (BL) and one or more Enhancement Layers (ELs) to encode the same video sequence. Each layer is encoded with a different Quantization Parameter (QP) to achieve several video representations with different picture quality [2]. The identical encoding process as in HEVC is applied on the BL, and the additional Inter-layer prediction is employed to encode the ELs, where Inter-Layer Reference (ILR) modes are introduced. Given that HEVC is quite complex, the computational complexity of QSHVC is even much greater than that of HEVC. This seriously prevents the extensive deployment of QSHVC in real-time scenarios. It is therefore of great significance to develop fast coding algorithms for QSHVC [3]–[5].

This paper proposes a novel probability-based zero-block early termination algorithm to alleviate the computational complexity of the QSHVC encoder. We first obtain the Sum of Squared Differences (SSD)-based All-Zero Block (AZB) decision condition, which is combined with the probabilities of various modes and depths being chosen as the best one, to derive the corresponding early termination conditions for individual modes and depths, thus terminating the evaluations early. For the remaining coding modes and depths, we then propose a Partial-Zero Block (PZB) detection method to obtain the proportion of zero-valued coefficients. The obtained proportion is employed on top of the possibility of each mode or depth being the best one to further refine the early termination conditions for the subsequent coding process. **To the best of our knowlege, this is the first paper use both AZB and PZB and their probabilities to early terminate coding mode and depth selections**.

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 presents an overview of the proposed algorithm. Section 4 proposes the probability-based AZB early termination. Section 5 introduces the probability-based PZB early termination. Section 6 discusses and analyses the experimental results. Finally, Section 7 concludes this research and indicates the plan for future work.

II. Related work

In SHVC, each Coding Tree Unit (CTU) can be divided using a quad-tree structure at four coding depths (0, 1, 2, 3) corresponding to the CU sizes of 64×64, 32×32, 16×16 and 8×8, respectively. Both ILR and Intra modes for all coding depths are evaluated to determine the best coding

depth and mode. This encoding process ensures the best coding efficiency at the expense of significantly increased computational complexity. To speed up the encoding process, it is straightforward to early terminate the evaluations of coding modes and depths. The Rate-Distortion (RD) cost and residual coefficients are often considered to determine whether coding modes and depths can be early terminated.

(1) RD costs-based early termination

Generally, a CU that generates a large RD cost is likely to be subdivided to a next coding depth, and vice versa. In [6], if the aggregation of the estimated RD cost of the sub-CUs in the current CU is larger than the RD cost of the current CU, the subdivision for the current CU is terminated. Cho et al. [7] used a Bayes decision rule to evaluate the probability of early termination for the current coding depth according to the RD cost, thus achieving the fast CU splitting and pruning. Zhang et al. [8] utilized the conditional random field to examine whether it is appropriate to apply early CU decisions. In [9], Zhang et al. employed the ratio of the RD cost between the current CU and its neighboring CUs as the feature, and the weighted Support Vector Machine (SVM) is used to determine whether to proceed with next coding depths. To achieve an early termination for coding mode evaluation, Wang et al. [10] proposed a few conditions based on the RD cost differences among square coding modes. Wang et al. [11]–[13] first investigated RD cost distributions of ILR mode and Intra mode, and then used the distributions to determine whether ILR mode is the best mode thus skipping Intra mode.

(2) Residual Coefficients-based Early Termination

To effectively exploit residual coefficients in early termination algorithms, research tends to focus on the expectation, variance, and distribution of residual coefficient, as well as AZB.

Expectation and variance of residual coefficients: In [14], a given CU comprising residual coefficients is divided into two parts in two structures: the top half and the bottom half, and the left side and the right side. The differences in variances between the two parts in both structures are calculated, respectively. If either of the differences is significant, the CU is divided further, and vice versa. Wang et al. [15] split a CU in the same way as in [14], however, hypothesis testing is introduced to examine the significance of the difference in variances.

Distribution of residual coefficients: Wang et al. [10] [15] used the "kurtosis skewness test" to observe whether residual coefficients obey the Gaussian distribution. In the affirmative, the current mode is considered the best one and the evaluation of the remaining modes is therefore skipped. Otherwise, the remaining modes are required to undertake further evaluations. In [16], the Chi-square distribution fitting test method was employed to check the distribution of residual coefficients, so as to decide whether the early termination is to be applied in the coding depth decision or not.

Usage of AZBs: When an AZB is presented, it generally indicates that a CU can be accurately predicted using the current coding mode and depth. The remaining selection process for coding mode and depth can therefore be skipped. In [17], Jung et al. utilized the inter-layer and spatial correlations to detect AZBs, thus terminating the mode decision process. Pan et al. [18] [19] proposed that the coding mode is considered the best one when the quantized residual coefficients generated by the corresponding merge mode are all zero, then the mode decision process can be early terminated. Chiang et al. [20] determined whether the inter-coding is an AZB based on the Sum-of-Absolute-Difference (SAD) value to terminate the depth selection process early. Jiang et al. [21] derived the decision condition for the AZB detection to predict if the current mode is the best one. For three types of transform and quantization methods, Wang et al. [22] theoretically derive the decision condition for DCT coefficients to be quantized to zeros. For different types of DCT and quantization implementations, Wang et al. [23] first used the Gaussian distribution to propose multiple thresholds, and then further optimize the thresholds to sufficiently detect AZB. Wang et al. [24] proposed to use the Sum of Absolute Transformed Difference (SATD) to detect AZB. Lee et al. [25] used the relationship between Walsh Hadamard and DCT transform kernels to derive AZBs. Methods in [26]–[28] divided CUs into Genuine AZBs (G-AZB) and Pseudo AZBs (P-AZBs). For G-AZB, the AZB decision condition is directly derived from the residual distribution. For P-AZBs, the AZB decision condition is proposed according to the RD cost.

Despite the aforementioned research efforts, the following issues still need to be investigated.

(1) Decision conditions for AZB are usually developed based on SAD or SATD. HEVC, SHVC and VVC quantize the sum of actual residual coefficients, while SAD-based decision conditions quantize the sum of absolute residual coefficients to predict AZBs. Obviously, the sum of absolute residual coefficients is greater than or equal to the sum of actual residual coefficients. Therefore, SAD-based decision conditions are too strict. SATD needs to further use additional Hadamard Transform (HT), which will increase the coding complexity.

(2) If an AZB is generated, this indicates a CU is predicted accurately. However, there is usually a limited number of AZBs only in a given CU. The coding mode or depth is also possibly optimal if the blocks of residual coefficients generated in the quantization operation contain largely zero-valued coefficients. These blocks are denoted as PZBs. The more zero-valued quantized coefficients, the greater the probability of the current mode or depth being chosen as the best one. Although we can directly obtain zero-valued quantized coefficients by quantizating DCT coefficients, the DCT process is very complex. Therefore, the improvement in encoding time reduction was limited as well.

(3) The coding mode or depth selection is closely related to the probabilities of each mode or depth being selected as the best one. The above algorithms do not consider this feature, which might limit the improvement of the coding speed.

The above existing issues motivate us to propose a new zero-block early termination algorithm based on the probability analysis of coding modes and depths. The major novelty and contributions of this paper are summarized below:

(1) We derive a SSD-based AZB decision condition. Generally, if a CU is predicted with great accuracy, the generated residual coefficients follow the Laplace distribution. According to this distribution, we derive a SSD-based AZB decision condition. Given a set of data, the square of the sum is generally greater than the sum of squares. The square root of SSD is smaller than SAD, thus SSD-based AZB decision is looser than SAD-based AZB decision. Therefore, more AZBs can be obtained. Compared to SATD-based AZB decision, SSD-based AZB decision does not need HT cost, and naturally decreases the coding complexity.
(2) We derive a HT based zero-valued quantized coefficient decision condition. Since HT is similar to but much simpler than the DCT transform, for both the HT coefficient and the DCT coefficient at the same position, if they are sufficiently close, we can use HT rather than DCT to obtain the zero-valued coefficient. Therefore, we derive the HT based zero-valued quantized coefficient decision condition. If the decision condition is satisfied, HT can be used instead of DCT, and the corresponding transform can be significantly simplified. In turn, the coding time is further reduced.
(3) The probability of each coding mode and coding depth being chosen as the best one is combined with both AZBs and PZBs to derive a probability-based AZB and PZB early termination conditions. For a mode or depth with a higher probability, looser early termination conditions can be used to improve coding speed; while for a mode or depth with a smaller probability, stricter early termination conditions are used to maintain coding speed.

III. Overview of the Proposed Probability-Based Zero-Block Early Termination

To early terminate the mode and depth selection processes, we propose a probability-based zero-block early termination algorithm, which includes a probability-based AZB early termination (PBAET) strategy and a probability-based PZB early termination (PBPET) strategy, respectively. The overview of the algorithm is illustrated in Fig. 1. The PBAET strategy and PBPET strategy as well as its corresponding sub-strategies are shown in the left part of Fig. 1.

We first use PBAET to determine whether the current mode or depth is the best one. In the affirmative, we further use PBPET to early terminate mode or depth selection. In PBAET, we first derive the SSD-based AZB determining condition based on residual coefficients following Laplace distribution. It is subsequently combined with the probabilities of modes and depths to study the corresponding early termination decision conditions. If this decision condition is satisfied, the coding process can be early terminated. Otherwise, PBPET is subsequently used to early terminate the mode or depth selection process.

In PBPET, we first predict zero-valued quantized coefficients based on SSD. As discussed in section 5, this prediction cannot obtain all zero-valued quantized coefficients. Therefore, we further derive the PZB prediction condition based on HT and predict zero-valued quantized coefficients by HT. In this process, if the subsequent HT coefficients are to be quantized to zero, we can early terminate the calculation for these HT coefficients. Therefore, we propose early termination of PZB based on energy conservation. Through the above process, the percentage of zero-valued coefficients in a CU is obtained. It is subsequently combined with the probabilities of modes and depths to obtain the corresponding early termination decision conditions to terminate the mode or depth selection process.

Extensive experiments are undertaken to justify the proposed algorithm. To cover various spatial resolutions, two video sequences were selected from each of video classes: B, C, D, and E. In particular, we used the following video sequences for validation: Sunflower and Tractor in class B; Park and Town in class C; Flowervase and PartyScene in class D; RaceHorses and BlowingBubbles in class E. These video sequences are representative, as they contain different levels of motion and textual complexity. Experiment parameters are set according to the Common Test Conditions (CTC) [29]. Specifically, the QP values of $(26, 30, 34, 38)$ are used for the BL while the QP values of $(20, 24, 28, 32)$ and $(22, 26, 30, 34)$ are adopted for the EL. As the divergence between $(20, 24, 28, 32)$ in EL and $(26, 30, 34, 38)$ in BL is greater than that between $(22, 26, 30, 34)$ in EL and $(26, 30, 34, 38)$ in BL, if the algorithm performs well when QP values of $(20, 24, 28, 32)$ are employed in EL, there are grounds to believe that it performs better when QP values of $(22, 26, 30, 34)$ are used in EL. Therefore, only QP values of $(20, 24, 28, 32)$ are used for EL in the training phase, while both QP values of $(22, 26, 30, 34)$ and $(20, 24, 28, 32)$ are used for EL in the validation phase.

IV. Probability-Based AZB Early Termination

If residual coefficients presented in a CU are all zero after quantization, it indicates an accurate prediction is achieved. The corresponding mode or depth is very likely to be the best, and the mode or the depth selection process can therefore be early terminated. To achieve this, the condition to decide if a CU is an AZB should first be established. The AZB decision condition is so strict that the improvement in encoding time reduction is limited if the condition is applied directly for early termination. The decision condition of early termination is closely related to the probability of each mode or depth being selected as the best one.

For the modes or depths that are more likely to be chosen, we can set a looser decision condition to improve the encoding speed. Meanwhile, a stricter condition is chosen for the modes or depths that are less likely to be selected to maintain the coding efficiency. Therefore, the probability of each mode or depth being selected as the best one should be taken into consideration when we further develop the corresponding early termination conditions. The proposed probability-based AZB early termination algorithm is described below.

A. SSD-based AZB decision condition

Generally, if a CU is predicted with great accuracy, the generated residual coefficients follow the Laplace distribution

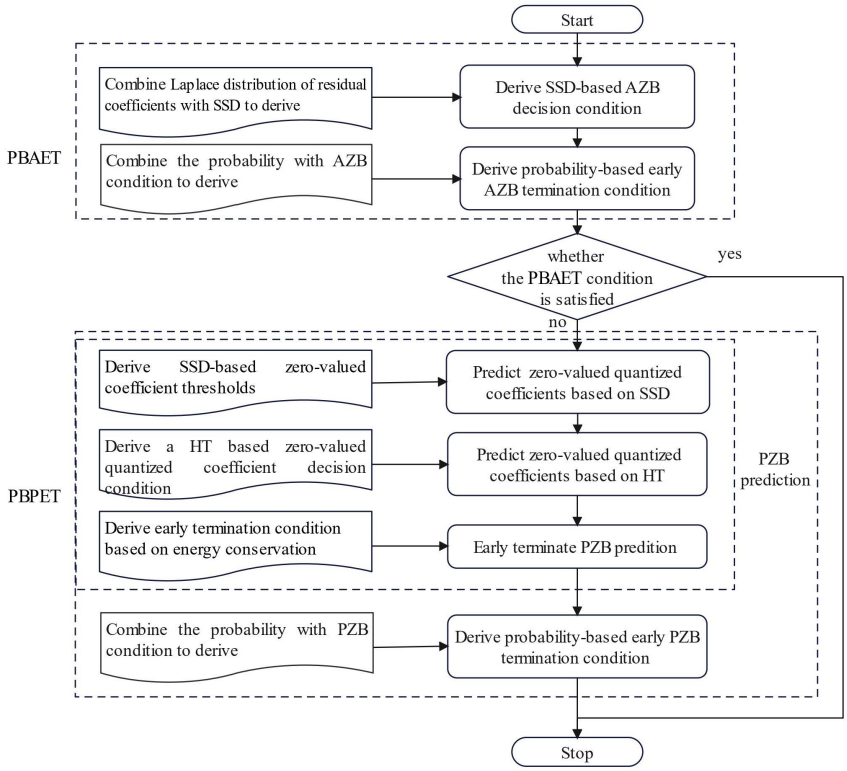

Fig. 1: Flowchart of the overall algorithm

with the expectation of 0 and the variance of σ^2 , and the corresponding probability density function is [30]

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|} \tag{1}$$

where λ represents the scale parameter of the Laplace distribution. As the expectation is 0, the corresponding variance σ^2 is:

$$D(x) = E\left(x^2\right) - (E(x))^2 = \frac{\lambda}{2}\int_{-\infty}^{\infty} x^2 e^{-\lambda|x|} dx - 0 = \frac{2}{\lambda^2} \tag{2}$$

According to Eq. (2), we have

$$\lambda = \frac{\sqrt{2}}{\sigma} \tag{3}$$

Assuming that x_{ij} is the residual value at the position of (i, j), the corresponding variance σ^2 is calculated as follows:

$$\sigma^2 = D(x_{ij}) = E(x_{ij}^2) - \left(E(x_{ij})\right)^2 \tag{4}$$

where $E(x_{ij})$ and $D(x_{ij})$ represent the expectation and the variance of x_{ij} , respectively. As the expectation of the residual coefficient is 0, namely $E(x_{ij})$ is $0.E(x_{ij}^2)$ can be calculated as follows:

$$E(x_{ij}^2) = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} x_{ij}^2}{N^2} \tag{5}$$

SSD is represented by $\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} x_{ij}^2$. Combining Eqs. (4) and (5), we have

$$\sigma^2 = \frac{SSD}{N^2} \tag{6}$$

The variance of the DCT coefficients at position (u, v) is

$$\sigma_{dct}^2(u, v) = \sigma^2[\mathrm{ARA}^T]_{uu}[\mathrm{ARA}^T]_{vv} \tag{7}$$

where$[\cdot]_{u,u}$is the value of the coefficient at position (u, u) in the matrix, A is an DCT matrix, A^T represents the transpose of A, and R is the correlation matrix. For a CU of size 8×8, the corresponding R is

$$\mathrm{R} = \begin{bmatrix} 1 & \rho & \cdots & \rho^7 \\ \rho & 1 & \cdots & \rho^6 \\ \vdots & \vdots & \vdots & \vdots \\ \rho^7 & \cdots & \rho & 1 \end{bmatrix} \tag{8}$$

where ρ is the correlation coefficient of the matrix of residual coefficients, and its value ranges from 0.4 to 0.75 [30]. The value of ρ used in [30] is 0.6. The average of σ_{dct}^2 is:

$$\sigma_{\mathrm{dct}}^2 = E[\sigma^2[\mathrm{ARA}^T]_{uu}[\mathrm{ARA}^T]_{vv}] = \sigma^2 \tag{9}$$

Combining Eqs. (3),(6) and (9), we obtain the parameters of the Laplace distribution based on the DCT transform as follows:

$$\lambda_{dct} = \frac{\sqrt{2}N}{\sqrt{SSD}} \tag{10}$$

For the residual coefficients that follow the Laplace distribution, the number of non-zero coefficients generated from quantization is [31]:

$$R(q) = N \times N \times e^{-(1-f)q\lambda \mathrm{dc}t} \tag{11}$$

where f is the rounding offset parameter in the quantization process. In Intra prediction, f generally takes the value of 1/3; while in Inter prediction, f is 1/6. q is the quantization step size. If $R(q)$ in the above equation is less than 1, it means that the coefficients in a block of size N×N are all quantized to zero. Therefore, we can use the following conditions to determine if a block is an AZB:

$$\left(N^2 e^{-(1-f)q\lambda_{\mathrm{dct}}} + 0.5\right) < 1 \tag{12}$$

where 0.5 is the rounding offset. If Eq. (12) holds, the quantized coefficients in an N×N block are entirely 0. Combining Eqs. (10) and (12), the decision condition for AZB based on SSD is deduced as:

$$\sqrt{\mathrm{SSD}} < \frac{\sqrt{2}\,(1-f)\,qN}{\ln 2 N^2} \tag{13}$$

The above AZB decision condition is deduced for 8×8 CUs. Similar decision conditions can be obtained for 16×16, 32×32 and 64×64 CUs.

As mentioned above, the coding mode or depth selections are closely related to their probabilities of being selected as the best ones. We need to combine the probability of each mode or depth being selected as the best one to obtain the corresponding decision condition. Suppose α is a parameter, which is determined according to the probability of the current CU adopting a particular mode or depth, Eq. (13) can be revised as:

$$\sqrt{\mathrm{SSD}} < \alpha \cdot \frac{\sqrt{2}\,(1-f)\,qN}{\ln 2 N^2} \tag{14}$$

It is a key problem to obtain the optimal value for α. Since there are mode selection and depth selection in SHVC, we combine with probabilities of mode and depth to obtain their corresponding optimal α in the following sub-sections, respectively.

B. Mode probability-based ILR mode early termination

In our previous research [16], we discussed how to obtain the probability of each particular mode being selected as the best one. We divided the probability range of [0%~100%] into 5 equal intervals, namely [0%~20%], [20%~40%], [40%~60%], [60%~80%], and [80%~100%]. When the range is divided into more equal intervals, the calculation process is more complex while the accuracy is only improved negligibly. However, when fewer equal intervals are divided, the accuracy is significantly degraded. Therefore, we divide the probability range of [0%~100%] into 5 equal intervals. Since probabilities in different intervals are significantly different, these five intervals are tested individually to obtain their corresponding optimal values.

As the coding process of ILR mode is relatively simple and the ILR mode is often adopted by most CUs as the best mode [15], we first evaluate the ILR mode to determine whether the early termination condition is satisfied, thus skipping the Intra prediction. To achieve the decision condition of AZB, α in Eq. (14) is set to common values, e.g. 1, 2, 3, 4, 5, 6, etc when the probability of the ILR mode being selected as the best is 40%~60%. For various α values, the corresponding coding efficiency is shown in Table I, where BDBR [32] measures the bitrate difference when identical PSNR is achieved in the EL. A positive or negative BDBR indicates either a loss or an increase in coding efficiency, respectively.

In general, the turning point is usually chosen as the optimal value of α. If α is greater than the turning point, its corresponding BDBR will change. In other words, its corresponding accuracy will also change. Therefore, this α is selected as the best one. As can be seen from Table I, there is a turning point for the "Flowervase" and "BlowingBubbles" sequences when α is greater than 4. Therefore, $\alpha = 4$ is selected. In addition, as also can be seen in Table I, in many sequences, BDBR does not vary under different alpha values and only one decimal place changes. The reason for this is that the proportion of CUs from the interval of [40%~60%] is quite small. In particular, the proportions of CUs in Sunflower, tractor, park, town, Flowervase, PartyScene, BlowingBubbles and RaceHorses are 4.3%, 3.6%, 8.1%, 5.7%, 6.7%, 5.1%, 5.6% and 4.7%, respectively. BDBR does not vary significantly under different alpha values and only one decimal place changes due to the small proportions.

The experimental results show that the average precisions are 92.46%, 91.05%, 90.45%, 90.05%, 87.91% and 87.13% when the α values are 1, 2, 3, 4, 5 and 6, respectively; the average recalls are 25.80%, 40.00%, 51.55%, 60.79%, 67.91% and 73.91% when the α values are 1, 2, 3, 4, 5 and 6, respectively. We can observe that as α increases, the precision gradually decreases and the recall gradually increases. The reason for this is that in Eq. (14), as α increases, the early termination condition is gradually loosened and more CUs are terminated early. As a result, the corresponding precision gradually decreases and the recall gradually increases.

Since $\alpha = 4$ is not optimal for all probabilities in [40%~60%], we take $\alpha = 4$ as the optimal value for the median of [40%~60%], namely 50%. In a similar manner, the optimal α values for 10%, 30%, 70%, and 90% is 1, 3, 6, and 9, respectively. Concatenating the probability and its corresponding optimal value of α, we obtain a series of datasets: $(10\%, 1)$, $(30\%, 3)$, $(50\%, 4)$, $(70\%, 6)$, $(90\%, 9)$. To investigate the relationship between the probability x and the corresponding optimal value of α, the polynomial regression is applied as follows:

$$\alpha_i = k_0 + k_1 x_i + k_2 x_i^2 + \cdots + k_m x_i^m + \varepsilon_i \tag{15}$$

We use $i = 1, 2, \cdots, 5$ to denote the five datasets. x_i and α_i are the probability and its corresponding optimal α of the i-th set respectively. ε_i is the error of the i-th dataset. k_j is the j-th coefficient, and m is the greatest value of coefficient or power. To obtain the optimal value of k_j , and the minimal ε_i, we have:

$$f = \sum_{i=1}^{5} \varepsilon_i^2 = \sum_{i=1}^{5} \left(y_i - k_0 - k_1 x_i - k_2 x_i^2 - \cdots - k_m x_i^m\right)^2 \tag{16}$$

TABLE I: Coding efficiency, expressed by BDBR, under different α

Sequence \ α	1	2	3	4	5	6
sunflower	0.20%	0.20%	0.20%	0.20%	0.20%	0.20%
tractor	-0.10%	0.00%	0,00%	0.00%	0.00%	0.00%
park	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
town	-0.10%	-0.10%	-0.10%	-0.10%	-0.10%	-0.10%
Flowervase	-0.20%	-0.20%	-0.20%	-0.20%	-0.10%	-0.10%
PartyScene	-0.10%	-0.10%	-0.10%	-0.10%	-0.10%	-0.10%
BlowingBubbles	-0.10%	-0.10%	-0.10%	-0.10%	-0.20%	-0.20%
RaceHorses	-0.20%	-0.20%	-0.20%	-0.20%	-0.20%	-0.20%

where f is the sum of errors of the five datasets. To find the optimal value of k_j , we take the derivative of f with respect to k_j and set it equal to zero:

$$\frac{\partial f}{\partial k_j} = -2\sum_{i=1}^{5}\left(y_i - k_0 - k_1 x_i - k_2 x_i^2 - \cdots - k_m x_i^m\right)x_i^j = 0 \tag{17}$$

In order to achieve the optimal value of m, we have chosen a variety of different values, such as 1, 2, 3, 4, 5, 6, etc., for fitting. The fitting results suggest that the regression model has a similar fitting effect when $m \geq 3$. Therefore, we use $m = 3$ for fitting, and the corresponding expression of the fitting function is given as follows:

$$\alpha = 20.83 \cdot x^3 - 25.89 \cdot x^2 + 16.93 \cdot x - 0.4268 \tag{18}$$

Consequently, we can derive the corresponding optimal value of α for any probability of x. Combining Eqs. (14) and (18), we obtain the probability-based early termination conditions for ILR mode as shown Eq. (19).

$$\sqrt{SSD} < \left(20.83 \cdot x^3 - 25.89 \cdot x^2 + 16.93 \cdot x - 0.4268\right) \cdot \frac{\sqrt{2}\,(1-f)\,qN}{\ln 2N^2} \tag{19}$$

If the inequality in Eq. (19) holds, the condition of early termination is satisfied, thus the Intra mode can be skipped; otherwise, the Intra mode is required to be examined for further prediction.

C. Depth probability-based early termination

In our previous work [16], we managed to obtain the probability of each coding depth being selected as the best one. Again, we divided the probability range of [0%~100%] into 5 equal intervals, namely [0%~20%], [20%~40%], [40%~60%], [60%~80%], and [80%~100%]. Similar to the above ILR mode early termination, these five intervals were tested individually to obtain their corresponding optimal α.

Simulation results show that only very few CUs adopt the coding depth of 0. Therefore, we do not early terminate the coding depth of 0, as this only brings a limited improvement in coding speed. As the maximum coding depth is set to 3, it makes no sense to terminate the coding process at this depth. Therefore, we apply the early termination at coding depths of 1 and 2 only.

The same methodology as described in section (2) for ILR mode early termination is employed. When the coding depth is 1, we obtain a series of datasets: $(10\%, 1)$, $(30\%, 2)$, $(50\%, 3)$, $(70\%, 4)$, $(90\%, 7)$.

The polynomial regression is applied to the five datasets to investigate the relationship between the probability d_1 of coding depth 1 and its corresponding optimal value of α:

$$\alpha = 14.63d_1^3 - 10.47d_1^2 + 4.382d_1 - 0.8997 \tag{20}$$

For coding depth of 2, the relationship between the probability d_2 of coding depth 2 and its corresponding optimal value of α is given as

$$\alpha = 14.83d_2^3 - 13.86d_2^2 + 9.31d_2 - 1.071 \tag{21}$$

The optimal value of α for the coding depths of 1 and 2 can be obtained using Eq. (20) and Eq. (21), respectively. Substituting them into Eq. (14), we can derive the corresponding early termination decision conditions for coding depths to improve the coding speed.

Through the above process, we obtain AZB early termination conditions. However, the number of AZBs is limited. If blocks only contain partial zero quantized coefficients, their corresponding coding modes and depths are also possibly the best ones. Therefore, we further investigate PZB early termination as shown below.

V. Probability-based PZB early termination

As mentioned above, the decision condition of early termination is closely related to the probability of each mode or depth being selected as the best one. Therefore, we propose an effective PZB prediction process, and this is consequently combined with the probability of each coding mode and coding depth being chosen as the best ones to obtain the corresponding early termination conditions.

A. PZB prediction process

To accurately calculate the percentage of zero-valued coefficients, we first count the number of zero-valued coefficients based on SSD in a CU. The SSD of the residual coefficients is greater than or equal to the actual sum of the residual coefficients. Therefore, the corresponding decision conditions using the SSD of residual coefficients are too strict. In other words, if a coefficient is predicted to be zero based on SSD, it is certainly quantized to zero. Therefore, we first predict

$$\begin{bmatrix} 0.7219 & 0.9439 & 1.2095 & 1.5480 & 1.8720 & 2.1496 & 2.3606 & 2.4902 \\ 0.9439 & 1.2342 & 1.5814 & 2.0240 & 2.4476 & 2.8105 & 3.0864 & 3.2560 \\ 1.2095 & 1.5814 & 2.0265 & 2.5935 & 3.1364 & 3.6014 & 3.9550 & 4.1722 \\ 1.5480 & 2.0240 & 2.5935 & 3.3192 & 4.0140 & 4.6092 & 5.0616 & 5.3397 \\ 1.8720 & 2.4476 & 3.1364 & 4.0140 & 4.8541 & 5.5739 & 6.1211 & 6.4573 \\ 2.1496 & 2.8105 & 3.6014 & 4.6092 & 5.5739 & 6.4004 & 7.0287 & 7.4148 \\ 2.3606 & 3.0864 & 3.9550 & 5.0616 & 6.1211 & 7.0287 & 7.7187 & 8.1427 \\ 2.4902 & 3.2560 & 4.1722 & 5.3397 & 6.4573 & 7.4148 & 8.1427 & 8.5900 \end{bmatrix} \tag{24}$$

zero-valued quantized coefficients. However, if a coefficient is predicted to be non-zero based on SSD, it still may likely be quantized to zero. Therefore, we further use the actual sum of residual coefficients to calculate zero-valued coefficients. Since the HT is similar to but much simpler than the DCT transform, we then derive the PZB prediction condition based on HT. Finally, we propose early termination based on energy conservation to improve coding speed.

1) Prediction of zero-valued quantized coefficients based on SSD: As mentioned in Eq. (7), the expectation and variance of the DCT coefficient at position (u, v) are 0 and $\sigma_F(u, v)$, respectively. There is a probability of greater than 99% that a DCT coefficient locates in the interval of $(-3\sigma_F, 3\sigma_F)$ [30]. To ensure the DCT coefficient at position (u, v) being zero after quantization, the following conditions must be met

$$\frac{3\sigma_F(u,v)}{Q_{step}} + \frac{1}{6} < 1 \tag{22}$$

Combining Eqs. (6), (7) and (22), the PZB decision condition is derived as shown below.

$$\sqrt{SSD} \le \frac{20Q_{step}}{9\sqrt{[\mathrm{ARA}^T]_{uu}[\mathrm{ARA}^T]_{vv}}} \tag{23}$$

The right part in Eq. (23) is given by a matrix in a numerical form as shown in Eq. (24).

Given the obtained matrix of thresholds, the zero-valued elements can be produced for quantized coefficients. If SSD is smaller than a threshold value in the matrix, scanning in the Zigzag order starting from the upper left corner towards the bottom right corner, the DCT coefficients from the upper left corner to the position of the thresholded value are quantized to be non-zero values, and the subsequent coefficients are quantized to zero [30]. For example, the quantized residual coefficients are all zeros if $\sqrt{SSD} < 0.7219Q_{step}$ is applied; the quantized residual coefficient at position $(0, 0)$ is non-zero only if $\sqrt{SSD} < 0.9439Q_{step}$ is applied; the quantized residual coefficients at positions $(0, 0)$, $(0, 1)$ and $(1, 0)$ are non-zero if $\sqrt{SSD} < 1.2095Q_{step}$ is applied, and so forth. All zero-valued quantized coefficients in a matrix can be predicted in the same way based on the threshold matrix (24).

2) PZB prediction condition based on HT: As mentioned above, The SSD-based decision conditions are so strict that it is possible for some predicted non-zero DCT coefficients to be quantized to zero-valued coefficients. Therefore, for those that do not satisfy the above decision conditions, we need to further predict DCT coefficients. HT is quite similar to DCT but much simpler. For both the HT coefficient and the DCT coefficient at the same position, if they are sufficiently close, we use HT rather than DCT to obtain the zero-valued coefficient. We use z_1 and z_2 to denote the DCT transformed coefficient and its corresponding HT coefficient at position (u, v), respectively.

The corresponding decision condition is written as

$$\left| \frac{z_1}{Q_{step}} - \frac{z_2}{Q_{step}} \right| \le k \tag{25}$$

From Eq. (25), we have:

$$|z_1 - z_2| \le kQ_{step} \tag{26}$$

The DCT transformed coefficient z_1 and the HT transformed coefficient z_2 are given by

$$z_1 = \sum_{i=0}^{7}\sum_{j=0}^{7} d_{ui}x_{ij}d_{jv}, z_2 = \sum_{i=0}^{7}\sum_{j=0}^{7} e_{ui}x_{ij}e_{jv} \tag{27}$$

where d_{ui} and e_{ui} are the elements at position (u, i) in the DCT and HT matrices, respectively, x_{ij} is the residual coefficient at position (i, j) . From Eq. (27), we have

$$\begin{aligned} |z_1 - z_2| &= \left| \sum_{i=0}^{7}\sum_{j=0}^{7} d_{ui}x_{ij}d_{jv} - \sum_{i=0}^{7}\sum_{j=0}^{7} e_{ui}x_{ij}e_{jv} \right| \\ &= \left| \sum_{i=0}^{7}\sum_{j=0}^{7} \left(d_{ui}d_{jv} - e_{ui}e_{jv}\right) x_{ij} \right| \end{aligned} \tag{28}$$

According to the Cauchy's inequality, we have

$$\begin{aligned} &\left| \sum_{i=0}^{7}\sum_{j=0}^{7} \left(d_{ui}d_{jv} - e_{ui}e_{jv}\right) x_{ij} \right| \le \\ &\sqrt{\sum_{i=0}^{7}\sum_{j=0}^{7} \left(d_{ui}d_{jv} - e_{ui}e_{jv}\right)^2} \sqrt{\sum_{i=0}^{7}\sum_{j=0}^{7} x_{ij}^2} \end{aligned} \tag{29}$$

From Eq. (29), we get

$$\begin{aligned} &\sqrt{\sum_{i=0}^{7}\sum_{j=0}^{7} \left(d_{ui}d_{jv} - e_{ui}e_{jv}\right)^2} \sqrt{SSD} \le \\ &\sqrt{2\sum_{i=0}^{7}\sum_{j=0}^{7} \left(\left(d_{ui}d_{jv}\right)^2 + \left(e_{ui}e_{jv}\right)^2\right)} \sqrt{SSD} \end{aligned} \tag{30}$$

then

$$\sum_{i=0}^{7}\sum_{j=0}^{7} \left(\left(d_{ui}d_{jv}\right)^2 + \left(e_{ui}e_{jv}\right)^2\right) = \sum_{i=0}^{7} d_{ui}^2 \sum_{j=0}^{7} d_{jv}^2 + \sum_{i=0}^{7} e_{ui}^2 \sum_{j=0}^{7} e_{jv}^2 \tag{31}$$

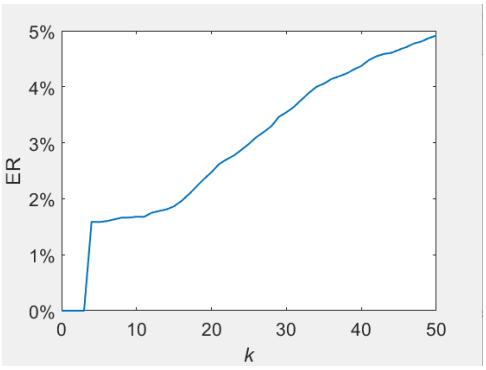

Fig. 2: Relationship between k and the error rate

As the sum of squares of the coefficients in each row of the DCT and HT matrices is always 1, we get:

$$\sum_{i=0}^{7} d_{ui}^2 \sum_{j=0}^{7} d_{jv}^2 + \sum_{i=0}^{7} e_{ui}^2 \sum_{j=0}^{7} e_{jv}^2 = 2 \tag{32}$$

Combining Eqs. (26), (28) and (32), we have:

$$2\sqrt{SSD} \leq kQ_{step} \tag{33}$$

It can be written as:

$$SSD \leq \frac{k^2}{4} Q_{step}^2 \tag{34}$$

If the condition in Eq. (34) holds, the quantized zero-valued coefficients can be calculated according to the HT transformed coefficients as shown below:

$$\frac{z_2}{Q_{step}} + \frac{1}{6} < 1 \tag{35}$$

If the condition in Eq. (35) holds, z_2 is quantized to 0. In this way, we can predict all of the coefficients that are to be quantized to zero.

To find the optimal value of k, we evaluate a number of values. The QP values used for the BL and the EL are 26 and 20, and the corresponding error rate (ER) which is the percentage of the numbers of mistakenly predicted CUs to the total number of CUs is shown in Fig. 2.

In Fig. 2, the x-axis denotes the value of k, and the y-axis is the corresponding ER. When k = 50, the corresponding ER is 4.9%. In Eq. (25), k is the absolute difference between the quantized HT coefficient and the quantized DCT coefficient at the same position. Obviously, the greater the k value, the greater the difference and the ER. Fig. 2 also verifies that ER increases gradually as the value of k increases. Therefore, when k is greater than 50, the corresponding ER should be greater than 4.9%. To ensure accuracy, k = 50 is chosen as the optimal value. For a CU satisfying the condition in Eq. (34), HT is used for the calculation to determine whether the obtained HT transformed coefficient is a quantized zero-valued coefficient.

Through this process, we can obtain the quantized zero-valued coefficients. In this way, each HT coefficient also takes HT calculation cost. After some HT coefficients have been obtained, if we can early predict that the subsequent HT coefficients will be quantized to zero, we can skip the calculation for these HT coefficients. Therefore, we propose early termination PZB as shown below.

3) Early termination based on energy conservation: According to the law of conservation of energy, the sum of the squares of the coefficients in the residual matrix is equal to that in the matrix of transformed coefficients, which can be written as

$$\sum_{i=0}^{N} r_i^2 = \sum_{i=0}^{N} T_i^2 = s \tag{36}$$

where r_i and T_i are the i-th residual coefficient and the i-th HT coefficient scanned in Zigzag order. N is the number of coefficients in the matrix. s is the sum of the squares of all the residual coefficients in the residual matrix or the sum of the squares of all the TH coefficients in the TH coefficients matrix. This forms the basis of the proposed early termination.

We first calculate the sum of the squares of all the residual coefficients in the residual matrix s, and the coefficients in the HT transformed matrix are then scanned in Zigzag order to examine if the condition in Eq. (37) holds or not.

$$\sum_{i=0}^{k} T_i^2 + T_k^2 \geq s \tag{37}$$

Combining Eqs. (36) and (37), we have:

$$T_k^2 \geq \sum_{i=k+1}^{N} T_i^2 \tag{38}$$

Therefore, if Eq.(38) holds, it means that T_k is greater than the subsequent HT coefficients. If T_k is zero after quantization, its subsequent HT quantized coefficients must be zero. The calculation process for the subsequent HT coefficients can be terminated.

This early termination is developed for the HT coefficients. If a CU does not satisfy Eq. (34), we can calculate the DCT coefficients instead of the HT coefficients.

The above process deduces the PZB decision condition for 8×8 CUs. For 16×16, 32×32 and 64×64 CUs, they can be divided into several 8×8 CUs for calculation.

B. Mode probability-based PZB early termination

Following the above process, we can obtain the number of zero-valued quantized coefficients for the entire CU. According to the percentage of zero-valued quantized coefficients, we divide it into eight intervals, namely, [0~0.125], [0.125~0.25], [0.25~0.375], [0.375~0.5], [0.5~0.625], [0.625~0.75], [0.75~0.875], and [0.875~1]. Similar to the previous section, we divide the probability of the ILR mode being selected as the best one into five intervals: [0%~20%], [20%~40%], [40%~60%], [60%~80%], and [80%~100%]. Consequently, each of these intervals is evaluated to select the optimal zero-valued quantized coefficient interval. When the probability of the ILR mode being selected as the best one is in the range of between 40% and 60%, we test the above eight intervals respectively. The corresponding coding efficiency is presented in Table II.

TABLE II: Coding efficiency, expressed by BDBR, for various percentage intervals of zero-valued coefficients

Sequence \ Intervals	0~0.125	0.125~0.25	0.25~0.375	0.375~0.5	0.5~0.625	0.625~0.75	0.75~0.875	0.875~1
sunflower	-0.10%	-0.10%	-0.10%	-0.10%	0.00%	0.00%	0.00%	0.10%
tractor	-0.20%	-0.20%	-0.20%	-0.20%	-0.10%	-0.10%	0.00%	0.00%
park	-0.20%	-0.20%	-0.20%	-0.20%	-0.10%	-0.10%	0.00%	0.00%
town	-0.40%	-0.40%	-0.40%	-0.30%	-0.30%	-0.20%	-0.10%	0.00%
Flowervase	0.80%	0.80%	0.70%	0.50%	0.40%	0.30%	0.10%	0.00%
PartyScene	-0.60%	-0.60%	-0.60%	-0.40%	-0.30%	-0.30%	0.00%	0.00%
BlowingBubbles	0.60%	0.60%	0.60%	0.50%	0.40%	0.20%	0.10%	0.10%
RaceHorses	-0.40%	-0.40%	-0.40%	-0.30%	-0.30%	-0.20%	-0.10%	-0.10%

It can be seen from Table II that a lower percentage of zero-value coefficients can lead to performance improvements. However, these performance gains come from prediction errors. We have analyzed the reasons for this in the last paragraph of this manuscript. Prediction errors sometimes decrease and sometimes increase the BDBR. For example, "Flowervase" and "BlowingBubbles" have positive BDBR, but the other sequences have negative BDBR. Clearly, a lower percentage of zero-valued coefficients does not give very good performance for all sequences. As the percentage of zero-valued coefficients increases, the corresponding CU is predicted more accurately by the current mode or depth, so the interval of 0.875~1 can obtain very good performance across all sequences. Therefore, we choose the interval of [0.875~1].

The experimental results show that the average precision under the intervals of [0~0.125], [0.125~0.25], [0.25~0.375], [0.375~0.5], [0.5~0.625], [0.625~0.75], [0.75~0.875], and [0.875~1] are 80.97%, 82.40%, 83.72%, 87.15%, 89.75%, 92.18%, 96.80% and 98.75% respectively; and the corresponding average recalls are 4.81%, 5.14%, 6.94%, 8.51%, 11.62%, 16.51%, 21.44% and 25.04% respectively. We can observe that, as the percentage interval of zero-valued coefficients increases, their corresponding precision gradually increases. This is because as the percentage of zero-value coefficients increases, the corresponding CU is predicted more accurately by the current mode or depth.

In a similar way as PBAET, the median of the interval of [0.875~1], namely 0.9375, is taken as the optimal value for the mode probability between 40% and 60%. It is not suitable to take 0.9375 as the optimal value for the entire of [40%~60%], so it is taken as the optimal value for the median of [40%~60%], namely 50%. In a similar manner, we obtain five sets of data: $(10\%, 1)$, $(30\%, 1)$, $(50\%, 0.9375)$, $(70\%, 0.8125)$, $(90\%, 0.6875)$. Again, the polynomial regression is applied to establish the relationship between the probability of the ILR mode being selected as the best coding mode and the percentage of zero coefficients in an 8×8 CU:

$$\beta = 1.628 \cdot x^4 - 2.604 \cdot x^3 + 0.6185 \cdot x^2 + 0.02604 \cdot x + 0.9937 \quad (39)$$

where x is the probability of the ILR mode being selected as the best coding mode, and β represents the proportional threshold of zero-valued coefficients in an 8×8 CU. According to Eq. (39), we obtain the proportional threshold of the quantized zero-valued coefficient ratio of the probability of the current ILR mode. If the percentage of zero-valued coefficients in the current CU is greater than β, we skip the Intra mode. Otherwise, we need to use the Intra mode for further prediction.

C. Depth probability-based PZB early termination

The early termination is applied at the coding depths of 1 and 2 using the same process of the ILR mode. The probability of the current CU adopting the coding depth of 1 is divided into five intervals. The percentage of quantized zero-valued coefficients is divided into eight intervals. We obtain five datasets:$(10\%, 0.9375)$, $(30\%, 0.8125)$, $(50\%, 0.8125)$, $(70\%, 0.6875)$, and $(90\%, 0.4375)$.

The polynomial regression is applied to the five datasets to obtain the relationship between the probability (d_1) of coding depth 1 being selected as the best depth and the corresponding percentage of quantized zero-valued coefficients (β):

$$\beta = -2.107 \cdot d_1^3 + 2.43 \cdot d_1^2 - 1.116 \cdot d_1 + 1.008 \quad (40)$$

In the same way, five datasets for coding depth 2 are obtained, namely $(10\%, 0.9375)$, $(30\%, 0.6875)$, $(50\%, 0.5625)$, $(70\%, 0.4375)$ and $(90\%, 0.1875)$. Applying the polynomial regression to the five datasets, the relationship between the probability (d_2) of coding depth 2 being selected as the best depth and the corresponding percentage of quantized zero-valued coefficients (β) is obtained:

$$\beta = -1.076 \cdot d_2^3 + 1.427 \cdot d_2^2 - 1.317 \cdot d_2 + 1.009 \quad (41)$$

The thresholds of the zero-valued coefficient percentage of coding depths 1 and 2 can be obtained through Eqs. (40) and (41). If the percentage of quantized zero-valued coefficients in the current CU is greater than β , we can early terminate the evaluation on coding depth; otherwise, we need to use subsequent depths for further prediction.

VI. Experimental Results

To verify the effectiveness of the proposed algorithm, we employ the reference software (SHM 11.0) on a server with Intel (R) 2.0 GHz CPU and 30 GB memory for testing. The encoding parameters specified in CTC [29] are used. The QP factors for the BL are (26, 30, 34, 38). Two sets of QP factors (20, 24, 28, 32) and (22, 26, 30, 34) are used for the EL. For simplicity, the QP factors of (20, 24, 28, 32) and (22, 26, 30, 34) are denoted as C1 and C2, respectively. The experimental performance is measured in terms of coding efficiency and coding speed.

TABLE III: Performance Comparison between PBAET and PBPET

Sequence	PBAET		PBPET	
	TS(%)	BDBR(%)	TS(%)	BDBR(%)
Traffic	79.40	-0.34	75.58	-0.36
PeopleOnStreet	80.97	-0.37	71.69	-0.37
Kimono	77.80	-0.31	83.31	-0.32
ParkScene	74.15	-0.27	73.34	-0.28
Cactus	72.06	-0.21	65.17	-0.19
BasketballDrive	72.65	-0.10	80.41	-0.09
BQTerrace	62.71	-0.09	52.85	-0.08
Average	74.25	-0.24	71.76	-0.24

The coding efficiency is measured by BDBR [32] and the coding speed is evaluated by TS, which represents the proportional reduction in encoding time for the EL.

The proposed algorithm includes the PBAET and the PBPET. It is observed that the two QP sets for the EL produced similar experimental results. We, therefore, present the results in Table III when the QP set of (22, 26, 30, 34) is used for the EL.

It is seen from Table III that the coding speed improved by PBAET and PBPET for the EL is 74.25% and 71.76%, respectively. The BDBR is decreased by 0.24% for both PBAET and PBPET. The experimental results show that the improvement in coding speed achieved by PBAET is more significant than that achieved by PBPET, and this is because PBPET needs additional computational overhead, which reduces the improvement of coding speed.

In order to further demonstrate the effectiveness of the proposed algorithms, the overall performance when combining the two proposed strategies, i.e. "PBAET" and "PBPET", is evaluated and compared with EMSIP [15], FDMDIP [16] and HAZSQBD [28]. The early termination for modes and depths in EMSIP [15] and FDMDIP [16] and the AZB decision conditions in HAZSQBD [28] are used to early terminate the mode and depth selection. To the best of our knowledge, EMSIP [15] and FDMDIP [16] represent the state-of-the-art performance in early termination for SHVC, and the best AZB decision condition is provided in HAZSQBD [28]. Table IV and Table V compare the coding performance of the proposed algorithms with that of the state-of-the-art methods.

As can be seen from Table IV, the computation time of the proposed algorithm is decreased by 79.76%, and the computation time of EMSIP, FDMDIP and HAZSQBD are reduced by 59.10%, 71.30% and 50.16%, respectively. The BDBR of the proposed algorithm is -0.25, and the BDBRs of EMSIP, FDMDIP and HAZSQBD are -0.20, -0.25 and -0.22, respectively. Compared with EMSIP, FDMDIP and HAZSQBD, the proposed algorithm saves additional encoding time of 20.66%, 8.46% and 29.60%, with 0.05, 0 and 0.03 BDBR savings, respectively. As can be seen from Table V, the computation time of the proposed algorithm is decreased by 81.45%, and the computation time of EMSIP, FDMDIP and HAZSQBD is reduced by 59.82%, 74.68% and 51.84%. The BDBR of the proposed algorithm is -0.34, and the BDBRs of EMSIP, FDMDIP and HAZSQBD are -0.11, -0.36 and -0.23, respectively. Compared with EMSIP, FDMDIP and HAZSQBD, the proposed algorithm saves additional encoding time of 18.27%, 5.21% and 31.85% with 0.23, -0.02 and 0.11 BDBR savings, respectively.

Therefore, the proposed algorithm is effectively faster than the other three methods under the conditions of C1 and C2. This is because more unlikely modes and depths are skipped. Thus, the coding speed is significantly improved. In Table IV and Table V, the coding speed of all the four algorithms under the condition of C2 is slightly higher than that under the condition of C1. Because QP factors used in the condition of C2 are greater than those in C1, more AZB and PZB appear under the condition of C2.

From Table IV and Table V, we can observe that the coding speed is significantly improved. In quality SHVC, for a CU in the EL and its co-located CU in the BL, only the QP factors used in the different layers are different. Consequently, there is a strong inter-layer correlation and the ILR mode can be used to accurately predict the corresponding CU in the EL. Therefore, a great number of CUs adopt the ILR mode for prediction at a low coding depth, such as the coding depth of 0 or 1. Compared with the Intra mode and the deep coding depths, the ILR mode and low coding depths generally require less time for evaluation. The proposed PBAET and PBPET strategies can effectively skip the time-consuming evaluations for the Intra mode and the deep coding depths, thus significantly improving the coding speed.

The improvement in coding speed generally leads to a decrease in coding efficiency, namely an increase in BDBR. However, it is observed from Tables I to V that coding efficiency is occasionally improved compared with the SHM reference software. This can be attributed to the Intra prediction process, where CUs are predicted by reference pixels. It is apparent that when the texture of CUs and their reference pixels are more similar, the corresponding Intra prediction is more accurate, and the generated RD costs are smaller accordingly, and vice versa. For a particular CU to be coded, different methods lead to different neighboring CUs, which contain different reference pixels, and in turn, lead to different RD costs for the CU to be coded. In a word, this is caused by RD dependency among coding CUs [33] [34]. We have analyzed the reason why our method decreases and increases the BDBR at times in detail [16]. To further enhance coding efficiency, some optimized bit allocation schemes may be considered [35] [36].

VII. Conclusion

In order to improve the speed of the Intra coding process in QSHVC, this paper proposes an efficient probability-based zero block early termination algorithm. We first derive the AZB decision conditions based on the distribution of residual coefficients and develop the PZB predict process to obtain the proportion of zero-valued quantized coefficients, and then combine mode and depth probabilities to early terminate the process of mode and depth selections. The experimental results demonstrate that the proposed algorithm can significantly improve the speed of video coding with a noticeable increase

TABLE IV: Performance comparison among different algorithms under the condition of C1

Sequence	Proposed		EMSIP [15]		FDMDI [16]		HAZSQBD [28]	
	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Traffic	-0.37	79.67	-0.29	53.38	-0.44	77.18	-0.27	49.45
PeopleOnStreet	-0.38	79.68	-0.26	46.13	-0.40	76.64	-0.27	49.11
Kimono	-0.31	80.06	-0.21	69.17	-0.36	74.68	-0.21	51.47
ParkScene	-0.29	78.41	-0.23	55.23	-0.29	67.61	-0.24	49.28
Cactus	-0.16	80.46	-0.28	62.38	-0.23	69.06	-0.25	48.23
BasketballDrive	-0.09	80.51	-0.03	66.41	0.01	69.41	-0.16	51.05
BQTerrace	-0.12	79.53	-0.08	60.99	-0.06	64.53	-0.12	52.52
Average	-0.25	79.76	-0.20	59.10	-0.25	71.30	-0.22	50.16

TABLE V: Performance comparison among different algorithms under the condition of C2

Sequence	Proposed		EMSIP [15]		FDMDIP [16]		HAZSQBD [28]	
	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)	BDBR(%)	TS(%)
Traffic	-0.30	82.35	-0.17	49.95	-0.57	69.67	-0.27	53.12
PeopleOnStreet	-0.20	81.67	-0.07	52.54	-0.39	73.95	-0.29	51.56
Kimono	-0.20	81.91	-0.26	68.95	-0.40	77.84	-0.22	52.52
ParkScene	-0.40	80.74	-0.04	58.12	-0.37	75.28	-0.27	52
Cactus	-0.50	81.57	-0.09	64.58	-0.28	75.56	-0.25	51.21
BasketballDrive	-0.40	80.50	-0.14	62.23	-0.37	75.29	-0.18	48.65
BQTerrace	-0.40	81.39	-0.01	62.35	-0.13	75.14	-0.15	53.81
Average	-0.34	81.45	-0.11	59.82	-0.36	74.68	-0.23	51.84

in coding efficiency. These results show that the proposed algorithm can efficiently simulcast different quality versions of the same video content in one bit-stream, which is of great use for broadcasters and end-users alike. In future, we plan to develop zero-block algorithms for 3D-HEVC [37] [38], Multiview Video Coding [39] and Screen Content Coding [40] [41].

REFERENCES

[1] D. Wang, Y. Sun, J. Liu, F. Dufaux, X. Lu, and B. Hang, "Probability-Based Fast Intra Prediction Algorithm for Spatial SHVC," *IEEE Transactions on Broadcasting*, vol. 68, no. 1, pp. 83–96, 2022.

[2] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, 2016.

[3] H. R. Tohidypour, M. T. Pourazad, and P. Nasiopoulos, "Probabilistic Approach for Predicting the Size of Coding Units in the Quad-Tree Structure of the Quality and Spatial Scalable HEVC," *IEEE Transactions on Multimedia*, vol. 18, no. 2, pp. 182–195, 2016.

[4] H. R. Tohidypour, M. T. Pourazad, and P. Nasiopoulos, "An Encoder Complexity Reduction Scheme for Quality/Fidelity Scalable HEVC," *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 664–674, 2016.

[5] H. R. Tohidypour, H. Bashashati, M. T. Pourazad, and P. Nasiopoulos, "Online-Learning-Based Mode Prediction Method for Quality Scalable Extension of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 10, pp. 2204–2215, 2017.

[6] H. Zhang and Z. Ma, "Fast Intra Mode Decision for High Efficiency Video Coding (HEVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 660–668, 2014.

[7] S. Cho and M. Kim, "Fast CU Splitting and Pruning for Suboptimal CU Partitioning in HEVC Intra Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 9, pp. 1555–1564, 2013.

[8] J. Zhang, S. Kwong, and X. Wang, "Two-Stage Fast Inter CU Decision for HEVC Based on Bayesian Method and Conditional Random Fields," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 11, pp. 3223–3235, 2018.

[9] T. Zhang, M.-T. Sun, D. Zhao, and W. Gao, "Fast Intra-Mode and CU Size Decision for HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 8, pp. 1714–1726, 2017.

[10] D. Wang, Y. Sun, C. Zhu, W. Li, and F. Dufaux, "Fast Depth and Inter Mode Prediction for Quality Scalable High Efficiency Video Coding," *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 833–845, 2020.

[11] D. Wang, Y. Sun, W. Li, L. Xie, X. Lu, F. Dufaux, and C. Zhu, "Hybrid Strategies for Efficient Intra Prediction in Spatial SHVC," *IEEE Transactions on Broadcasting*, pp. 1–14, 2022.

[12] D. Wang, X. Wang, Y. Sun, W. Li, X. Lu, and F. Dufaux, "Gaussian Distribution-based Mode Selection for Intra Prediction of Spatial SHVC," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 2711–2715.

[13] D. Wang, Y. Sun, W. Li, L. Xie, X. Lu, F. Dufaux and C. Zhu, "A NOVEL MODE SELECTION-BASED FAST INTRA PREDICTION ALGORITHM FOR SPATIAL SHVC," in *The 48th IEEE International Conference on Acoustics, Speech, & Signal Processing (ICASSP)*, 2023(submitted).

[14] H. L. Tan, C. C. Ko, and S. Rahardja, "Fast Coding Quad-Tree Decisions Using Prediction Residuals Statistics for High Efficiency Video Coding (HEVC)," *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 128–133, 2016.

[15] D. Wang, C. Zhu, Y. Sun, F. Dufaux, and Y. Huang, "Efficient Multi-Strategy Intra Prediction for Quality Scalable High Efficiency Video Coding," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2063–2074, 2019.

[16] D. Wang, Y. Sun, C. Zhu, W. Li, F. Dufaux, and J. Luo, "Fast Depth and Mode Decision in Intra Prediction for Quality SHVC," *IEEE Transactions on Image Processing*, vol. 29, pp. 6136–6150, 2020.

[17] S.-W. Jung, S.-J. Baek, C.-S. Park, and S.-J. Ko, "Fast Mode Decision Using All-Zero Block Detection for Fidelity and Spatial Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 2, pp. 201–206, 2010.

[18] 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 Transactions on Broadcasting*, vol. 60, no. 2, pp. 405–412, 2014.

[19] Z. Pan, P. Zhang, B. Peng, N. Ling, and J. Lei, "A CNN-Based Fast Inter Coding Method for VVC," *IEEE Signal Processing Letters*, vol. 28, pp. 1260–1264, 2021.

[20] P.-T. Chiang and T. S. Chang, "Fast zero block detection and early CU termination for HEVC Video Coding," in *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 1640–1643.

[21] L. Fu-jiang and Z. Gang, "Predictive skip mode based all zero block detection for AVS encoder," in *2009 ISECS International Colloquium on Computing, Communication, Control, and Management*, vol. 2, 2009, pp. 490–493.

[22] H. Wang, S. Kwong, and C.-W. Kok, "Efficient prediction algorithm of integer DCT coefficients for H.264/AVC optimization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 4, pp. 547–552, 2006.
[23] H. Wang and S. Kwong, "Hybrid Model to Detect Zero Quantized DCT Coefficients in H.264," *IEEE Transactions on Multimedia*, vol. 9, no. 4, pp. 728–735, 2007.
[24] H. Wang and S. Kwong, "Prediction of Zero Quantized DCT Coefficients in H.264/AVC Using Hadamard Transformed Information," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 510–515, 2008.
[25] B. Lee, J. Jung, and M. Kim, "An All-Zero Block Detection Scheme for Low-Complexity HEVC Encoders," *IEEE Transactions on Multimedia*, vol. 18, no. 7, pp. 1257–1268, 2016.
[26] H. Fan, R. Wang, L. Ding, X. Xie, H. Jia, and W. Gao, "Hybrid Zero Block Detection for High Efficiency Video Coding," *IEEE Transactions on Multimedia*, vol. 18, no. 3, pp. 537–543, 2016.
[27] S. Ahn, B. Lee, and M. Kim, "A Novel Fast CU Encoding Scheme Based on Spatiotemporal Encoding Parameters for HEVC Inter Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 422–435, 2015.
[28] J. Cui, R. Xiong, X. Zhang, S. Wang, S. Wang, S. Ma, and W. Gao, "Hybrid All Zero Soft Quantized Block Detection for HEVC," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 4987–5001, 2018.
[29] X. Li, J. Boyce, P. Onno, and Y. Ye, "JCTVC-M1009: Common SHM test conditions and software reference configurations," 08 2018.
[30] I.-M. Pao and M.-T. Sun, "Modeling DCT coefficients for fast video encoding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, pp. 608–616, 1999.
[31] B. Lee, M. Kim, and T. Q. Nguyen, "A Frame-Level Rate Control Scheme Based on Texture and Nontexture Rate Models for High Efficiency Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 3, pp. 465–479, 2014.
[32] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T VCEG-M33, April, 2001*, pp. 1–4, 2001.
[33] S. Li, C. Zhu, Y. Gao, Y. Zhou, F. Dufaux, and M.-T. Sun, "Inter-frame dependent rate-distortion optimization using lagrangian multiplier adaption," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, 2015, pp. 1–6.
[34] Y. Gao, C. Zhu, S. Li, and T. Yang, "Temporally Dependent Rate-Distortion Optimization for Low-Delay Hierarchical Video Coding," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4457–4470, 2017.
[35] F. Song, C. Zhu, Y. Liu, Y. Zhou, and Y. Liu, "A new GOP level bit allocation method for HEVC rate control," in *2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, 2017, pp. 1–4.
[36] H. Guo, C. Zhu, S. Li, and Y. Gao, "Optimal Bit Allocation at Frame Level for Rate Control in HEVC," *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 270–281, 2019.
[37] Z. Pan, F. Yuan, W. Yu, J. Lei, N. Ling, and S. Kwong, "RDEN: Residual Distillation Enhanced Network-Guided Lightweight Synthesized View Quality Enhancement for 3D-HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 6347–6359, 2022.
[38] Z. Pan, H. Zhang, J. Lei, Y. Fang, X. Shao, N. Ling, and S. Kwong, "DACNN: Blind Image Quality Assessment via a Distortion-Aware Convolutional Neural Network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7518–7531, 2022.
[39] J. Lei, Z. Zhang, Z. Pan, D. Liu, X. Liu, Y. Chen, and N. Ling, "Disparity-Aware Reference Frame Generation Network for Multiview Video Coding," *IEEE Transactions on Image Processing*, vol. 31, pp. 4515–4526, 2022.
[40] L. Zhao, T. Lin, D. Zhang, K. Zhou, and S. Wang, "An Ultra-Low Complexity and High Efficiency Approach for Lossless Alpha Channel Coding," *IEEE Transactions on Multimedia*, vol. 22, no. 3, pp. 786–794, 2020.
[41] L. Zhao, K. Zhou, J. Guo, S. Wang, and T. Lin, "A Universal String Matching Approach to Screen Content Coding," *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 796–809, 2018.