IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 February 2024, accepted 12 March 2024, date of publication 18 March 2024, date of current version 22 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3378285

RESEARCH ARTICLE

Multi-Layer Features Fusion Model-Guided Low-Complexity 3D-HEVC Intra Coding

CHANG LIU[1] AND KEBIN JIA[2], (Member, IEEE)
[1]Research Center for Intelligent Information Technology, Nantong University, Nantong 226019, China
[2]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Corresponding author: Chang Liu (liuchang@ntu.edu.cn)

This work was supported in part by Beijing Natural Science Foundation under Grant 4212001, and in part by Natural Science Research Program for Higher Education in Jiangsu Province under Grant 23KJB520030.

ABSTRACT Three-dimensional (3D) video with depth information is essential for many applications in the consumer electronics industry. The 3D-high efficiency video coding (3D-HEVC) is the latest 3D video coding standard. Nonetheless, it utilizes various complex coding techniques to create extra intermediate views for better representation of 3D videos, which imposes significant challenges for real-time 3D video applications. Specifically, the high complexity of 3D-HEVC intra coding could be a significant barrier to the adoption of 3D video in consumer electronics. Therefore, in this research, a low-complexity 3D-HEVC intra coding technique is proposed. Firstly, we perform a complexity analysis of 3D-HEVC intra coding. Secondly, we develop a multi-layer features fusion (MLFF) model to estimate the optimal coding tree unit (CTU) depth and prediction unit (PU) mode. Thirdly, to improve the model's prediction accuracy, we incorporate two external features into the model: the quantization parameter (QP) and texture complexity. Finally, we embed the MLFF model into the 3D-HEVC test platform. The experimental results demonstrate that the suggested method can effectively reduce the 3D-HEVC intra coding time with a small amount of rate-distortion (RD) performance loss while maintaining the subjective quality of the synthesized view.

INDEX TERMS 3D video, 3D-HEVC, intra coding, low-complexity, multi-layer features fusion model.

I. INTRODUCTION

With the rapid development of Internet and multimedia technology and the wide popularity of intelligent terminals, information visualization has become the main trend in the development of the global consumer electronics industry. As a result, video has become an important medium for transmitting information. Faced with increasing demands for video telepresence, realism, and interactivity, three-dimensional (3D) video [1] with depth information has emerged as a hotspot in video applications.

To promote the wide application of 3D video in the consumer electronics industry, some members of the motion picture expert group (MPEG) and video coding expert group (VCEG) cooperated to form a joint collaborative team on three dimensional video (JCT-3V) to jointly develop the 3D video coding standard [2], named as 3D-high efficiency video coding (3D-HEVC) [3]. 3D-HEVC, an extension standard of HEVC, was finalized in 2015. Up to now, 3D-HEVC is the newest 3D video coding standard, which uses multiview video plus depth (MVD) [4] format to represent 3D scene. The MVD format includes two or three views, each of which contains the texture video and its corresponding depth video. In the MVD, the intermediate views can be synthesized by decoder through depth image based rendering (DIBR) [5] technology. Furthermore, while coding 3D video in the MVD format, if intra coding is utilized, it can not only encode and decode independently, but also generate high-quality video sequences. Consequently, 3D video often utilizes 3D-HEVC intra coding to ensure viewing quality [6].

When coding the texture video in 3D-HEVC intra coding, it uses the same quad-tree partition structure [7] and prediction method [8] as HEVC [9]. However, it does not remain similar when coding the depth video. In contrast to

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

the texture video, the depth video [10] represents the depth information of the target object in the scene, which has a large surface with a smooth texture and a sharp edge. If the depth video is still coded with the original HEVC coding tools [11], a significant amount of compression artifacts will occur in the synthesized virtual view during decoding, resulting in greater subjective quality distortion [12]. As a result, in order to ensure the quality of the synthesized virtual view, it is important to include coding technologies that are more consistent with depth video during the 3D-HEVC intra coding process. To achieve this objective, 3D-HEVC uses a variety of complex coding techniques [13], [14], [15] for depth video. However, it results in considerable computational complexity, which seriously restricts the adoption of 3D-HEVC in real-time applications.

Therefore, developing a low-complexity 3D-HEVC intra coding technique is a critically important research task. To address the real-time application requirement for 3D video, researchers are focusing on how to accelerate the 3D-HEVC intra coding process while ensuring subjective quality. Several studies have presented strategies to minimize the computational complexity of 3D-HEVC. Some of these techniques focus on the partitioning of the coding tree unit (CTU), while others are concerned with the mode selection of the prediction unit (PU). Few studies, however, are capable of simultaneously and accurately predicting CTU partition depth and PU mode in advance.

Deep learning's remarkable feature extraction abilities have inspired researchers to gradually apply it to the field of video coding. Motivated by this observation, this study presents a multi-layer feature fusion (MLFF) model-guided low-complexity 3D-HEVC intra coding method.

The main contributions are summarized as follows:

1) We perform a complexity analysis of 3D-HEVC intra coding utilizing statistical techniques, then propose the complexity-guided fast 3D-HEVC intra coding method.

2) We suggest a preliminary estimation of the complexity level of CTU based on its texture complexity and quantization parameters (QP) features.

3) We develop an MLFF model that introduces two external features, QP and texture complexity, to predict the CTU depth and PU mode in advance, which reduces the complexity of 3D-HEVC at the intra-mode.

4) We embed the constructed MLFF model into the 3D-HEVC testing platform. The experimental results indicate that the proposed MLFF model-guided low-complexity 3D-HEVC intra coding approach can efficiently reduce coding time with a small rate-distortion (RD) performance loss while ensuring the subjective quality of the synthesized view.

The rest of this paper is organized as follows. Section II reviews previous research on 3D-HEVC complexity reduction. The examination of 3D-HEVC intra coding complexity is reported in Section III. Section IV describes the MLFF model-guided low-complexity 3D-HEVC intra coding method. Section V summarizes the experimental findings. Finally, Section VI concludes this paper.

II. RELATED WORKS

In this section, we classify the current 3D-HEVC low-complexity intra coding works into three categories: CTU partition decision, PU mode selection, and a combination of CTU partition decision and PU mode selection.

The CTU partition decision can be roughly classified into traditional-based methods and learning-based methods. The early CTU partition decision mainly relied on traditional methods. Among them, Zhang et al. [16] skipped some tree blocks based on the correlations of inter-view, spatial-temporal, and texture-depth. Fu et al. [17] designed a two-step adaptive selection method, which was used to split the CTU. Zhang et al. [18] proposed to skip some CTUs based on statistically analysis of spatial and inter-view correlations in 3D video content.

However, the above-mentioned methods mainly employed the decision rules by manually extracting texture complexity, RD cost, and other features to determine the CTU partition depth in advance and reduce the 3D-HEVC intra coding complexity. They are not suitable for video sequences with complex textures. Recently, with the rapid development of deep learning in video coding, some researchers utilized deep learning to design decision rules. Saldanha et al. [19] utilized data mining to build the static decision trees that defined the CTU partition. This approach, however, is only suitable for video sequences with a relatively simple background. It is impossible to mine the general video features of a video sequence with a complex background. Given this, Zhang et al. [20] introduced a convolutional neural network (CNN) into 3D-HEVC for the first time. Later, Liu et al. [21] proposed a CNN network with adaptive QP to address the problem of high computational complexity caused by CTU iterative partition. Zhang et al. [22] proposed an adaptive coding unit (CU) size CNN network to predict CTU partition. In addition, Li and Yang [23] proposed a depth edge classification CNN framework, utilizing post-processing to improve the classification result.

The PU mode selection can be roughly classified into edge-based methods and non-edge-based methods. For edge-based methods, a low complexity intra mode decision method was proposed by Hamout and Elyousfi [24], which skipped unnecessary modes by detecting the flat area and edge direction. In addition, Hamout and Elyousfi [25] proposed a depth learning network based on edge detection to reduce the computational complexity caused by the traversal selection of prediction modes. For non-edge-based methods, Wang et al. [26] utilized the automatic merging possibility clustering method to construct an intra prediction model, which was used to predict intra prediction mode. Furthermore, Zou et al. [27] implemented tensor feature extraction and data analysis to skip conventional depth modelling modes (DMMs). Lin et al. [28] proposed a depth intra mode decision method with multiple strategies, including rough mode decision termination, candidate mode reduction and fast DMM decision.

For a combination of CTU partition decision and PU mode selection, Chen et al. [29] presented early mode selection and adaptive CTU pruning termination based on the Bayesian decision rule. Song et al. [30] utilized a human visual system to reduce PU mode and additionally combined the boundary continuity and RD cost to accelerate the CTU partition decision. In addition, to accelerate the selection of CTU depth and PU mode together, Shen et al. [31] proposed to reduce the intra coding time based on boundary continuity. In this method, the thresholds of the total sum of squares (TSS) of PU boundaries and the RD cost were designed for CTU early termination.

Although the methods mentioned above have reduced the 3D-HEVC intra coding time, yet have the potential for improvement. The shortcomings of the above-mentioned studies mainly include: 1) only using CTU partition decision or PU mode selection has limitations in reducing the 3D-HEVC intra coding complexity; 2) although there are currently 3D-HEVC intra complexity reduction methods that include CTU partition decision and PU mode selection, these methods are usually divided into two steps and cannot be implemented at one time. Therefore, more research is required to determine CTU partition decision and PU mode selection simultaneously.

In this paper, we propose an MLFF model-guided low-complexity 3D-HEVC intra coding method. We develop a deep learning model that can fully mine the internal features of 3D video, enabling low-complexity 3D-HEVC intra coding with a small RD performance loss. Compared with traditional methods, the proposed method overcomes the problem of relying on manual feature extraction. Compared with the methods based on machine learning techniques, the proposed method addresses the problem of being unable to mine generalized video features. Compared with the methods based on deep learning, the proposed method makes full use of the internal characteristics in 3D video. Most importantly, the proposed method, which introduces two external features composed of QP and texture complexity into the model, has never been investigated before.

III. 3D-HEVC INTRA CODING COMPLEXITY ANALYSIS

To determine the reason for the high computational complexity of 3D-HEVC intra coding, we analyzed the two important processes of 3D-HEVC intra coding, CTU partition decision and PU mode selection.

To obtain the optimal partition result for the CTU partition decision, for a CTU of size 64 × 64, it will be split into four sub-CUs of size 32 × 32. For each 32 × 32 CU, it will be split into four sub-CUs of size 16 × 16. For each 16 × 16 CU, it will be split or not. Based on this, it can be calculated that for size 16 × 16 CU, there exists 2 partition structures. For size 32 × 32 CU, there exists 17 partition structures. For the CTU of size 64 × 64, there exists 83,522 partition structures. Table 1 shows the probability distribution of CTU partition depth. Table 1 indicates that the CTU partition depth changes with QP. The video sequences coded with small QP usually select CTU with large depth. On the contrary, the video sequences coded with large QP usually select CTU with small depth. Therefore, it is not necessary to code CTU in full traversal in all cases. If the partition depth of the current CTU can be predicted in advance, some unnecessary CTU partition processes can be skipped, reducing 3D-HEVC intra coding complexity.

For PU mode selection, it mainly includes two steps. The first step is rough mode decision (RMD), which is used to select N modes from 35 intra prediction modes and add them to the candidate mode list (RD-List). The second step is rate-distortion optimization (RDO), which is used to calculate the RD cost of all modes in the RD-List. If the partition depth of the current CTU does not reach the maximum partition depth, the current CTU is partitioned into 4 sub-CUs, and then the PU mode selection process is performed again until the size of the sub-CUs reaches 8 × 8. Table 2 illustrates the calculation times of RMD and RDO. It should be noted that in the process of RDO for the depth video, it is necessary to traverse DMMs mode, which is a unique intra coding mode for the depth video and has higher computational complexity [15]. Table 3 depicts the probability distribution of the PU mode. It can be observed that the PU mode being selected varies significantly. Among them, the Planar and DC modes, which are mainly used to encode the smooth texture, have the highest probability to be selected. However, the Angular

TABLE 1. Distribution probability of CTU partition depth.

Sequence	QP	Depth 0	Depth 1	Depth 2
Balloons (1024×768)	(25,34)	28.13%	1.04%	**70.84%**
	(40,45)	**55.21%**	13.54%	31.25%
Poznan_Street (1920×1088)	(25,34)	10.78%	3.53%	**85.69%**
	(40,45)	**78.43%**	8.63%	12.94%
Average	-	**43.14%**	6.69%	**50.18%**

TABLE 2. Calculation times of RMD and RDO in PU mode selection.

PU Size	Number of Sub PU	RMD times	RDO times	
			Texture	Depth
64×64	1	35	3	5
32×32	4	140	12	20
16×16	16	560	48	80
8×8	64	2,240	512	640
4×4	256	8,960	2,048	2,560
Total	341	11,935	2,623	3,305

TABLE 3. Distribution probability of PU mode.

Sequence	Planar/DC	Angular	DMMs
Balloons (1024×768)	77.30%	13.72%	8.98%
Poznan_Street (1920×1088)	91.75%	5.46%	2.79%
Average	**84.53%**	**9.59%**	**5.88%**

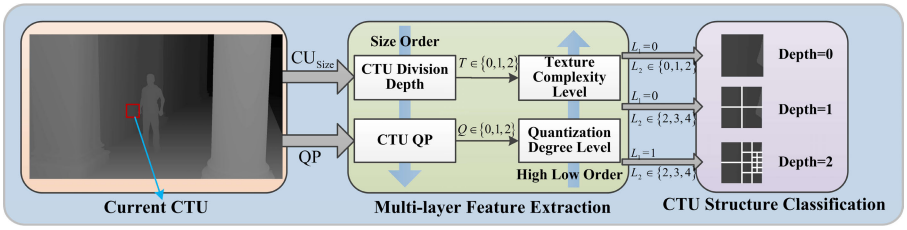

FIGURE 1. The process of CTU complexity level classification.

and DMMs mode, which are mainly used for coding the sharp edge, have a small probability to be selected. Therefore, if the PU mode can be predicted in advance, some unnecessary PU mode calculation processes can be skipped, thus reducing 3D-HEVC intra coding complexity.

Based on the above analysis, it is clear that the process of CTU partition decision and PU mode selection is filled with uncertainty. If the CTU partition depth and PU mode can be directly predicted, the full traversal process of CTU partition and PU mode selection can be avoided, reducing the complexity of 3D-HEVC intra coding.

IV. PROPOSED MLFF-GUIDED LOW-COMPLEXITY 3D-HEVC INTRA CODING METHOD

A. CLASSIFICATION OF CTU COMPLEXITY LEVEL

The aforementioned analysis of 3D-HEVC intra coding complexity indicates that QP and texture are important for CTU partition decision and PU mode selection. Motivated by this finding, we attempt to determine the complexity level of CTU by utilizing QP and texture. Fig. 1 shows the process of CTU complexity level classification. The specific process is as follows.

Firstly, we divide the texture complexity of CTU into three categories based on the size of encoded CTUs, and record it as texture complexity parameter T.

$$T_n^k = \begin{cases} 0, & \text{if } \mathrm{CU_{size}} = 64 \times 64 \\ 1, & \text{if } \mathrm{CU_{size}} = 32 \times 32 \\ 2, & \text{if } \mathrm{CU_{size}} = 16 \times 16 \text{ or } 8 \times 8 \end{cases} \tag{1}$$

where the subscript $n \in \{0, 1, 2, 3\}$ is the index of CTU partition depth and the superscript $0 \le k \le 63$ is the index of sub-CUs split from the CTU. If the size of encoded CU is 64×64, the current CU almost has no texture details and the texture complexity is the lowest. Set T to 0 at this time. If the size of encoded CU is 32, the current CU has fewer texture details and the texture complexity is medium. Set T to 1 at this time. If the size of encoded CU is 16×16 or 8×8, the current CU has the most texture details and the texture complexity is the highest. Set T to 2 at this time.

Secondly, considering that the quantization step (Q_{step}) of CU has a direct impact on the quality of encoded video, the relationship between QP and Q_{step} is defined as follows:

$$Q_{\text{step}} \approx 2^{(\mathrm{QP}-4)/6} \tag{2}$$

We divide the quantization degree of CU into three categories according to the QP range of 0-51, and record it as quantization degree parameter Q.

$$Q_n^k = \begin{cases} 0, & \text{if QP} = 45 \sim 51 \\ 1, & \text{if QP} = 35 \sim 44 \\ 2, & \text{if QP} = 0 \sim 34 \end{cases} \tag{3}$$

where the subscript $n \in \{0, 1, 2, 3\}$ is the index of CTU partition depth and the superscript $0 \le k \le 63$ is the index of sub-CUs split from the CTU. If the QP range is 45-51, the current CU belongs to rough quantization with the lowest quantization degree. Set Q to 0 at this time. If the QP range is 35-44, the current CU belongs to general quantization, and the quantization degree is medium. Set Q to 1 at this time. If the QP range is 0-34, the current CU belongs to fine quantization with the highest quantization degree. Set Q to 2 at this time.

Finally, we define two complexity control parameters, $L_1 = T \&\& Q$ and $L_2 = T + Q$, to classify the CTU complexity level. It is worth noting that, in the proposed MLFF model, L_1 and L_2 are added as external features and participate in training together.

B. PROPOSED MLFF MODEL

In this section, an MLFF model is developed to predict the CTU depth and PU mode in advance. Here, we integrate the texture and QP of CTU into multi-layer features. The MLFF model is designed as shown in Fig. 2 according to the complexity control parameters that represent the CTU complexity level in Fig. 1.

In general, the proposed MLFF model includes three channels. From top to bottom, each channel corresponds to each layer of the CTU partition structure. The input of the model is the CTU of size 64×64. The output of the model corresponds to the probability prediction results of Depth=0, Depth=1, and Depth=2, respectively. If the prediction probability for a certain depth is greater than 0.5, the prediction process for the next layer is terminated and the current depth is used as the optimal partition depth. The structure of the proposed MLFF model is described in detail below. Specifically, the proposed MLFF model consists of three preprocessing modules, nine convolution layers, four concatenate layers and nine fully connected layers.

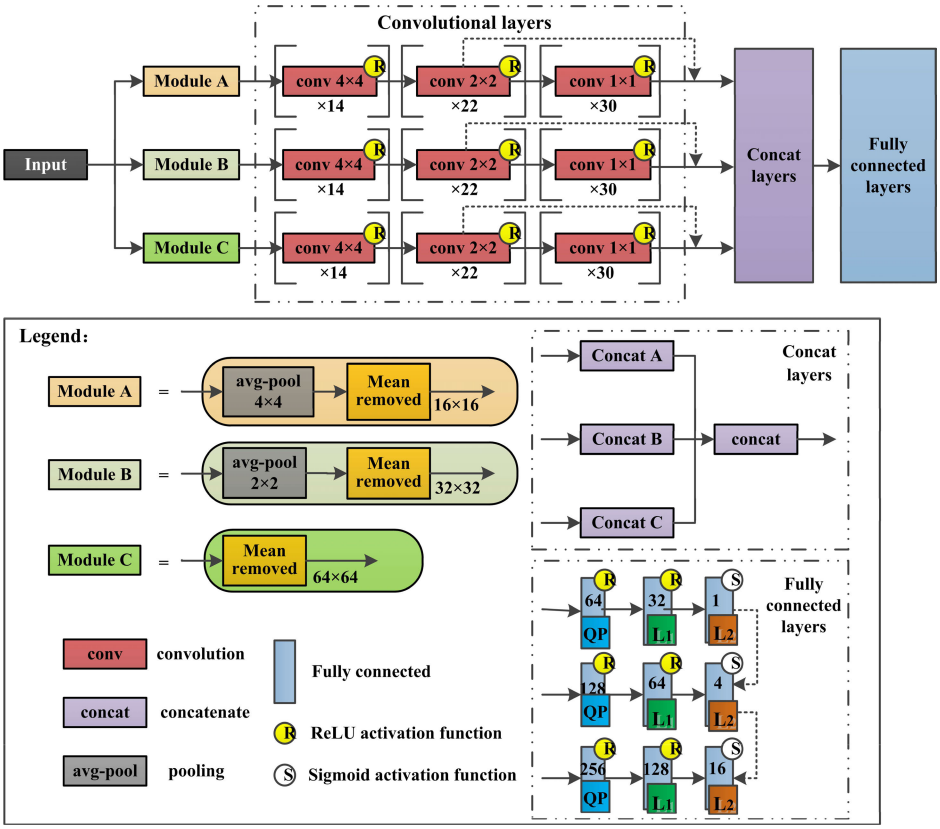

FIGURE 2. The proposed MLFF model.

First, the input CTU must be preprocessed in order to make the model output in the form of a CTU partition structure. In Fig. 2, Module A, Module B, and Module C are all preprocessing modules. In these three preprocessing modules, we adopt average pooling and mean removed to reduce the feature dimensions and interference information.

Second, as compared to advanced features such as image semantics, texture is considered a shallow feature that CNN can obtain through a simple network structure. Therefore, the convolutional operation after the preprocessing module is the core part of the model. The convolution core size of the first group of 14 convolution layers is 4×4. The dimensions of the feature maps are 13×13, 29×29, and 61×61, respectively. The convolution core size of the second group of 22 convolution layers is 2×2. The dimensions of the feature maps are 12×12, 28×28, and 60×60, respectively. The convolution core size of the third group of 30 convolution layers is 1×1. The dimensions of the feature maps are 12×12, 28×28, and 60×60, respectively. Considering that the proposed model is essentially a classification problem, the activation function utilized in the convolution layer is the Rectified Linear Unit (ReLU). The specific calculation process can be expressed as follows:

$$C_m\left(\mathrm{CTU}_n\right) = \begin{cases} \mathrm{CTU}_n, & m = 0 \\ \mathrm{ReLU}\left(W_m C_{m-1}\left(\mathrm{CTU}_n\right) + B_m\right), & m \le M \end{cases} \tag{4}$$

where C_m represents the convolutional layer after the preprocessing module, M represents the total number of convolutional layers, and m represents the current processing layer. W_m is the weight matrix and B_m denotes the offset. In addition, n denotes the CTU that is currently being processed. If the processed video resolution is 1920×1088, and there are 510 CTUs in a frame, the range of n is $1 \le n \le 510$. If the processed video resolution is 1024×768, and there are 192 CTUs in a frame, the range of n is $1 \le n \le 192$. It is worth noting that the higher the video resolution, the more CTUs are included in a frame, but also the larger the inaccuracy of model prediction.

Third, following the convolutional layers, the extracted features from the second and third groups of convolutional layers are passed to the concatenate layer. The fully connected layers that learn the correlation of features from different channels and convolutional layers. Moreover, we introduce two external features, L_1 and L_2, to the fully connected layers. At the same time, in order to avoid overfitting, 50% of the first fully connected layer features will be randomly dropped out.

Finally, because the binary cross entropy loss function is frequently used in binary classification, and since the model proposed in this research is essentially a binary classification problem, we use binary cross entropy as the loss function of the MLFF model. The specific calculation of loss

function L is defined as follows:

$$L = -\frac{1}{N}\sum_{n=1}^{N}(l_n \times \ln Y_n + (1 - l_n) \times \ln(1 - Y_n)) \quad (5)$$

where l_n represents whether the currently encoded CTU is partitioned, Y_n is the output of the model, N denotes the total number of CTUs to be processed, and n represents the index of the CTUs currently being processed.

C. FLOWCHART OF THE PROPOSED METHOD

Fig. 3 gives the flowchart of the proposed MLFF model-guided low-complexity 3D-HEVC intra coding method.

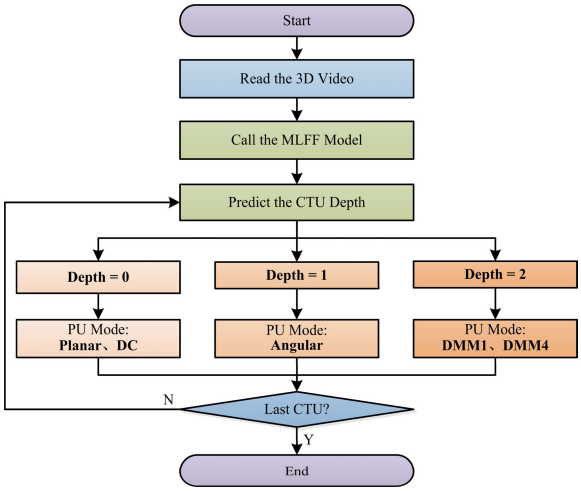

FIGURE 3. Flowchart of the proposed method.

From Fig. 3, firstly, we read the 3D video to be encoded and call the proposed MLFF model. Secondly, we read the CTU to be encoded and use the model to predict CTU partition depth. Finally, we determine the selection range of PU mode based on CTU depth.

V. EXPERIMENTAL RESULTS

A. SETTINGS

1) TRAINING ENVIRONMENT

Table 4 presents the hardware and software environments. It is worth noting that the GPU is only utilized to accelerate model training, not the coding process. Model training conditions are used as follows: set the batch size to 64, set the number of iterations to 10,000, and the initial learning rate to 0.01. The learning rate drops at a rate of 0.1 for every 4,000 iterations as learning times increase.

2) CODING CONFIGURATION

The test experiment is conducted in Common Test Conditions (CTC). We use All Intra (AI) mode to test the performance of 3D-HEVC intra coding. The test platform is HTM-16.0, the compiling software is Visual Studio 2010, and the configuration file is *baseCfg_3view+ depth_AllIntra*. The standard test sequence and specific coding configuration parameters used in this paper are shown in Table 5. It is worth mentioning that the dataset of the MLFF model was collected from standard test sequences *Kendo* and *Undo_Dancer*, hence these two video sequences are not included in the test sequence.

TABLE 4. Training environment.

Hardware	
Processor	13th Gen Intel(R) Core(TM) i9-13900K 3.00 GHz
Random Access Memory (RAM)	32.0 GB
Graphic Processing Unit (GPU)	NVIDIA GeForce RTX 3080
Software	
Python	3.5
Tensorflow	1.4.0
CUDA	8.0

TABLE 5. Standard test sequence coding configuration.

Standard test sequence			
Sequence	Resolution	Frame rate	View
Balloons	1024×768	30	3 1 5
Newspaper	1024×768	30	4 2 6
Poznan_Hall2	1920×1088	25	6 7 5
Poznan_Street	1920×1088	25	4 5 3

3) EVALUATION CRITERIA

The coding complexity of the proposed method is measured as follows:

$$\Delta T_{\text{total}} = \frac{T_{\text{HTM-16.0,total}} - T_{\text{proposed,total}}}{T_{\text{HTM-16.0,total}}} \times 100\,(\%) \quad (6)$$

where $T_{\text{HTM-16.0,total}}$ represents the coding time of HTM-16.0 and $T_{\text{proposed,total}}$ represents the coding time of the proposed method. Fig. 4 depicts the computation process for calculating the PSNR of the synthesized view. PSNR of the synthesized view is the average PSNR of the six virtual views. The specific calculation formula can be expressed as follows:

$$\text{PSNR}_{\text{synth}} = \frac{\sum \text{PSNR}_n}{6},\ n = 0.25, 0.5, 0.75, 1.25, 1.5, 1.75 \quad (7)$$

where $\text{PSNR}_{\text{synth}}$ represents PSNR of the synthesized view, and PSNR_n represents PSNR of the six virtual views.

B. TRAINING PERFORMANCE EVALUATION

Fig. 5 (a) and Fig. 5 (b) show the loss and accuracy curves in predicting CTU partition depth, respectively. The loss is calculated by (5), and the accuracy is computed by comparing the prediction results with the ground truth. It can be seen that as the number of iterations increases, the prediction loss gradually decreases, and the prediction accuracy gradually increases.

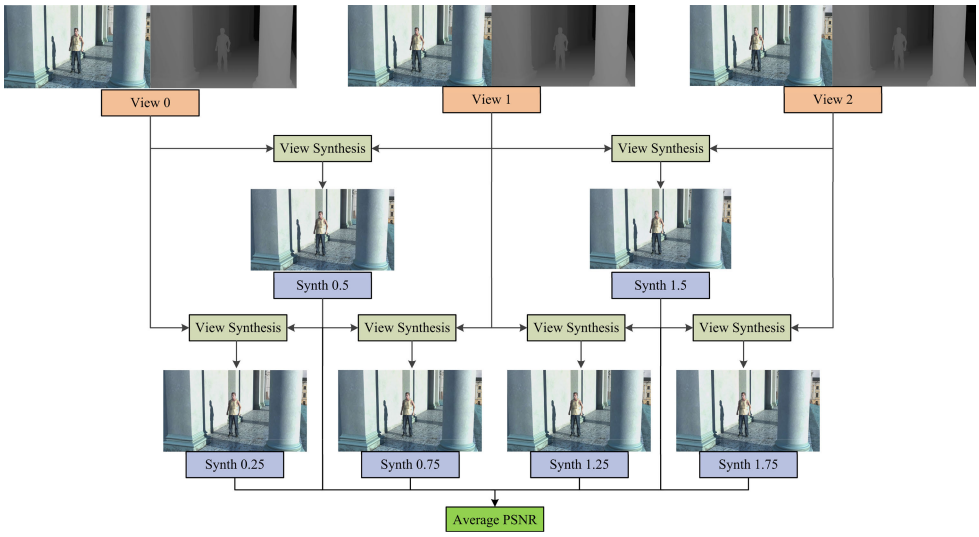

FIGURE 4. PSNR calculation process of the synthesized view.

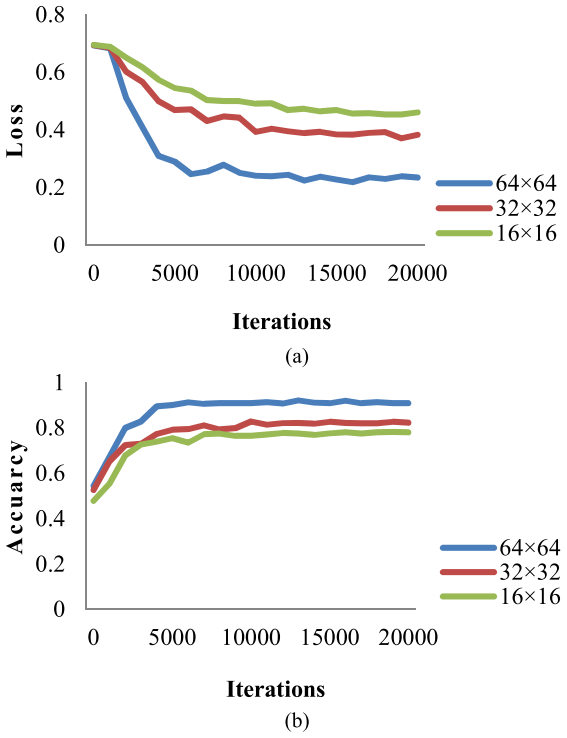

FIGURE 5. PSNR calculation process of the synthesized view.

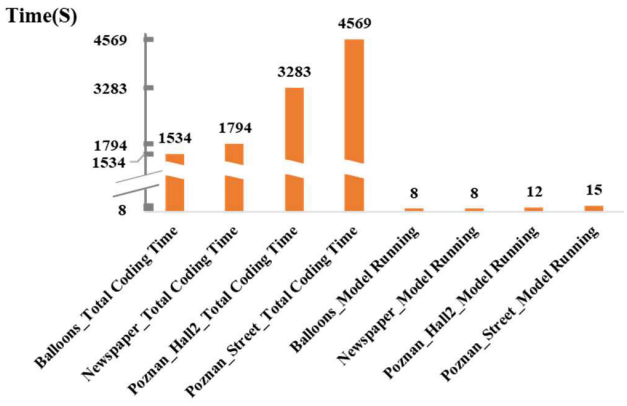

FIGURE 6. PSNR calculation process of the synthesized view.

In addition, we also recorded the model running time and the total coding time of the proposed method. As shown in Fig. 6, the running time of the model accounts for just about 0.4% of the entire coding time. As a result, introducing the MLFF model will not add too much coding time.

C. OBJECTIVE PERFORMANCE EVALUATION

The test experiment employs four video sequences that are not included in the dataset. As demonstrated in Table 6, the proposed method can save 37.4% coding time when compared to HTM-16.0. More coding time will be saved in particular for video sequences such as *Newspaper* that have a mostly unchanged background and a slowly changing foreground. For instance, the time saving for *Newspaper* can reach to 45.2%.

Comparative experiments were carried out to evaluate the proposed method's performance in more detail. Having comparison to the test sequences at resolution 1024 × 768, Table 7 indicates that the coding time is 9.4%, 3.7% and 4.8% less than that of reference [26, 27, and 30]. Furthermore, in test sequences with all other resolution cases, our method performs better in 36.6% [24], 33.7% [25], and 32.6% [28].

Regarding coding performance, the synthesized view's quality has more significance than that of the texture and depth videos. Hence, we examine the BDBR of the synthesized view using several approaches. The term "Synth PSNR/total bitrate" refers to the BDBR that is

FIGURE 7. Comparison of the subjective quality for the synth 0.25 of "*Balloons*", "*Newspaper*", "*Poznan_Street*", and "*Poznan_Hall2*".

TABLE 6. Coding time comparison of the proposed method to original HTM-16.0.

Sequences	Coding Time (s)		ΔT_t (%)
	HTM-16.0	Proposed	
Balloons	2293.7540	1534.1218	**33.1**
Newspaper	3274.6653	1793.6025	**45.2**
Poznan_Hall2	5184.2390	3283.4488	**36.6**
Poznan_Street	6976.4223	4569.0478	**34.5**
Average	4432.2700	2795.0550	**37.4**

determined by comparing the total bitrate to the synthesized view PSNR. Table 7 demonstrates that the suggested method's synthesized view BDBR loss is 0.31% at all resolutions and just 0.17% at 1024 × 768 resolution, which is far better than most of the comparison methods.

D. SUBJECTIVE PERFORMANCE EVALUATION

Fig. 7 illustrates the subjective quality of the synthesized view of the proposed and the original HTM-16.0 methods. Fig. 7 indicates that, despite a slight decrease in objective data, there is no decline in subjective. This further proves that the proposed method can reduce 3D-HEVC intra coding

TABLE 7. Coding time comparison of the proposed method to others under HTM-16.0.

Sequences	1024×768 resolution							
	[26]		[27]		[30]		Proposed	
	ΔT_2 (%)	BDBR(%)	ΔT_3 (%)	BDBR (%)	ΔT_4 (%)	BDBR (%)	ΔT_1(%)	BDBR (%)
Balloons	28.1	0.20	32.3	0.37	34.3	0.12	33.1	0.14
Newspaper	31.4	0.13	38.7	0.17	34.4	0.35	45.3	0.20
Average	**29.8**	**0.17**	**35.5**	**0.27**	**34.4**	**0.24**	**39.2**	**0.17**
	All resolutions							
Sequences	**[24]**		**[25]**		**[28]**		**Proposed**	
	ΔT_5 (%)	BDBR (%)	ΔT_6 (%)	BDBR (%)	ΔT_7 (%)	BDBR (%)	ΔT_1(%)	BDBR (%)
Balloons	37.9	0.40	31.9	0.61	33.0	0.62	33.1	0.14
Newspaper	37.0	0.40	30.2	0.37	33.8	0.65	45.3	0.20
Poznan_Hall2	35.7	1.13	36.0	0.82	32.1	0.47	36.7	0.52
Poznan_Street	35.7	0.19	36.7	0.57	31.4	0.29	34.7	0.38
Average	**36.6**	**0.53**	**33.7**	**0.59**	**32.6**	**0.51**	**37.4**	**0.31**

complexity without compromising the quality of synthesized views.

VI. CONCLUSION

Existing 3D-HEVC intra coding approaches use a variety of complicated coding techniques to provide extra intermediate views for better representations of 3D videos, posing significant challenges for real-time 3D video applications. To address this issue, we propose in this research, a low-complexity 3D-HEVC intra coding method. The complexity of 3D-HEVC intra coding was examined first in the proposed method. Then, instead of performing a full traversal search, an MLFF model was built that learns to predict the best CTU partition depth and PU mode at the same time. Then, two external features composed of QP and texture were introduced into the model, to improve its prediction accuracy. The experimental findings confirmed that the proposed method is better than existing methods in terms of complexity reduction and far more effective in terms of RD performance.

In future work, we will continue to improve our proposed model. Considering the characteristic differences between the texture video and the depth video, we intend to extract features from the texture video and the depth video independently to implement feature fusion.

REFERENCES

[1] J. Artois, G. van Wallendael, and P. Lambert, "360DIV: 360° video plus depth for fully immersive VR experiences," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Las Vegas, NV, USA, Jan. 2023, pp. 1–2.

[2] G. Tech, Y. Chen, K. Müller, J.-R. Ohm, A. Vetro, and Y.-K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 35–49, Jan. 2016, doi: 10.1109/TCSVT.2015.2477935.

[3] 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 Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 6347–6359, Sep. 2022, doi: 10.1109/TCSVT.2022.3161103.

[4] P. Gao, W. Xiang, and D. Liang, "Texture-distortion-constrained joint source-channel coding of multi-view video plus depth-based 3D video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 11, pp. 3326–3340, Nov. 2019, doi: 10.1109/TCSVT.2018.2877903.

[5] G. Wang, Z. Wang, K. Gu, K. Jiang, and Z. He, "Reference-free DIBR-synthesized video quality metric in spatial and temporal domains," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1119–1132, Mar. 2022, doi: 10.1109/TCSVT.2021.3074181.

[6] G. Sanchez, J. Silveira, L. V. Agostini, and C. Marcon, "Performance analysis of depth intra-coding in 3D-HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2509–2520, Aug. 2019, doi: 10.1109/TCSVT.2018.2865645.

[7] Y. Zhang, G. Wang, R. Tian, M. Xu, and C. C. J. Kuo, "Texture-classification accelerated CNN scheme for fast intra CU partition in HEVC," in *Proc. IEEE Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2019, pp. 241–249.

[8] Z. Chen, J. Shi, and W. Li, "Learned fast HEVC intra coding," *IEEE Trans. Image Process.*, vol. 29, pp. 5431–5446, 2020, doi: 10.1109/TIP.2020.2982832.

[9] 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, doi: 10.1109/TCSVT.2012.2221191.

[10] J. Huo, X. Zhou, H. Yuan, S. Wan, and F. Yang, "Fast rate-distortion optimization for depth maps in 3-D video coding," *IEEE Trans. Broadcast.*, vol. 69, no. 1, pp. 21–32, Mar. 2023, doi: 10.1109/TBC.2022.3192992.

[11] J. R. Atencia, O. L. Granado, M. P. Malumbres, M. O. Martínez-Rach, and G. van Wallendael, "Analysis of the perceptual quality performance of different HEVC coding tools," *IEEE Access*, vol. 9, pp. 37510–37522, 2021, doi: 10.1109/ACCESS.2021.3062938.

[12] L. Li, Y. Huang, J. Wu, K. Gu, and Y. Fang, "Predicting the quality of view synthesis with color-depth image fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 7, pp. 2509–2521, Jul. 2021, doi: 10.1109/TCSVT.2020.3024882.

[13] M. Zhang, X. Zhai, Z. Liu, and C. An, "Fast algorithm for HEVC intra prediction based on adaptive mode decision and early termination of CU partition," in *Proc. Data Compress. Conf.*, Snowbird, UT, USA, Mar. 2018, p. 434.

[14] V. Borges, M. Perleberg, V. Afonso, M. Porto, and L. Agostini, "A hardware design for 3D-HEVC depth intra skip with synthesized view distortion change," in *Proc. 33rd Symp. Integr. Circuits Syst. Design (SBCCI)*, Campinas, Brazil, Aug. 2020, pp. 1–6.

[15] P. Liu, G. He, S. Xue, and Y. Li, "A fast mode selection for depth modelling modes of intra depth coding in 3D-HEVC," in *Proc. Vis. Commun. Image Process. (VCIP)*, Chengdu, China, Nov. 2016, pp. 1–4.

[16] Q. Zhang, Y. Wang, T. Wei, L. Huang, and R. Su, "A fast and efficient 3D-HEVC method for complexity reduction based on the correlations of inter-view, spatio-temporal, and texture-depth," *IEEE Access*, vol. 8, pp. 129075–129086, 2020, doi: 10.1109/ACCESS.2020.3009424.

[17] C.-H. Fu, Y.-L. Chan, H.-B. Zhang, S. H. Tsang, and M.-T. Xu, "Efficient depth intra frame coding in 3D-HEVC by corner points," *IEEE Trans. Image Process.*, vol. 30, pp. 1608–1622, 2021, doi: 10.1109/TIP.2020.3046866.
[18] Q. Zhang, Y. Wang, L. Huang, T. Wei, and R. Su, "Fast coding scheme for low complexity 3D-HEVC based on video content property," *Multimedia Tools Appl.*, vol. 80, no. 7, pp. 25909–25925, Apr. 2021, doi: 10.1007/s11042-021-10961-6.
[19] M. Saldanha, G. Sanchez, C. Marcon, and L. Agostini, "Fast 3D-HEVC depth map encoding using machine learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 850–861, Mar. 2020, doi: 10.1109/TCSVT.2019.2898122.
[20] H. Zhang, W. Yao, H. Huang, Y. Wu, and G. Dai, "Adaptive coding unit size convolutional neural network for fast 3D-HEVC depth map intracoding," *J. Electron. Imag.*, vol. 30, no. 4, Jun. 2021, Art. no. 041405, doi: 10.1117/1.jei.30.4.041405.
[21] C. Liu, K. Jia, and P. Liu, "Fast depth intra coding based on depth edge classification network in 3D-HEVC," *IEEE Trans. Broadcast.*, vol. 68, no. 1, pp. 97–109, Mar. 2022, doi: 10.1109/TBC.2021.3106143.
[22] R. Zhang, K. Jia, P. Liu, and Z. Sun, "Fast intra-mode decision for depth map coding in 3D-HEVC," *J. Real-Time Image Process.*, vol. 17, no. 5, pp. 1637–1646, Oct. 2020, doi: 10.1007/s11554-019-00920-8.
[23] Y. Li and J. Yang, "Fast intra coding algorithm for depth map in 3D-HEVC," *J. Optoelectron. Laser*, vol. 31, no. 2, pp. 222–228, Feb. 2020, doi: 10.16136/j.joel.2020.02.0344.
[24] H. Hamout and A. Elyousfi, "Fast depth map intra-mode selection for 3D-HEVC intra-coding," *Signal, Image Video Process.*, vol. 14, no. 7, pp. 1301–1308, Oct. 2020, doi: 10.1007/s11760-020-01669-5.
[25] H. Hamout and A. Elyousfi, "Fast depth map intra coding for 3D video compression-based tensor feature extraction and data analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 7, pp. 1933–1945, Jul. 2020, doi: 10.1109/TCSVT.2019.2918770.
[26] C. Wang, G. Feng, C. Cai, X. Han, and H. Cao, "Multi-strategy depth intra mode decision algorithm in 3D-HEVC," *Multimedia Tools Appl.*, vol. 79, nos. 13–14, pp. 8841–8861, Apr. 2020, doi: 10.1007/s11042-019-7715-0.
[27] D. Zou, P. Dai, and Q. Zhang, "Fast depth map coding based on Bayesian decision theorem for 3D-HEVC," *IEEE Access*, vol. 10, pp. 51120–51127, 2022, doi: 10.1109/ACCESS.2022.3174119.
[28] J.-R. Lin, M.-J. Chen, C.-H. Yeh, Y.-C. Chen, L.-J. Kau, C.-Y. Chang, and M.-H. Lin, "Visual perception based algorithm for fast depth intra coding of 3D-HEVC," *IEEE Trans. Multimedia*, vol. 24, pp. 1707–1720, 2022, doi: 10.1109/TMM.2021.3070106.
[29] M.-J. Chen, J.-R. Lin, Y.-C. Hsu, Y.-S. Ciou, C.-H. Yeh, M.-H. Lin, L.-J. Kau, and C.-Y. Chang, "Fast 3D-HEVC depth intra coding based on boundary continuity," *IEEE Access*, vol. 9, pp. 79588–79599, 2021, doi: 10.1109/ACCESS.2021.3083498.
[30] W. Song, P. Dai, and Q. Zhang, "Content-adaptive mode decision for low complexity 3D-HEVC," *Multimedia Tools Appl.*, vol. 82, no. 17, pp. 26435–26450, Mar. 2023, doi: 10.1007/s11042-023-14874-4.
[31] L. Shen, K. Li, G. Feng, P. An, and Z. Liu, "Efficient intra mode selection for depth-map coding utilizing spatiotemporal, inter-component and inter-view correlations in 3D-HEVC," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4195–4206, Sep. 2018, doi: 10.1109/TIP.2018.2837379.

CHANG LIU received the Ph.D. degree in electronic science and technology from Beijing University of Technology, Beijing, China.

She is currently an Associate Professor with the Research Center for Intelligent Information Technology, Nantong University. Her research interests include two dimensional video coding, three dimensional video coding, and deep learning.

KEBIN JIA (Member, IEEE) received the M.S. and Ph.D. degrees in information and communication engineering from the University of Science and Technology of China, Anhui, China, in 1990 and 1998, respectively.

He is currently a Professor and the Director of the First-Class Disciplines Construction Office, Beijing University of Technology, Beijing, China. He is also a Full Professor with the Faculty of Information Technology, Beijing University of Technology, and the Director of the Institute of Multimedia Information Processing and Imaging Technology. He was a PI for more than 15 research projects from the National Natural Science Foundation of China (NSFC), the 973 National Basic Research Program, and the 863 Program. His research interests include multimedia and database systems, content-based image/video retrieval, image/video coding and processing, data mining, and pattern recognition. He is a Senior Member of the Chinese Institute of Electronics. His group has received the Award of Outstand and Innovator Group Award of the Committee of Education of Beijing.