Fast Depth Video Coding with Intra Prediction on VVC

Hongan Wei[1], Binqian Zhou[1], Ying Fang[1], Yiwen Xu[1*] and Tiesong Zhao[1]
[1] Fujian Key Lab for Intelligent Processing and Wireless Transmission of Media Information, College of Physics and Information Engineering, Fuzhou University
Fuzhou, China
[e-mail: {weihongan, N171127047, fangying, xu_yiwen, t.zhao}@fzu.edu.cn]
*Corresponding author: Yiwen Xu

Received January 22, 2020; revised May 12, 2020; accepted June 8, 2020; published July 31, 2020

Abstract

In the stereoscopic or multiview display, the depth video illustrates visual distances between objects and camera. To promote the computational efficiency of depth video encoder, we exploit the intra prediction of depth videos under Versatile Video Coding (VVC) and observe a diverse distribution of intra prediction modes with different coding unit sizes. We propose a hybrid scheme to further boost fast depth video coding. In the first stage, we adaptively predict the HADamard (HAD) costs of intra prediction modes and initialize a candidate list according to the HAD costs. Then, the candidate list is further improved by considering the probability distribution of candidate modes with different CU sizes. Finally, early termination of CU splitting is performed at each CU depth level based on the Bayesian theorem. Our proposed method is incorporated into VVC intra prediction for fast coding of depth videos. Experiments with 7 standard sequences and 4 Quantization parameters (Qps) validate the efficiency of our method.

Keywords: Depth video, intra prediction, Versatile Video Coding (VVC), fast coding algorithm, mode decision, early termination

This work is supported by National Natural Science Foundation of China (No.61671152)

http://doi.org/10.3837/tiis.2020.07.016 **ISSN : 1976-7277**

1. Introduction

Videos have become the indispensable part of human daily life. Consumers are no longer satisfied with the visual enjoyment provided by traditional flat videos, while expecting to enjoy the new visual experience from stereo perception. As a result, research on 3D vision has gradually become a hot topic. Depth video is an essential part of 3D scene representation. It provides distance information between photographed objects and the camera, thus stereoscopic perception is introduced to viewers. Actually, depth video is a kind of grayscale sequence, where the brighter area means the closer distance between object and camera. It has some different characteristics from traditional color video. Traditional color video focuses on detail and color, while depth video highlights geometric information about the relative distance of the objects to the cameras. Since the Human Visual System (HVS) is less sensitive to depth video than color video, depth video allows for more distortion and can be encoded and transmitted with less information. These characteristics are usually employed to achieve efficient fast coding in depth video. Due to the particularity of depth video, the design of video coding schemes is also somewhat different from traditional color video. For the purpose of improving the coding efficiency of stereoscopic video, a reasonable design scheme for efficient coding of depth video must be proposed based on the depth characteristics.

Consumers are not only pursuing high-impact and immersive visual experiences, but they also place a higher requirement on the image quality of videos. Due to higher resolution and higher quality videos, such as 360-degree, ultra-high-definition, high dynamic range and virtual reality, the demand for the next generation video compression has become urgent. To better address this issue, the Joint Video Exploration Team (JVET) has been developing the next generation video coding standard since 2015, officially named as the Versatile Video Coding (VVC) in 2018, and released the VVC Test Mode (VTM) [1].

A number of innovative techniques have been adopted in the VTM for improving the compression efficiency on the basis of HEVC. Among them, one of the most significant changes is the elimination of the concepts of Prediction Unit (PU) and Transform Unit (TU). Coding Unit (CU) is utilized directly for prediction and transform progressing without any further division. Another key innovation is that the partition method adopts the Quad-Tree with nested Multi-Type Tree (QTMTT) [2] coding block structure instead of the traditional QT structure. QTMTT includes QT, Ternary Tree (TT), and Binary Tree (BT) partition structure, which means that the shape of the segmentation blocks no longer has only a single square, but also a rectangle. This segmentation method can be more flexibly divided according to the texture of the picture, and better adapt to the feature of various regions. An example of QTMTT partition structure is illustrated in **Fig. 1**. In addition, to better capture any edge direction, VVC expands the number of Intra Prediction Modes (IPM) from 35 to 67, including 2 non-angle modes and 65 angle modes. Due to the increase of the number of prediction modes, the selection mechanism of the optimal intra mode has changed. Based on the Rough Mode Decision (RMD) mechanism in HEVC, the prediction of adjacent sub-modes of candidate modes has been added, thereby further improving the intra prediction accuracy.

In brief, the updated coding tools have fully demonstrated the VVC coding efficiency. What is more, they also have brought a huge computational burden to the emerging codecs, especially intra prediction. Therefore, to facilitate the development and application of VVC, it is essential to reduce the coding complexity. In this paper, based on the characteristics of depth video, we propose a hybrid fast depth video coding algorithm for reducing the complexity of

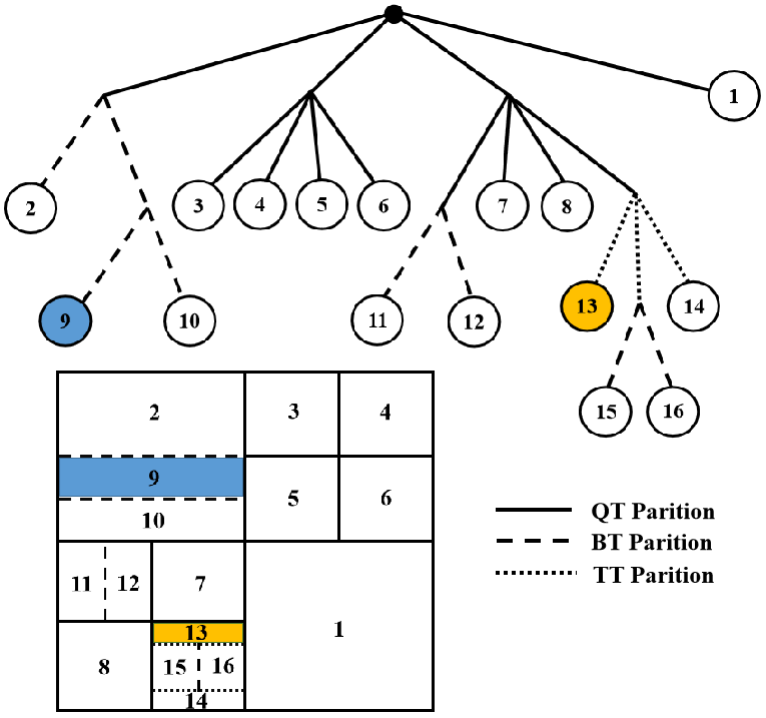


Fig. 1. An example of QTMTT partition structure

VVC intra prediction. Experimental results demonstrate that our method has strong robustness, and averagely gain 39.85% time saving with negligible quality loss.

The remains of this paper are organized as follows. Section 2 reviews previous work in depth video intra coding. Section 3 introduces the features of depth videos. Section 4 describes the details of the proposed method. Some experimental results and analyses are presented in Section 5. Finally, the conclusions are made in Section 6.

2. Related Work

Until now, enormous works have contributed to fast intra coding of depth videos under Advance Video Coding (AVC) [3] and High Efficiency Video Coding (HEVC) [4]. Among them, two types of methods are more popular: candidate mode elimination before intra prediction and early termination during intra prediction.

The former is mainly used to optimize the number of intra prediction modes. In [5, 6], the coding blocks are classified to speed up the mode decision process of the depth map. The difference is that Yoon et al. divided depth maps into the continuous areas and the discontinuous areas [5], while Sanchez et al. Classified Coding Unit (CU) blocks as the edges or nearly constant regions [6]. Da Silva et al. utilized the correlation between depth and texture map to limit the number of candidate modes [7]. In [8], the authors decreased the computational complexity by reducing the number of modes in the Rate Distortion (RD) list. Shen et al. allocated different intra candidate modes by judging the complexity of CUs, and skipped unnecessary CU sizes based on RD cost [9]. Depending on the specific characteristics of depth map, Merkle et al. designed intra prediction modes based on geometric primitives along with a residual coding method in the spatial domain, replacing the traditional intra

prediction modes and transform coding, respectively [10]. Zhang et al. [11] designed a low complexity intra mode selection algorithm by detecting the flat area and the texture direction of depth map. Based on edge detection, Wang et al. [12] proposed a fast intra modes decision algorithm for depth map by determining whether the Depth Modeling Modes (DMMs) need to be traversed. Hamout et al. [13] utilized fast depth map intra algorithms based on tensor feature extraction and data analysis to skip unnecessary conventional intra prediction modes and DMMs.

In early termination algorithms, time saving is obtained via early terminating some uncritical processes. For instance, Conceição et al. proposed an adaptive threshold model for 3D depth coding by utilizing early skip and early depth intra skip scheme [14]. Kim et al. skipped the sub CUs by judging RD cost of the intra skip mode and the variance value of the coding block [15]. Chen et al. proposed an early termination method for the block segmentation, which depended on the relationship between the current block and its first sub block [16]. In [17], the proposed algorithm consisted of a directional intra prediction frame and a simple residual coding method combined with an optimized flexible block partitioning scheme. In order to terminate CU partition early, Chung et al. proposed a fast depth map QT structure determination algorithm [18]. In [19], using the total square sum and RD cost criteria for complexity check, an early termination algorithm is proposed to speed up the intra encoding process. Based on the coding information from the spatial neighboring depth map treeblocks and co-located texture video treeblocks, Zhang et al. introduced an efficient early termination algorithm for 3D-HEVC depth map encoding [20]. In [21], the coding stage of Intra 2Nx2N, Intra NxN or CU partition for the CUs could be early skipped based on several decision trees.

From the above algorithms, various achievements in fast coding of depth video have been obtained. However, all the above schemes were implemented on the previous generation of video coding standard, such as HEVC and AVC. While a few works were done for the new coding standards VVC. In this paper, we present a hybrid fast depth video coding algorithm, which is subsequently employed in the intra prediction of VVC to ulteriorly improve coding performance. The simulation results show that our method can reduce the encoding time by 39.85% on average under the premise of ensuring a certain image quality.

3. The Features of Depth Video

3.1 Differences between depth video and color video

Depth videos and the corresponding traditional color videos are used to represent 3D videos. But depth videos have their unique characteristics, which is different from traditional color videos. The traditional color videos focus on the contour, texture, and chrominance information of objects in the scene, while depth videos mainly describe the geometric positional relationship in the scene. To better compare the differences in texture complexity between color and depth videos, we calculate their Temporal Information (TI) and Spatial Information (SI) [22]. Generally, complex scenes have high SI values, and sequences with vigorous motion have high TI values. The SI and TI values of color videos and their corresponding depth videos in several sequences are shown in **Fig. 2**, where the diamond represents the color sequence and the circle represents the depth sequence. It is clearly observed that the complexity of color videos is much higher than that of depth videos. Among them, the relatively distinct difference emerges in the *UndoDancer* sequence, whose color map and depth map are as shown in **Fig. 3**. It can be seen from the figure that there are more details in

the color map, such as the patterns of the floor, the grids of the window, and so on. However, the same areas in the corresponding depth map are flat and homogeneous. This is because depth videos only provide depth information but ignore the details of videos. When several objects are in the same plane or close in the distance, there may appear to be a large-scale homogeneous region. These features result in a stronger correlation between adjacent CUs in the depth video, which provides more room for video compression.

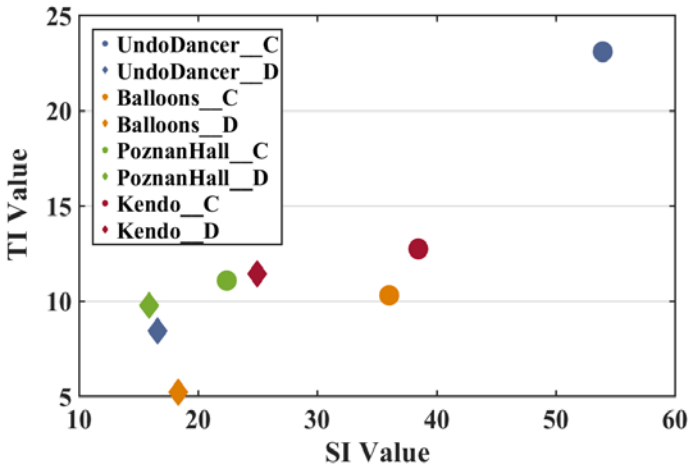
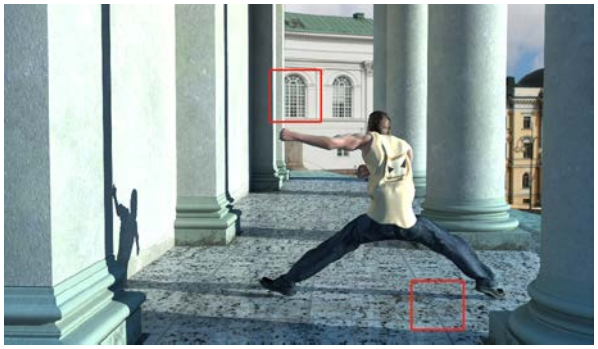
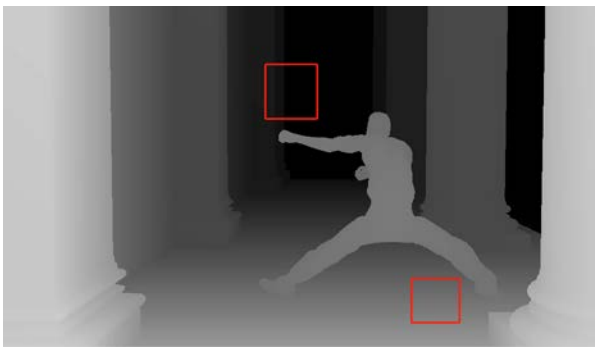


Fig. 2. The SI and TI values of depth and corresponding color sequences



(a) color map



(b) depth map

Fig. 3. The color map and depth map of the *UndoDancer* sequence

In addition, depth videos have penetrated into many popular areas such as 3D scene reconstruction. Commonly, they are not displayed to viewers [23]. Due to the different uses of color videos and depth videos, viewers have lower quality requirements on depth videos than color videos. Besides, human eyes are less sensitive to depth videos than color videos. Therefore, depth videos allow more distortion and can be encoded with less bit-rate. This is also a useful rule to optimize intra prediction in depth videos.

3.2 Complexity Analysis of Depth Video Coding

For a video sequence, if the details of a CU are richer, CU tends to be divided into the smaller size, and the probability that the angle modes being selected as the optimal mode is high. Conversely, if the texture of a CU is relatively smooth, it is highly possible to split CU into the larger sizes and select non-angle modes as the optimal mode. From the previous part, we observed that depth video has the characteristics of simple structure and obvious boundaries. In order to further study the coding complexity of depth video, we have performed experiments on different standard depth sequences. First, **Table 1** shows the occupation ratio of the optimal mode for each CU. It can be clearly seen that in various types of depth sequences with different resolutions, the probability that the non-angle modes (Planar mode and DC mode) and two special angle modes (Horizontal mode and Vertical mode) are selected as the optimal mode is about 50%. Among 67 prediction modes, these 4 modes are always selected as the optimal mode with one-half probability, which is quite high. In other words, these modes are of great importance. Thus, to guarantee the accuracy of the optimal mode, Planar mode, DC mode, Horizontal mode and Vertical mode are added in the initialization candidate list.

Furthermore, in order to determine the best way to split the Coding Tree Unit (CTU), it is necessary to recursively traverse all possible split depth levels in video coding. **Fig. 4** shows the CU final partition size distribution for 4 different depth sequences. We can observe that the probability that CTU is eventually split into larger sub-CUs is relatively high. For *PoznanHall*, the proportion of 64x64 and 32x32 CU is high, accounting for about 80% of the total. Due to many luminance gradient areas in *Kendo* and *UndoDancer*, the occurrence probability of smaller CU sizes is slightly increased. Besides, the texture of the *Balloons* is more complicated, so the CU needs to be divided with smaller CU sizes. Therefore, the features of depth video coding can be fully utilized to terminate some unnecessary partition processes early, thereby achieving the purpose of effectively saving encoding time of depth video.

Table 1. Average probability of candidate modes as the optimal mode

Prediction mode / Sequences	0 (Planar)	1 (DC)	18 (Horizontal)	50 (Vertical)	Other modes
PoznanHall	15.7%	9.1%	4.8%	36.3%	34.1%
UndoDancer	14.1%	4.2%	4.8%	33.3%	43.6%
Kendo	19.7%	10.9%	4.2%	14.0%	51.2%
Balloon	21.6%	11.4%	2.0%	13.4%	51.6%

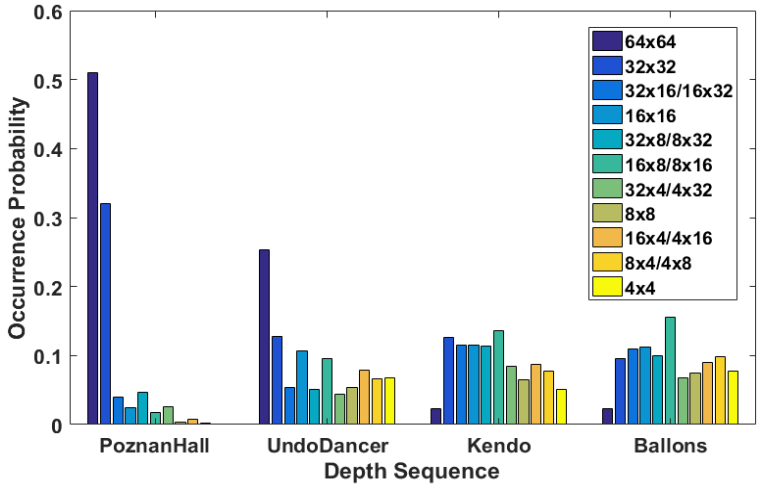


Fig. 4. the distribution of the optimal CU sizes of four depth sequences

4. The Proposed Method

Compared with HEVC, the new VVC standard has introduced some new features into the process of intra prediction. For example, the number of intra prediction modes is extended from 35 to 67, QTMTT structure is used in CU partition, and CU can be used directly for prediction and transform processes, and so on. All the above changes are taken into account in our algorithm.

The presented method is comprised of three stages. Firstly, we design a model to predict the HAD costs of each prediction mode, and select the first 10 modes with lower HAD cost as the initial candidate modes. Secondly, we further optimize the candidate list by statistically analyzing the probabilities that candidate modes become the best mode. Finally, we utilize the Log-normal function to fit the RD cost distribution of split and non-split CUs, and subsequently employee the Bayesian theorem to guide the early termination of CU partition.

4.1 Review of Previous Work

A video sequence is composed of a plurality of still images in a certain order. The temporal correlation and spatial correlation are the basic characteristics of a video sequence. Due to neighboring blocks usually hold similar textures and directional structures, there are strong correlations between coding modes of spatial and temporal neighboring CUs. Take into account this observation, we designed a probability-based scheme that benefits a low complexity HEVC intra encode in our previous work [24]. First, through the estimation model based on the spatiotemporal correlation, the probability of the intra mode of the PU becoming the best mode is predicted to construct an optimal mode candidate list for fast mode decision. Then, the further partition of CU depends on the RD cost distribution of the split and non-split CUs. Then, an early termination scheme based on RD cost is proposed. When combined with early termination techniques, the computational complexity by the mode decision process are significantly reduced.

In this paper, the estimation model is also applied as the basis of the proposed fast intra mode decision algorithm. Different from the previous work [24], the statistical characteristics of depth video are fully studied. According to the characteristics of depth video, the number of candidate modes is optimized. In addition, in the early termination technique, we propose a

fast CU partitioning scheme based on Bayesian theorem, and the adaptive threshold replaces the feature value of the previous offline training. Our proposed method is incorporated into VVC intra prediction for fast coding of depth videos.

4.2 Candidate Modes Initialization

According to the analyses of the previous section, the spatial-temporal correlation of depth video is stronger than that of color videos. In order to acquaint the correlation between adjacent blocks of the depth videos in VVC, we calculate HAD costs of CUs in different types of video sequences with All-Intra (AI) configuration. **Fig. 5** is a schematic diagram of adjacent reference CU, defined as the Upper CU (UCU), the Left CU (LCU) and the co-located CU (coCU), respectively. And the four curves in **Fig. 6** demonstrate the HAD cost distribution in the original 35 prediction modes among neighboring CUs. We can easily find that the distribution trends of HAD costs composed of Planar mode, DC mode and even number prediction modes have obvious similarity among adjacent CUs. Consequently, the cost distributions of neighboring blocks can be used to predict the distribution curve of the current block. Combining with our previous work, the HAD cost estimation model of the current CU

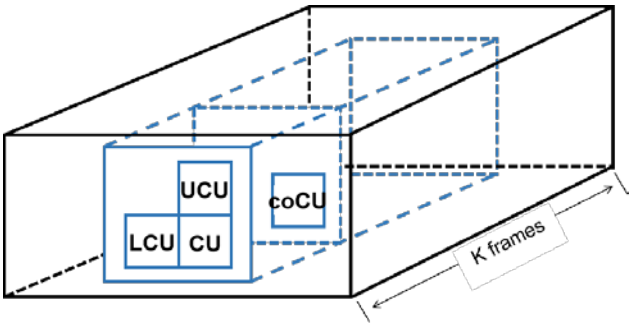


Fig. 5. Schematic diagram of adjacent reference CU for intra prediction

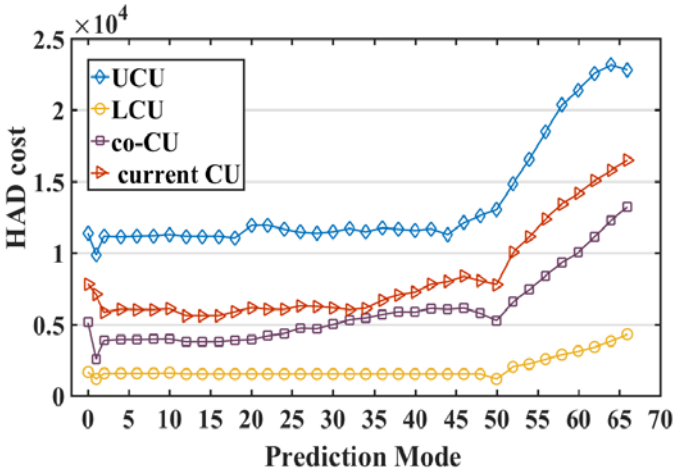


Fig. 6. The distribution of HAD cost among neighboring CUs

is described as:

$$C(t) = \varphi_t \cdot C_l(t) + (1 - \varphi_t) \cdot C_u(t), \tag{1}$$

where C_l and C_u present the distribution of HAD costs in LCU and UCU, respectively. φ_t is the weight for the current CU. According to experimental experiences, we select the first 10 prediction modes with lower cost to initialize the candidate list, and join the four primary modes mentioned in Section 3 to enter the RMD process. This means that we reduced the prediction mode from 35 to 14, which reduces coding complexity. In VVC intra prediction, adjacent blocks may not exist, which results in being failure to fully obtain the parameters required by the model. Hence, the above candidate modes initialization method is only employed in the case that both LCU and UCU exist simultaneously. Otherwise, the current CU will be encoded by the original VVC encoder.

4.3 Optimization of Candidate Modes

As mentioned in Section 2, depth videos have special texture features. They have a large number of smooth areas and sharp edges. In general, for a CU with the simple texture, the probability of the first candidate mode being selected as the optimal mode is much higher than a CU with the complex texture. Based on the above analyses, we calculate the probabilities of each mode in the candidate list being selected as the optimal mode by testing in several different types of sequences. The average results are listed in **Table 2**. For different CU sizes, we find that the average probability of the first candidate mode being selected as the optimal mode is more than 70%, and the first two candidate modes are approximately 90% possibility to be selected as the best mode. That is, when the number of candidate modes is 2, relatively accurate prediction can be achieved.

The analyses in Section 2 suggest us that viewers can tolerate relatively more distortion in depth videos. Inspired by this, we further optimize the candidate mode list. We try several combinations of different candidate numbers based on Part 4.1, and compare them with the original VVC Test Model 2.0.1 (VTM2.0.1) encoder. **Table 3** shows the average performance results for four kinds of combinations. It can be seen that combination III makes an optimal

Table 2. Average probability of candidate modes as the optimal mode

Candidate Mode	First	Second	Third
64x64	72%	17%	11%
32x32	71%	18%	11%
16x16	70%	20%	10%
8x8	75%	17%	8%
4x4	78%	15%	7%
4x8/8x4	76%	16%	8%
4x16/16x4	74%	17%	9%
4x32/32x4	72%	18%	10%
8x32/32x8	70%	20%	10%
16x32/32x16	70%	20%	10%
8x16/16x8	73%	18%	9%

Table 3. Comparison of combinations for different candidate numbers

Candidate Number	Combinations			
	Ⅰ	Ⅱ	Ⅲ	Ⅳ
64x64	2	2	2	1
32x32	2	2	2	1
16x16	2	2	1	1
8x8	2	1	1	1
4x4	2	1	1	1
4x8/8x4	2	1	1	1
4x16/16x4	2	1	1	1
4x32/32x4	2	1	1	1
8x32/32x8	2	2	1	1
16x32/32x16	2	2	1	1
8x16/16x8	2	1	1	1
BDPSNR	-0.036	-0.061	-0.079	0.094
BDBR	0.569	1.024	1.339	1.557
TS	12%	20%	27%	28%

trade off between the coding performance and the computational complexity. Hence, according to the number (denoted as *M*) listed in combination III, we select the first *M* modes from the initial candidate list as the final candidate modes used in the successive mode decision process.

4.4 Early Termination of CU partition

In VVC, RD cost is calculated to determine whether the CU is split or not, which leads to increasing complexity in intra coding. To solve it, we propose an early termination scheme based on Bayesian theorem as follows.

We consider CU partition as a dichotomous problem, namely $I = \{s, n\}$, where s and n denote the cases of continuing to split CU and terminating CU partition, respectively. Then,

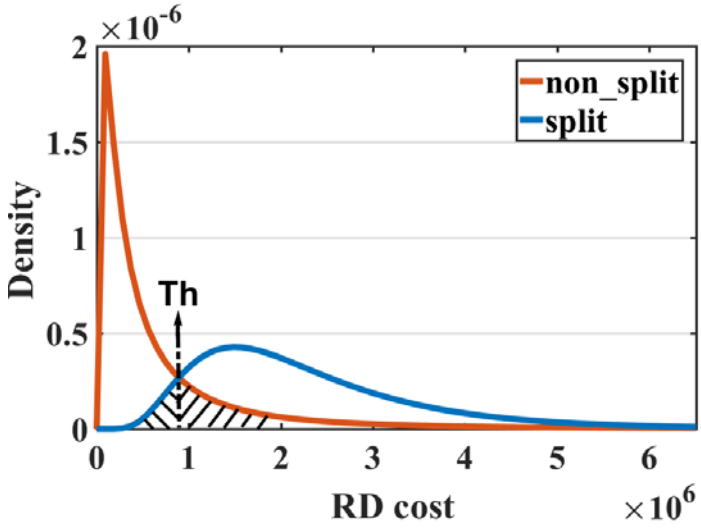


Fig. 7. The distribution of HAD cost among neighboring CUs

inspired by [25], we study the RD cost distribution of split and non-split CUs by performing a lot of simulations in several sequences with different Qps. Based on the experimental results, it is found that the probability distributions for split and non-split CUs can be well fitted by the Log-normal distribution. Taking *UndoDancer* as an example, the results are shown in **Fig. 7**. Therefore, the expression for the Log-normal Density Function can be written as:

$$f_i(r) = \frac{1}{\sqrt{2\pi} r \sigma_i} \exp\left[-\frac{(\ln r - \mu_i)^2}{2\sigma_i^2} \right] \quad i \in \{s, n\}, \tag{2}$$

where r represents RD cost, $f_i(r)$ is the probability density distribution of RD cost in the case of i. μ is the mean and σ^2 is the variance, which can be calculated by the maximum likelihood estimation.

The two marked areas in **Fig. 7** indicate the two error cases (the partition is terminated when CU needs to be split, or CU is no need to be split but is judged to continue splitting) occur in CU partition. Thus the total error probability of CU partition, denoted as P_e, can be calculated by:

$$P_e = P(s) \cdot \int_{-\infty}^{Th} f_s(r) dr + P(n) \cdot \int_{Th}^{+\infty} f_n(r) dr, \tag{3}$$

where $P(*)$ is the prior probability of the case of s or n, and $P(s)+P(n)=1$. Th is the RD cost threshold to decide whether CU needs to be split or not. Apparently, the minimum P_e emerges in the case that the two curves intersect, and the intersection is the optimal threshold Th_{opt}. Hence, the early termination of CU partition in our algorithm can be described as:

$$H = \begin{cases} H_n & if \ r \le Th_{opt} \\ H_s & if \ r > Th_{opt} \end{cases}, \tag{4}$$

where H_n indicates the case of terminating CU partition and H_s means that the CU continues to be split.

A series of experiments have proved that Th_{opt} plus a certain offset value can further optimize the CU partition process and reduce more coding time. Thus, in practice, Th_{opt} is rewritten as:

$$Th'_{opt} = \begin{cases} Th_{opt} + \alpha & if \ Th_{opt} < Th_{\max} - \alpha \\ Th_{\max} & if \ Th_{opt} \ge Th_{\max} - \alpha \end{cases}, \tag{5}$$

where Th_{max} is the maximum threshold for preventing false judgments caused by excessive thresholds. α is the offset value from offline training, which can adapt to different Qps and CU sizes.

In practice, in order to obtain the prior probability $P(*)$ and the statistical parameters (μ and σ^2 of $f_s(r)$ and $f_n(r)$), the proposed algorithm for early termination of CU partition alternates between the online update phase (J frames) and fast CU partitioning phase (K frames) to make the algorithm more adaptable. In the online update phase, the prior probability and the statistical parameters will be online learning. After each learning phase, the optimal threshold

Th_{opt} is updated for fast CU partitioning phase.

4.5 Overall algorithm

The overall algorithm is summarized as follows.

Algorithm 1 Fast Intra Mode Decision

```
if LCU and UCU exist then
    Calculate C by (1)
    Retain the first 10 prediction modes
    Add Planar, DC, 18, 50 mode into the list
    Perform RMD// Reduce the number of prediction modes
else
    Perform the original RMD
end if
Select M modes to join the candidate list
Merge the MPMs into the list
Perform RDO      //Determine the optimal mode
```

Algorithm 2 Early Termination of CU Partition

```
if In online update phase then
    Train P(s), P(n), statistical parameters μ and σ^2
else // Fast CU partitioning phase
    Obtain Th_opt by (3), (4), (6)
    if r ≤ Th_opt then
        H=H_n
    else
        H=Hs
    end if
end if
```

In Algorithm 1, when both LCU and UCU exist, the HAD costs for prediction modes of the current CU are calculated by the estimation model (presented in Section 3 Part 1), and 10 modes with lower cost are retained to initialize the list. Then add four important modes into list to perform RMD. Otherwise, the original RMD process is performed. After that, M modes with lower cost are selected as candidate modes to join in the successive RDO process.

In Algorithm 2, the prior probability and the statistical parameters are learned online during the online update phase. After that, the optimal threshold Th_{opt} is updated to determine whether to terminate the CU partition early.

5. Experimental Results and Analysis

In order to verify the effectiveness of our proposed algorithm, reference software VTM 2.0.1 of VVC is adopted as the software platform. All experimental tests are conducted on a platform with a CPU of Intel i7@3.2GHz and RAM of 32.0 GB, and the development environment is the Microsoft Visual Studio 2015. We set the encoding parameters according to the original VVC configuration file, using the AI configuration [26]. QP settings and Test sequences are inherited from the common test condition (CTC). Qp is set as 22, 27, 32 and 37, respectively, and seven widely used standard depth videos are as test sequences, namely

PoznanStreet, *PoznanHall*, *UndoDancer*, *GhostTownFly*, *Newspaper*, *Balloons*, *Kendo*. **Table 4** shows the configuration parameters of sequences.

Table 4. Parameters of test sequences

Resolution	Sequence	Total Frames	Frame Rare(fps)
1024x768	*Balloons*	300	30
1024x768	*Kendo*	300	30
1024x768	*Newspaper*	300	30
1920x1088	*PoznanHall*	200	25
1920x1088	*PoznanStreet*	250	25
1920x1088	*GhostTownFly*	250	25
1920x1088	*UndoDancer*	250	25

The RD performance of the proposed algorithm is evaluated by the Bitrate increase (ΔBR, %), Peak-Signal-to-Noise-Ratio decrease (ΔPSNR, dB), BDPSNR (dB), BDBR (%) [27], and Time Saving (TS, %). TS represents the entire encoding time saving of the proposed algorithm compared to the original VTM2.0.1 encoder, which is defined as follow,

$$TS = \frac{Time_{ori} - Time_{prop}}{Time_{ori}} \times 100\%, \tag{6}$$

where $Time_{ori}$ and $Time_{prop}$ are the entire encoding time of the VTM 2.0.1 encoder and the proposed algorithm.

Table 5 shows the performance comparison between the proposed algorithm and the original algorithm under different QPs. Especially, there are 150 frames coded for each sequence in this paper. It can be clearly seen that compared with the original VVC algorithm, the proposed algorithm achieves 39.85% depth coding time reduction on average. At the same time, BR of the proposed algorithm increases by 0.71% on average. And for some sequences, such as *PoznanStreet*, *Newspaper*, BR decreases when QP is 37. PSNR decreases by a minimum of 0.04 dB and a maximum by 0.20 dB.

Table 6 illustrates the comparison between the proposed algorithm and BDPSNR and BDBR of VTM 2.0.1 under different sequences. Among them, ATS is the average value of TS. From the results, our fast algorithm can bring 26.26%- 45.29% time saving for each sequence, while it incurs the negligible performance loss, which the average BDPSNR loss of -0.12 dB and the average BDBR increase of 2.10% on average.

Among them, *PoznanHall* saves the least coding time but causes a relatively large performance loss. The reason may be that *PoznanHall* is darker than other sequences and has barely obvious object contours, which leads to incorrect selection of the best prediction mode and consequently decreases the accuracy of intra prediction. Meanwhile, *PoznanHall* is flatter than other sequences, which brings about the difficulty of satisfying the early termination condition and thus results in lower time saving. As a whole, the performance results indicate that our algorithm has strong robustness, for the little fluctuations in the results.

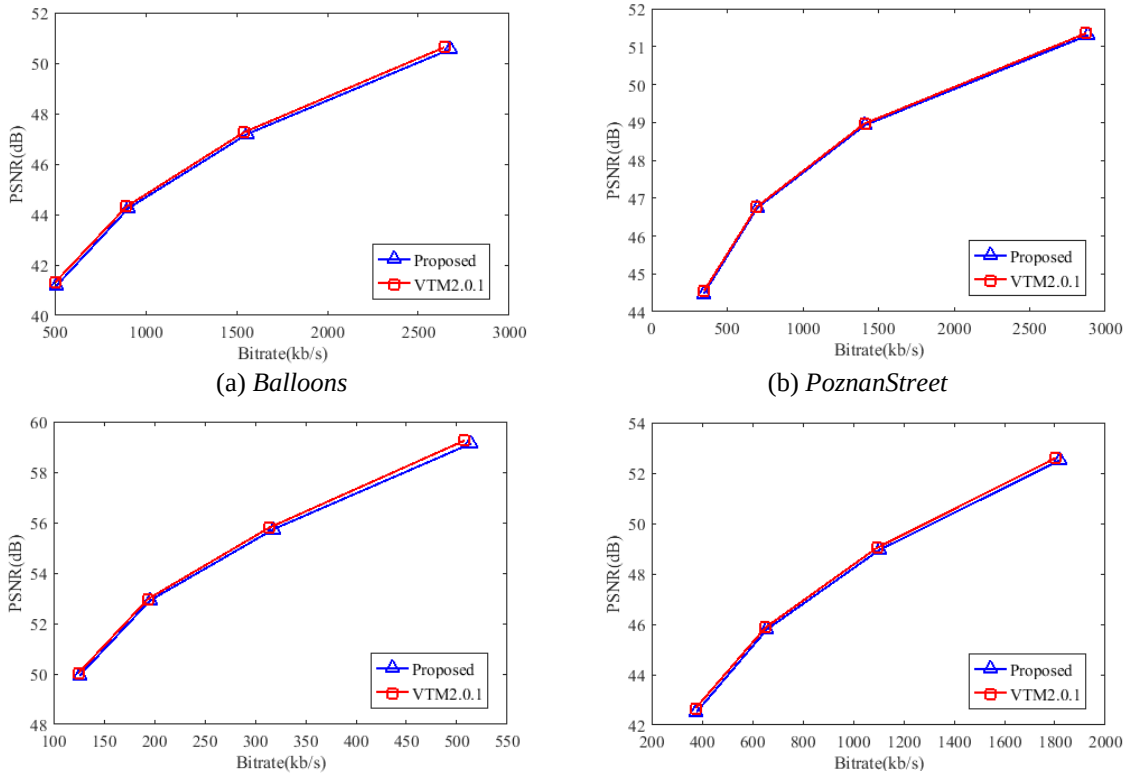
In addition, **Fig. 8** shows the RD curves of VTM 2.0.1 and our proposed algorithm for 4 sequences. It can clearly see that the differences between the two curves in each figure are minimal. This indicates that the algorithm can ensure the stability of the RD performance of the entire system, and is highly effective for improving the performance of depth video coding.

Table 5. The performance comparison between the proposed algorithm and the original algorithm under different QPs

Resolution	Sequence	QP	$Time_{ori}$	$Time_{prop}$	ΔPSNR(dB)	ΔBR(%)	TS (%)
1024x768	*Kendo*	22	4408.48	2411.71	-0.08	1.05	45.29
		27	2579.74	1458.12	-0.09	0.83	43.48
		32	1497.67	892.38	-0.08	0.84	40.42
		37	682.86	373.04	-0.16	0.86	45.37
	Balloons	22	6368.38	3318.65	-0.07	1.25	47.89
		27	3798.45	2102.65	-0.07	1.20	44.64
		32	2156.01	1277.21	-0.07	1.13	40.76
		37	977.23	509.61	-0.15	0.91	47.85
	Newspaper	22	7852.63	4132.24	-0.07	1.17	47.38
		27	4275.12	2338.14	-0.06	1.00	45.31
		32	2419.63	1441.96	-0.05	0.95	40.41
		37	1088.91	566.39	-0.20	-0.31	47.99
1920x1088	*PoznanHall*	22	982.22	634.22	-0.12	1.30	35.43
		27	520.52	356.41	-0.06	1.36	31.53
		32	350.54	265.92	-0.06	1.01	24.14
		37	245.30	211.12	-0.09	1.04	13.93
	UndoDance	22	2068.24	1210.38	-0.10	1.84	41.48
		27	1429.38	883.72	-0.18	0.28	38.17
		32	990.80	654.93	-0.12	0.11	33.90
		37	585.94	411.27	-0.55	-0.07	29.81
	PoznanStreet	22	8006.50	4554.03	-0.05	0.48	43.11
		27	3220.75	1888.76	-0.04	0.30	41.70
		32	1701.20	1066.18	-0.04	0.13	37.97
		37	768.78	457.00	-0.09	-0.80	40.32
	GhostTownFly	22	982.22	634.22	-0.06	0.91	45.02
		27	520.52	356.41	-0.07	0.39	42.03
		32	350.54	265.92	-0.06	0.48	39.09
		37	245.30	211.12	-0.13	0.38	41.54
Average					**-0.09**	**0.71**	**39.85**

Table 6. Comparison of the overall performance of the proposed algorithm and the original algorithm

Resolution	Sequence	BDPSNR(dB)	BDBR(%)	ATS(%)
1024x768	*Kendo*	-0.15	2.36	43.64
	Balloons	-0.14	2.58	45.29
	Newspaper	-0.12	2.31	45.27
1920x1088	*PoznanHall*	-0.15	2.29	26.26
	UndoDance	-0.14	1.80	35.84
	PoznanStreet	-0.05	1.52	40.78
	GhostTownFly	-0.10	1.86	41.92
Average		**-0.12**	**2.10**	**39.85**



(a) *Balloons* (b) *PoznanStreet*

(c) *PoznanHall* (d) *Kendo*

Fig. 8. RD Curves of VTM 2.0.1 and our proposed algorithm

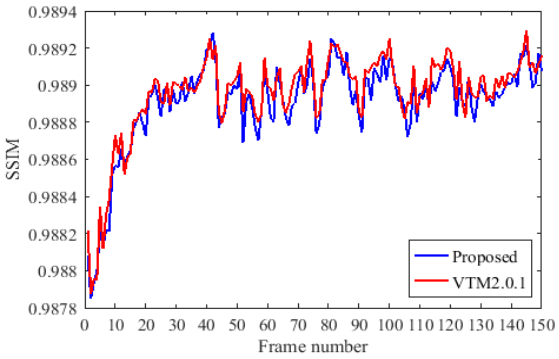
We also compared the performance in terms of SSIM. PSNR is the most commonly used objective evaluation index for images, which is based on the error between corresponding pixels. However, it does not consider the visual characteristics of the human eye. SSIM is a method based on structural information, which is used to measure the similarity between the original signal and the processed signal. The SSIM value can reflect the subjective perception of the human eye better. Therefore, in order to more comprehensively evaluate the coding performance of the proposed algorithm, we provide SSIM comparison between the original encoder and the algorithm. The results of the experiments are summarized in **Tables 7**.

The SSIM value is calculated as the average SSIM of 150 frames of the sequence. It can be observed from the experimental results that SSIM decreases by a maximum of 0.00051, which is insignificant. Therefore, it is proved that the proposed algorithm saves coding time significantly while maintaining almost the same image quality as the original algorithm.

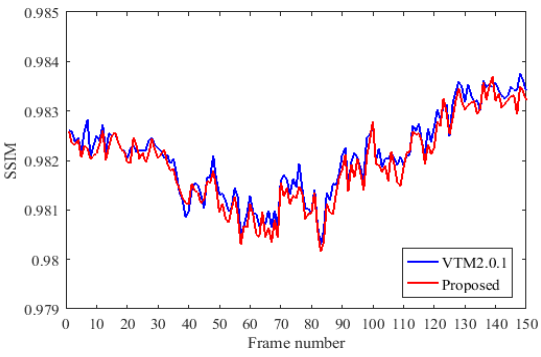
Fig. 9 shows the SSIM value of the first 150 frames for each sequence. QP is 32 for all tests. The SSIM-bitrate curves of VTM 2.0.1 and our proposed algorithm are compared in Fig. 10. It can be observed that both SSIM curve for each frame and SSIM-bitrate curve, the curve of our algorithm is very close to that of the original algorithm. It further proves that the proposed algorithm only causes negligible image quality loss.

Table 7. SSIM comparison between the proposed algorithm and the original algorithm under different QPs

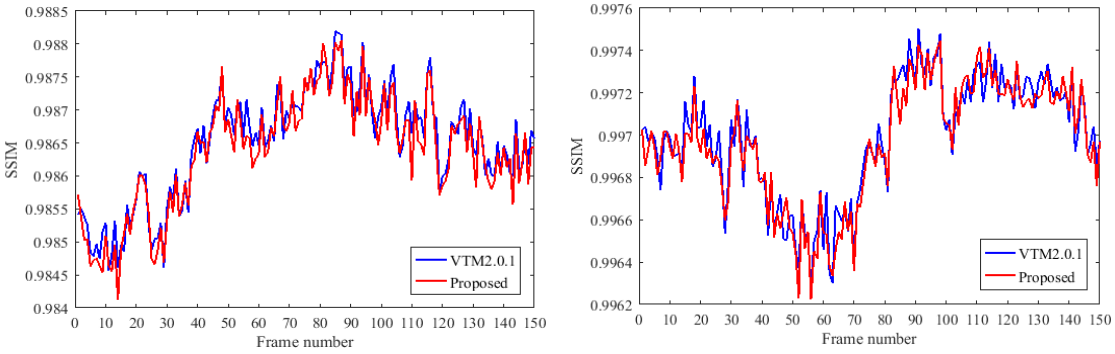
Resolution	Sequence	QP	VTM2.0.1		Proposed		ΔSSIM
			SSIM	Bitrate	SSIM	Bitrate	
1024x768	Kendo	22	0.99420	1802.7824	0.99416	1821.64	-0.00004
		27	0.99084	1093.64	0.99075	1102.7712	-0.00009
		32	0.98649	651.696	0.98637	657.16	-0.00012
		37	0.98016	374.5072	0.97982	377.736	-0.00033
	Balloons	22	0.99210	2643.3488	0.99210	2676.3616	0
		27	0.98763	1540.992	0.98754	1559.4192	-0.00009
		32	0.98209	894.7696	0.98194	904.8944	-0.00015
		37	0.97478	503.9984	0.97438	508.592	-0.00040
	Newspaper	22	0.98923	3176.5296	0.98916	3213.7504	-0.00007
		27	0.98403	1833.4688	0.98394	1851.7232	-0.00009
		32	0.97778	1041.9104	0.97760	1051.7744	-0.00018
		37	0.96955	573.1104	0.96904	571.3312	-0.00051
1920x1088	PoznanHall	22	0.99857	507.5707	0.99856	514.1627	-0.00001
		27	0.99780	313.6867	0.99779	317.9667	-0.00001
		32	0.99699	193.7293	0.99697	195.6933	-0.00002
		37	0.99595	124.5587	0.99590	125.848	-0.00005
	UndoDance	22	0.99756	877.8533	0.99755	894.012	-0.00001
		27	0.99667	598.04	0.99664	599.7293	-0.00003
		32	0.99577	386.1027	0.99573	386.5133	-0.00004
		37	0.99427	233.792	0.99412	233.6333	-0.00015
	PoznanStreet	22	0.96592	2873.4853	0.96601	2887.2987	-0.00003
		27	0.96316	1407.776	0.96319	1411.936	-0.00003
		32	0.96463	696.292	0.96467	697.2107	-0.00005
		37	0.96158	347.3013	0.9616	344.5147	-0.00010
	GhostTownFly	22	0.99678	1929.8987	0.99677	1947.42	-0.00001
		27	0.99527	1153.064	0.99523	1157.5933	-0.00004
		32	0.99300	653.888	0.99295	657.0547	-0.00005
		37	0.98954	364.9213	0.98934	366.292	-0.00020



(a) *PoznanStreet*

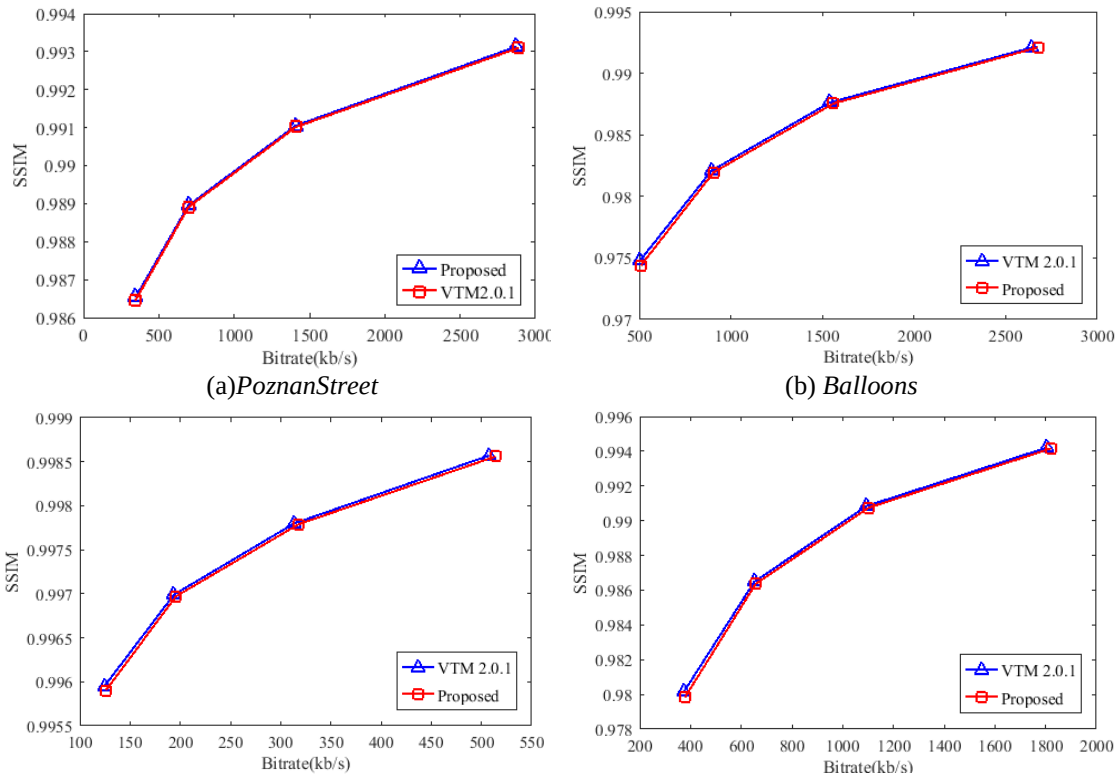


(b) *Balloons*



(c) *PoznanHall* (d) *Kendo*

Fig. 9. SSIM Curves of VTM 2.0.1 and our proposed algorithm

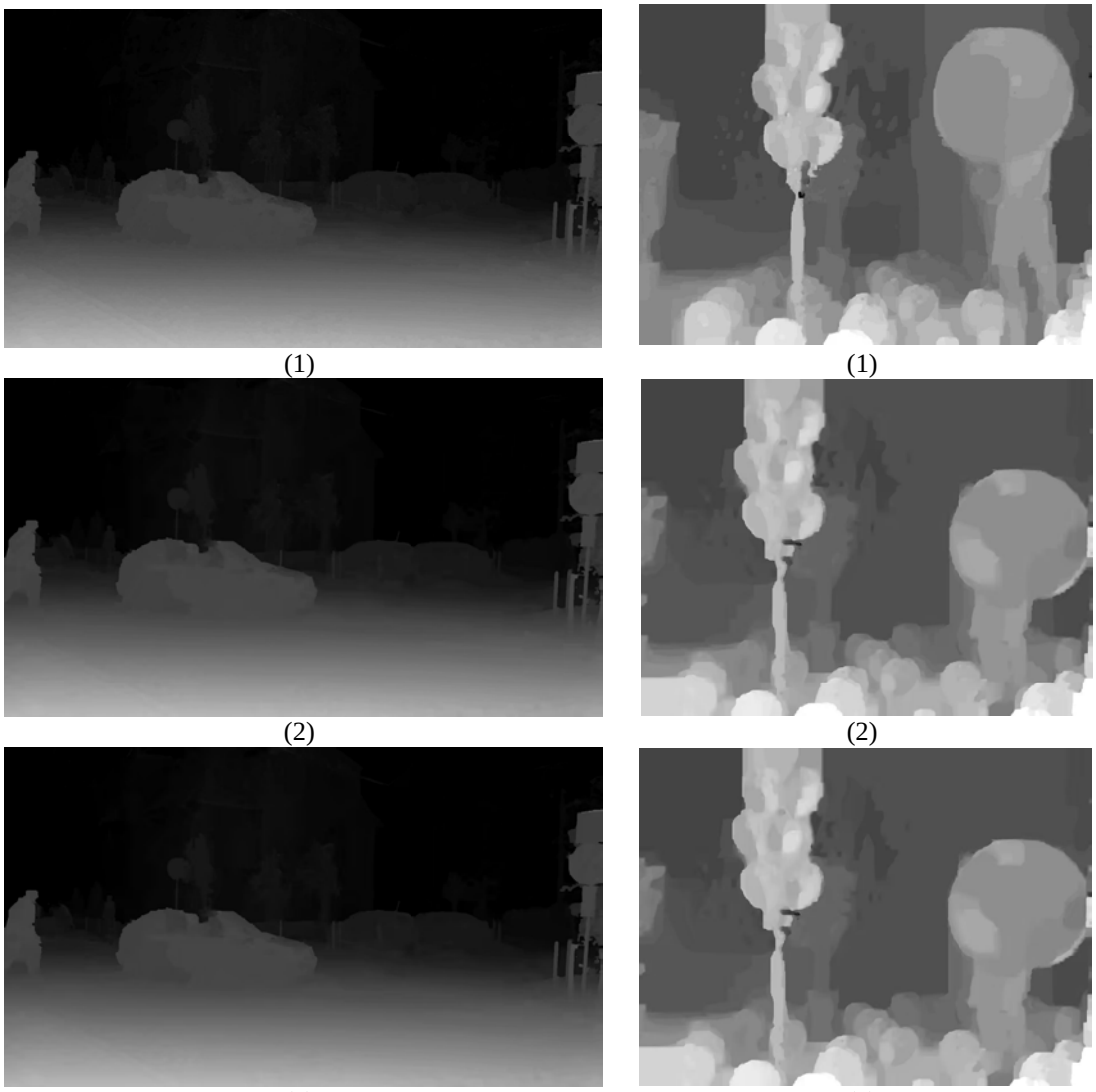


(a)*PoznanStreet* (b) *Balloons*

(c) *PoznanHall* (d) *Kendo*

Fig. 10. SSIM-bitrate Curves of VTM 2.0.1 and our proposed algorithm

Fig. 11 compares the visual quality of our algorithm with the VTM2.0.1 algorithm, where shows the 100th frame (cropped for visualization) of Balloons and *PoznanStreet*. It can be observed that under the condition of greatly reducing the encoding time, our algorithm guarantees that the visual quality is basically consistent with the original algorithm. The display of sequence frames further validates the effectiveness of the proposed method.



(1) (1)

(2) (2)

(3) (3)

(a) *PoznanStreet* (b) *Balloons*

Fig. 11. Visual quality comparison between the proposed algorithm and the VTM2.0.1 algorithm (1) Original. (2) Coded with the VTM2.0.1 algorithm.(3) coded with the proposed algorithm.

6. Conclusion

In this paper, we proposes a hybrid optimization approach for fast prediction of depth video coding in VTM 2.0.1, which aims to reduce the computational complexity of VVC. Based on the characteristics of depth video, firstly, a fast mode decision algorithm based on HAD cost is proposed to reduce the number of modes. Then a fast CU partition algorithm based on Bayesian theorem is developed to determine the CU partition early. The experimental results show that, compared with the original VTM encoder, the proposed algorithm effectively speeds up the encoding time of depth video about 39.85% on average, accompanied by the negligible loss of video quality.

The algorithm mainly uses some features and correlations which are inherent in the depth video. However, there is also a certain correlation between depth video and the corresponding color video. For future work, more encoding time saving would be achieved if depth video is combined with color video for joint coding.

References

[1] J. Chen, Y. Ye, S. Kim, "Algorithm description for Versatile Video Coding and Test Model 2 (VTM 2)," in *Proc. o f Document JVET-K1002-v2 11th JVET Meeting, Ljubljana, SI*, July, 2018.

[2] T. Lin, H. Jiang, J. Huang and P. Chang, "Fast intra coding unit partition decision in H.266/FVC based on spatial features," *Journal of Real-Time Image Processing*, 17, 493-510, 2020. Article (CrossRef Link)

[3] G. J. Sullivan and T. Wiegand, "Video compression-from concepts to the H.264/AVC standard," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 18–31, Jan. 2005. Article (CrossRef Link)

[4] G. J. Sullivan, J. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649-1688, Dec. 2012. Article (CrossRef Link)

[5] D. Yoon and Y. Ho, "Fast depth video coding method using adaptive edge classification," in *Proc. of 2011 Visual Communications and Image Processing (VCIP), Tainan*, pp. 1-4, 2011. Article (CrossRef Link)

[6] G. Sanchez, M. Saldanha, G. Balota, B. Zatt, M. Porto and L. Agostini, "Complexity reduction for 3D-HEVC depth maps intra-frame prediction using simplified edge detector algorithm," in *Proc. of 2014 IEEE International Conference on Image Processing (ICIP), Paris, France*, pp. 3209-3213, Jan. 2014. Article (CrossRef Link)

[7] T. L. Da Silva, L. V. Agostini and L. A. da Silva Cruz, "Fast mode selection algorithm based on texture analysis for 3D-HEVC intra prediction," in *Proc. of 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy*, pp. 1-6, June, 2015. Article (CrossRef Link)

[8] G. Sanchez, L. Agostini and C. Marcon, "Complexity reduction by modes reduction in RD-list for intra-frame prediction in 3D-HEVC depth maps," in *Proc. of 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD*, pp. 1-4, May. 2017. Article (CrossRef Link)

[9] 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 Transactions on Image Processing*, vol. 27, no. 9, pp. 4195-4206, Sept. 2018. Article (CrossRef Link)

[10] P. Merkle, K. Müller, D. Marpe and T. Wiegand, "Depth Intra Coding for 3D Video Based on Geometric Primitives," in *Proc. of IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 570-582, Mar. 2016. Article (CrossRef Link)

[11] R. Zhang, K. Jia, P. Liu and Z. Sun, "Edge-Detection Based Fast Intra-Mode Selection for Depth Map Coding in 3D-HEVC," in *Proc. of 2019 IEEE Visual Communications and Image Processing (VCIP), Sydney, Australia*, pp. 1-4, 2019. Article (CrossRef Link)

[12] H. Wang and Q. Li, "Fast Decision Algorithm for Intra Mode in Depth Map of 3D-HEVC," *Proc. of 2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP), Weihai, China*, pp. 327-331, 2019. Article (CrossRef Link)

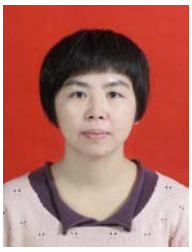
[13] H. Hamout and A. Elyousfi, "Fast Depth Map Intra Coding for 3D Video Compression Based Tensor Feature Extraction and Data Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 1933-1945, 2020. Article (CrossRef Link)

[14] R. Conceição, G. Avila, G. Corrêa, M. Porto, B. Zatt and L. Agostin, "Complexity reduction for 3D-HEVC depth map coding based on early Skip and early DIS scheme," in *Proc. of 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ*, pp. 1116-1120, 2016. Article (CrossRef Link)

[15] M. Kim, N. Ling, L. Song, "Fast single depth intra mode decision for depth map coding in 3D-HEVC," in *Proc. o f 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy*, pp. 1-6, 2015. Article (CrossRef Link)
[16] H. Chen, C. Fu, Y. Chan and X. Zhu, "Early intra block partition decision for depth maps in 3D-HEVC," in *Proc. of 2018 25th IEEE International Conference on Image Processing (ICIP), Athens*, pp. 1777-1781, 2018. Article (CrossRef Link)
[17] L. F. R. Lucas, K. Wegner, N. M. M. Rodrigues, C. L. Pagliari, E. A. B. da Silva and S. M. M. de Faria, "Intra predictive depth map coding using flexible block partitioning," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4055-4068, Nov. 2015. Article (CrossRef Link)
[18] K. Chung, Y. Huang, C. Lin and J. Fang, "Novel bitrate saving and fast coding for depth videos in 3D-HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 10, pp. 1859-1869, Oct. 2016. Article (CrossRef Link)
[19] Y. Hsu, J. Lin, M. Chen, C. Yeh, M. Lin and W. Lu, "Acceleration of Depth Intra Coding for 3D-HEVC by Efficient Early Termination Algorithm," in *Proc. of 2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Chengdu*, pp. 127-130, 2018.
Article (CrossRef Link)
[20] Q. Zhang, N. Li, L. Huang and Y. Gan, "Effective early termination algorithm for depth map intra coding in 3D-HEVC," *Electronics Letters*, vol. 50, no. 14, pp. 994-996, July 2014.
Article (CrossRef Link)
[21] C. Fu, H. Chen, Y. Chan, S. Tsang, H. Hong and X. Zhu, "Fast Depth Intra Coding Based on Decision Tree in 3D-HEVC," *IEEE Access*, vol. 7, pp. 173138-173147, 2019.
Article (CrossRef Link)
[22] Methodology for the subjective assessment of video quality in multimedia applications, ITU-R Rec. BT. 1788, 2007.
[23] Z. Zheng, J. Huo, B. Li and H. Yuan, "Fine virtual view distortion estimation method for depth map coding," *IEEE Signal Processing Letters*, vol. 25, no. 3, pp. 417-421, Mar. 2018.
Article (CrossRef Link)
[24] H. Wei, M. Wang, Y. Xu, Y. Liu, T. Zhao, "Probability-based intra encoder optimization in high efficiency video coding," in *Proc. of 2018 IEEE Visual Communications and Image Processing (VCIP), Tai Wan, China*, 1-4, 2018. Article (CrossRef Link)
[25] 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, Sept. 2013. Article (CrossRef Link)
[26] Common test conditions and software reference configurations , JVET J1010-v1, 2018.
[27] Calculation of average PSNR differences between RDcurves, VCEG M33, 2001.

Hongan Wei received the M.S. degree from the Fuzhou University in 2007. He is now an associate professor in the College of Physics and Information Engineering, Fuzhou University, Fujian, China. His research interests mainly include image/video compression, video codec and transmission.

Binqian Zhou is currently pursuing the M.S. degree in Electronic and Communication Engineering from Fuzhou University, Fujian, China. Her research interests include video coding.



Ying Fang is currently a Ph.D candidate with the College of Physics and Information Engineering, Fuzhou University, Fujian, China. Her research interests mainly include audio-visual systems, multimedia communication, quality of experience.

Yiwen XU received the Ph.D degree in the department of electronic engineering from Xiamen University, Xiamen, China, in 2012. He has been an Associate Professor with the College of Physics and Information Engineering, Fuzhou University, Fujian, China, since 2013. His research interests lie in multimedia information processing, video codec and transmission, and video quality assessment.

Tiesong Zhao received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2011. He has served as a Research Associate with the Department of Computer Science, City University of Hong Kong, from 2011 to 2012, a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, from 2012 to 2013, and a Research Scientist with the Ubiquitous Multimedia Laboratory, State University of New York at Buffalo until 2015. He is currently a Professor with the College of Physics and Information Engineering, Fuzhou University, Fujian, China. His research interests include image/video signal processing, visual quality assessment, and video coding and transmission.