Early Merge Mode Decision for Depth Maps in 3D-HEVC

Hao Chen, Chang-Hong Fu*, Yijin Zhang
School of Electronic and Optical Engineering
Nanjing University of Science and Technology
Nanjing, China
Email: enchfu@njust.edu.cn*

Yui-Lam Chan, Wan-Chi Siu
The Department of Electronic and Information Engineering
The Hong Kong Polytechnic University
Hong Kong, China

***Abstract*—With the introduction of several new coding tools and technology for depth maps in 3D-HEVC, the coding efficiency is improved at the expense of computational complexity. Meanwhile, the inter-view correlation is not considered and used well for depth maps, while it is developed maturely in texture video coding. Therefore, in this paper, we propose an early determination scheme for the best prediction mode of depth maps in dependent views. The proposed algorithm can early determine the Merge mode as the best one and if it is , all other candidates would be skipped by taking advantage of the inter-view information and the intermediate RD-cost comparison. Experimental results show that the proposed algorithm can provide about 22.1% remarkable time reduction with almost no BDBR increased compared with the state-of-the-art algorithm.**

Keywords—early mode decision, statistical analysis, inter-view correlation, 3D-HEVC, depth coding, dependent views

I. INTRODUCTION

With the rapid development of three-dimensional (3D) content capture and display technologies, 3D video system has gained great interest from everywhere in recent years. As a new famous efficient format of 3D data, multi-view video plus depth (MVD) is formed by texture video in several different views and their corresponding depth maps [1], [2]. In order to better reconstruct the 3D scene, additional required virtual views could be generated by exploiting Depth-Image-Based-Rendering (DIBR) [3] [4].

As a 3D video coding standard, 3D-HEVC basically adopts the key existing technology in HEVC [5]. Meanwhile, 3D-HEVC introduces disparity-compensated prediction (DCP) for dependent views referring to the reconstructed information in coded basic view. The reconstructed texture video information also could be used as motion parameter inheritance (MPI) for the coding of corresponding depth maps. And different from texture video, depth maps have large flat areas separated by sharp edges. If the encoder could take these characteristics of depth maps into account, depth coding could be more efficient. Hence, several new coding tools such as Depth Intra Single Mode (DIS), Depth Modelling Mode (DMM), Segment-wise DC coding (SDC) and View Synthesis Optimization (VSO) are designed for high efficient depth coding in 3D-HEVC. These new technologies and tools, together with traditional inter or intra coding methods, have achieved great coding efficiency for depth maps. However, these new tools also make the mode decision process more complex, which dramatically increases the computational complexity of 3D-HEVC. Obviously, it is meaningful to study fast coding algorithm to decrease the complexity of 3D-HEVC without damaging 3D video quality.

There already have been some fast algorithms [6, 7, 8, 16, 17] designed for depth coding in 3D-HEVC. Kang and Chung proposed a fast algorithm utilizing the coded corresponding texture intra mode information to early decide the depth intra mode [6]. In [7], an early detection of edges in depth coding unit was proposed, and different intra mode decision methods are taken for smooth blocks and edged blocks. In [8], Miok et al. proposed an early DIS mode decision by simply taking account of the variance of input depth map and model-based view synthesis distortion estimation.

However, the existing fast algorithms for depth coding seldom focus on the strong correlation between inter-views. And the large flat areas in depth maps have not been fully exploited. In this paper, a fast depth mode decision based on inter-view correlation is proposed to reduce the computational complexity of 3D-HEVC. The idea of inter-view prediction, which is originally employed in fast texture coding, is borrowed to depth coding. And the new introduced mode DIS in depth coding is utilized to further improve the scheme. The final proposed algorithm achieves a remarkable performance in decreasing the coding complexity with almost no loss of video quality.

The rest of the paper is organized as follows. Section II briefly introduces the depth mode decision process in 3D-HEVC and states our motivation by the statistics. In Section III, a fast method is proposed to early terminate the mode decision process for the dependent views of depth maps. Simulation results and conclusions are given in Sections IV and V, respectively.

II. BACKGROUND AND MOTIVATION

A. Depth mode decision in 3D-HEVC

In 3D-HEVC, the partitioning quadtree coding structure for texture video and depth maps is inherited from HEVC. The input video is divided into blocks called coding tree units

The work is supported in part by National Natural Science Foundation of China (No. 61301109), and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Grant No. PolyU 152052/15E)

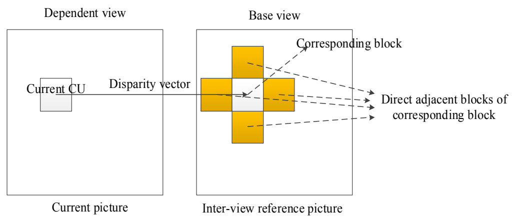

Fig. 1. Five nter-view neighboring blocks prposed in [10]

(CTUs) first. These CTUs are further split into sub-blocks named coding units (CUs) flexibly. The splitting depth of CUs is from 0 to 3. Before each splitting, a CU would define a region called prediction units (PU) sharing the same prediction mode. There are 12 PU modes for inter-frames in depth maps, consisting of Merge mode (including 2N×2N Merge mode and Skip mode, which is a special Merge mode without PU residuals), DIS, eight Inter modes (2N×2N, N×N, N×2N, 2N×N, 2N×nU, 2N×nD, nL×2N, and nR×2N), and two Intra modes (2N×2N, N×N).

Considering the depth maps having large smooth areas surrounded by sharp edges, 3D-HEVC adopted some new modes like DIS and DMM in the depth coding. DMM is a set of intra modes that partitions the area of the intra block into two non-rectangular regions, where each region is represented by a constant value, and it is more suitable for the sharp edges. The compression performance of depth map is improved by DMM, but more computational complexity is introduced at the same time. On the contrary, another new mode DIS will not introduce too much computational burden. It directly uses the reconstructed value of spatial neighboring blocks to represent the current coding block, which is useful in coding flat regions. Both DIS and DMM are added into the mode candidates for PU mode decision in depth coding together with other traditional modes.

For each splitting depth level of CUs, a best PU mode is determined by comparing the rate and distortion (RD) costs of all possible modes, which is also called rate distortion optimization (RDO) process. The RD cost calculation is defined as follow:

$$J(m) = D_{VSO}(m) + \lambda \cdot B(m) \qquad m \in C \tag{1}$$

where C is the set of all possible modes, $J(m)$ denotes the RD cost calculation, $D_{VSO}(m)$ is the distortion of mode m calculated by view synthesized optimization, λ is the Lagrange multiplier, $B(m)$ indicates the number of bits used to encode the current CU with mode m. Finally, the mode with the least RD cost is selected as the best PU mode for current CU.

B. Extremely Biased PU mode distirbution

In depth coding, Merge mode allows CUs to inherit the motion information from their reconstructed neighboring CUs, which could be spatial, temporal and texture neighbors. The CUs finally choosing Merge mode most likely have similar values with their neighboring CUs. Since depth maps usually have large flat areas surrounded by sharp edges, Merge mode is quite suitable for those blocks inside the same flat area [9]. Besides Merge mode, DIS is also an intra mode designed for representing flat areas in depth maps as mentioned before, while other inter or intra modes are used to predict more complicated regions.

TABLE I. PU MODE DISTRIBUTION OF DEPTH MAPS

Sequences	Merge	DIS	Other inter modes	Other intra modes
Balloons	83.65%	8.18%	1.41%	6.76%
Kendo	76.42%	14.74%	1.44%	7.40%
Newspaper	76.97%	12.09%	1.43%	9.51%
GT_Fly	97.32%	1.21%	0.49%	0.98%
Poznan_Hall2	87.42%	10.66%	0.82%	1.10%
Poznan_Street	90.53%	4.64%	0.86%	3.97%
Shark	94.68%	1.07%	2.42%	1.83%
Undo_Dancer	93.85%	2.87%	1.95%	1.34%
Average	87.61%	6.93%	1.35%	4.11%

Generally, all possible modes are examined in order to decide the best PU mode for each splitting CU in a depth map, which is a complex and time-consuming process. To explore the possibility of each mode and find whether there is some unnecessary work in RDO process that could be skipped, we analyzed the PU mode distribution of depth maps in the current reference software implementation of 3D-HEVC (HTM).

The statistical results of four categories (Merge mode, DIS, other inter modes and other intra modes) are shown in Table I. It could be observed that averagely 87.61% of CUs selecting Merge mode (Skip, 2N×2N) as their best PU mode, while 6.93% of CUs selecting DIS, and the CUs selecting the other inter or intra modes only account for 5.46%. The extremely biased distribution of prediction modes could be explained by the characteristic analysis about Merge mode above.

Obviously, if we could make an early decision whether Merge mode is the best PU mode, it is not necessary to check all other modes. As a result, the computational complexity of RDO process could be reduced dramatically.

III. PROPOSED EARLY MERGE MODE DECISION

As described in Section II, if we could find a proper condition to early distinguish the CUs (which are more likely to select Merge mode as their PU mode) from other CUs, the RDO process would be much simplified and huge calculation burden could be saved for those skipped candidates.

Recently, inter-view correlation has been considered in fast mode decision of texture video coding and achieved great improvement in reducing the computational complexity [10, 11, 18]. It is worth considering whether the inter-view correlation could also help early decide the Merge mode in depth maps. However, since DIS, DMM and some other new coding tools are introduced for depth maps, more factors could be taken into consideration for utilizing the inter-view correlation in depth mode decision.

A. Fast texture coding utiliziing inter-view correlation

3D-HEVC allows the reconstructed information in base view to be used in the coding process of CUs in dependent views. It is noted that depth maps in different views only have differences in camera angle. Since the same scene at the same time is recorded in depth maps in different views, there must be strong inter-view correlation that could be utilized.

In [10], five inter-view neighboring blocks (as shown in Fig.1) are defined. They are used to represent inter-view correlation and provide guidance information for coding texture video in dependent views. Two conditions for the early decision of Merge mode are proposed in [10]:

$C_{[10]}$: The condition set proposed in [10]

- All the five inter-view neighbouring blocks are coded as Merge modes.
- The RD cost of Skip mode is less than 2N×2N Merge mode for the current CU.

The first condition represents the inter-view information about Merge mode from neighboring CUs, obviously. For the second condition, as mentioned before, the only difference between the two kinds of Merge mode is that no residuals need to be coded in Skip mode, while in 2N×2N Merge Mode we need to code residuals. When the second condition is fulfilled, it could be assumed that the motion information obtained from neighbouring blocks is good enough in prediction, such that there are quite less residuals left.

For convenience, the two conditions in [10] are combined and named as $C_{[10]}$ in this paper. When $C_{[10]}$ is satisfied, the CUs of texture video in dependent views only check Merge mode and skip all other modes, which could reduce the encoding time by 32% with 0.0% BD-rate increase [10].

B. The inter-view correlation in depth maps

Due to the great performance in coding texture video, the fast algorithm in [10] has been integrated into HTM 16.1. We find that $C_{[10]}$ could also be considered as a good way to utilize the inter-view correlation of the depth maps in dependent views.

Table II shows the PU mode distribution when $C_{[10]}$ is applied in coding the depth maps in dependent views. From the table, we can see that 97.84% of CUs choose Merge mode (Skip) as their best PU mode, while the CUs finally choose other modes only account for 2.16% when the condition $C_{[10]}$ is satisfied. The accuracy of early deciding Merge mode looks convincing. But when we implement $C_{[10]}$ in the depth coding, the BD-rate increases up to 0.7% as shown in Table V. Therefore, we continue examining the performance of $C_{[10]}$ in depth coding, and try to further improve the accuracy.

As a new PU mode for depth coding, DIS even perform better than Merge mode in some flat areas. It could be observed from Table II that, for all CUs where $C_{[10]}$ is satisfied, 2.16% of them select Merge mode wrongly. Among the wrong selections, most of CUs (about 62%) choose DIS as the best prediction mode. As a result, we could further improve the accuracy of the early determination by checking the RD cost of DIS. The new condition C_{pro} is listed as follows:

C_{pro}: The proposed condition set

- All the five inter-view neighbouring blocks are coded as Merge modes.
- The RD cost of Skip mode is less than 2N×2N Merge mode for the current CU.
- The RD cost of Skip mode is less than DIS for the current CU.

In order to decide whether the condition C_{pro} could meet our needs, new experiments are conducted and the results are shown in Table III. It could be seen from the table that 99.35% of CUs on average finally selecting Merge mode (Skip) as their best PU mode when C_{pro} is satisfied, which is more accurate than $C_{[10]}$ from the fast texture coding.

TABLE II. PU MODE DISTRIBUTION OF DEPTH MAPS FOR DEPENDENT VIEWS WHEN $C_{[10]}$ IS SATISFIED

Sequence	Merge(%)	DIS(%)	Other modes(%)
Balloons	98.16	1.11	0.73
Kendo	94.45	4.66	0.89
Newspaper	98.27	0.87	0.86
GT_Fly	99.46	0.37	0.16
Poznan_Hall2	97.16	2.55	0.29
Poznan_Street	98.96	0.67	0.37
Shark	98.37	0.42	1.21
Undo_Dancer	97.92	1.09	0.99
Average	97.84	1.47	0.69
		2.16	

TABLE III. EXPERIMENTAL RESULTS WHEN C_{PRO} IS SATISFIED

Sequence	Merge(%)	P_{Cpro}(%)
Balloons	99.30	67.90
Kendo	99.12	54.69
Newspaper	99.16	67.26
GT_Fly	99.85	75.53
Poznan_Hall2	99.76	72.11
Poznan_Street	99.67	75.48
Shark	98.86	66.92
Undo_Dancer	99.09	67.80
Average	99.35	68.46

Table III also gives the percentage of CUs satisfying the condition C_{pro}, which is defined as:

$$P_{C_{pro}} = \frac{N(C_{pro} = true)}{N(all \quad CUs)} \times 100\% \qquad (2)$$

where N (…) represents the number of CUs that meet the condition in brackets.

The result of P_{Cpro} illustrates that CUs conforming to the condition C_{pro} occupy a large proportion (68.46% on average) in all CUs, which further demonstrates the feasibility of early judging Merge mode according to the condition C_{pro}.

The analysis above proves that the inter-view correlation is highly advantageous for mode decision of depth maps in dependent views. The RDO process could be much simplified by utilizing the proposed condition C_{pro} with no degradation in coding performance.

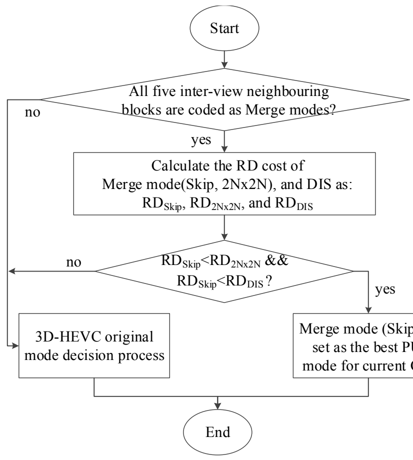

Fig. 2. Flow chart of proposed depth mode decision for each CU

C. The flow chart of proposed algorithm

It is illustrated in Part A and B that the strong inter-view correlation could also be utilized for depth maps besides texture video in dependent views, and the accuracy of early determination could be further improved to 99.35% by considering the special DIS mode of depth map. In this section, a fast depth mode decision algorithm utilizing the inter-view correlation is proposed for depth maps in dependent views. The details are presented as follows.

Based on the statistics in Table III, it is reasonable to only check Merge mode and DIS, and early terminate the RDO process when the condition C_{pro} is satisfied. The flow chart of our fast algorithm is shown in Fig.2. First, the reconstructed information from base view is collected. If the five inter-view neighbouring blocks are all coded as merge modes, then Merge mode and DIS are examined for current CU. If the RD cost of Skip mode is less than the cost of 2N×2N Merge mode and DIS for the current CU, then the Merge mode (Skip) would be set as the best PU mode for current CU and the RDO process would be early terminated. Otherwise, if any of the two conditions is not satisfied, original HTM encoder would be employed for current CU.

IV. SIMULATION RESULTS

In order to evaluate the performance of the proposed fast mode decision algorithm of depth maps in dependent views, we implement the algorithm in HTM-16.1, which has already adopted the state-of-the-art coding methods [12], [13]. The original depth mode decision process in HTM-16.1 has been used as an anchor. The fast mode decision method for texture video [10] is applied in depth coding and compared to the anchor together with the proposed scheme. Table IV shows the main experimental conditions, which follows the Common Test Condition (CTC) defined by JCT-3V for HTM experiments [14]. Test sequences include *Balloons*, *Kendo*, *Newspaper*, *GT_Fly*, *Poznan_Hall2*, *Poznan_Street*, *Shark and Undo_Dancer*. For each sequence, three-view texture videos and their corresponding depth maps are used for the experiment. The three views include an independent (center) view and two dependent (side) views. The experiments are conducted on the platform with the CPU of Intel Xeon(R) E3-1230 @3.3GHz and RAM 16.0GB in a clear environment with all other unnecessary processes terminated.

TABLE IV. EXPERIMENTAL CONDITIONS

Profile	Main
Group size	8
Intra Period	24
CTU size	64× 64
QP pairs	(25,34), (30,39), (35,42), (40,45)
Coding Structure	Hierarchical B frame structure
Fast Options-QTL	Enabled
Software Version	HTM 16.1

TABLE V. CODING RESULTS FOR DEPTH MAPS IN DEPENDENT VIEW

Sequence	**Zhang [10] vs HTM 16.1**		**Proposed vs HTM 16.1**	
	BDBR(Syn.)(%)	**ΔT(%)**	**BDBR(Syn.)(%)**	**ΔT (%)**
Balloons	-0.1	-22.1	0	-21.1
Kendo	0	-17.5	-0.1	-17.4
Newspaper	-0.1	-23.4	0	-22.0
GT_Fly	0	-26.3	0	-24.5
Poznan_Hall2	**+0.7**	-22.9	+0.1	-20.9
Poznan_Street	+0.1	-27.3	0	-25.2
Shark	+0.1	-24.0	+0.1	-22.8
Undo_Dancer	+0.2	-25.2	0	-22.9
Average	**+0.11**	-23.6	**+0.01**	-22.1

The complexity reduction and coding efficiency of coding results for depth maps in dependent views are taken into account to evaluate the performance of the fast algorithm. The coding efficiency is evaluated by the BDBR [15], which is calculated by the PSNR of synthesized views and the total bitrate of depth and texture videos. As the main concern for us, the average of encoding time saving ΔT is used to evaluate the complexity reduction, which is computed as:

$$\Delta T = \frac{T_P - T_O}{T_O} \times 100\% \tag{3}$$

where T_P represents the encoding time of depth maps in dependent views by the proposed fast algorithm, and T_O is the encoding time of depth maps in dependent views by original encoder HTM 16.1.

The simulation results are shown in Table V. It could be observed that the proposed method could effectively achieve time saving with high-quality coding performance for depth map coding. The proposed method reduces the encoding time by 22.1% on average for depth maps in dependent views. Meanwhile, the proposed method presents only 0.01% BDBR increase on average (usually no damage of video quality in test sequences), which means the fast method can protect the original RD performance very well.

The method from [10] could also achieve great improvement in reducing the computational complexity (23.6% time saving on average). And the average BDBR increase looks not serious as 0.11%. However, achieving a near real-time encoder in 3D-HEVC requires a few number of fast algorithms from different aspects working together. To prevent the accumulated degradation from different methods, it is worth to keep the influence on the coding performance as low as possible for each single fast algorithm. Besides, the degradation of BDBR in [10] is not consistent among all sequences. For example, it results in 0.7% BDBR increase for Poznan_Hall2, which is not negligible. It is because the method in [10] is designed for texture video and does not consider the characteristics of depth coding, where new coding tools like DIS could better present the depth blocks and protect the video quality. By introducing the DIS mode in the fast checking mode list, the proposed method could hold a more stable coding performance for all the test sequences.

V. CONCLUSION

In this paper, with the purpose of reducing the computational complexity of 3D-HEVC encoding process, a fast mode decision algorithm is proposed for depth maps in dependent views by utilizing inter-view correlation and the intermediate RD cost of Merge mode (Skip, 2N×2N) and DIS mode. Simulation results show that the proposed method can reduce the encoding time by 22.1% for dependent depth maps with no coding performance degradation.

REFERENCES

[1] P. Merkle, K. Müller, and T. Wiegand, “3D video: acquisition, coding, and display,” IEEE Trans. Consumer Electronics, vol. 56, no. 2, pp. 946-950, May 2010.

[2] Y. Chen, M. M. Hannuksela, T. Suzuki, S. Hattori. “Overview of the MVC + D 3D video coding standard,” J. Vis. Commun. Image. R., vol. 25, no. 4, pp. 679-688, May 2014.

[3] L. Levkovich-Maslyuk, A. Ignatenko, A, Zhirkov, et al, "Depth image-based representation and compression for static and animated 3·D objects," IEEE Trans, all Circrtits and Systems for Video Technology, vol. 14, no, 7, pp. 1032·1045, July 2004

[4] A. I. Purica, E. G. Mora, B. Pesquet-Popescu, et al. “Multiview plus depth video coding with temporal prediction view synthesis,” IEEE Trans. Circuits Syst. Video Technol., vol. 26, no. 2, pp. 360-374, Feb. 2016.

[5] K. Muller, H. Schwarz, D. Marpe, et al. “3D high-efficiency video coding for multi-view video and depth data,” IEEE Trans. Image Process., vol. 22, no. 9, pp. 3366-3378, Sep. 2013.

[6] J. M. Kang, K. D. Chung, “A fast mode decision using texture information for intra depth coding in 3DVC,” in 19th IEEE Symposium on Computers and Communications (ISCC)., Madeira, Portugal, June 2014, pp.1-5.

[7] Z. Wang, L. Mo, D. Li, T. Xu, “Fast depth map intra prediction algorithm based on edge detection and CU partition,” in 11th International Conference on Wireless Communications., Shanghai, China, Sep 2015, pp.1-5.

[8] Miok. K., Nam. L., and Li. S, “Fast single depth intra mode decision for depth map coding in 3D-HEVC,” Int. Conf. on ICMEW., Turin, Italian, June 2015, pp. 1-6

[9] P. Helle, S. Oudin, B. Bross, et al. “Block merging for quadtree-based partitioning in HEVC,” IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1720-1731, Dec. 2012.

[10] N. Zhang, Y. W. Chen, S. W. Lei,. et al.: “Fast encoder decision for texture coding,” ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-E0173, July 2013.

[11] T. Yu, Y. Zhang, and P. C. Cosman, “Classification based fast mode decision for stereo video coding,” in Proc. 20th IEEE Int. Conf. Image Process., Sep. 2013, pp. 1724–1728.

[12] G. Chi, X. Jin, Q. Dai, “A quad-tree and statistics based fast CU depth decision algorithm for 3D-HEVC,” in Proc. IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW), pp. 1-5, Sep. 2014.

[13] P. Merkle, K. Muller, and T. Wiegand, “Coding of depth signals for 3D video using wedgelet block segmentation with residual adaptation,” in Proc. IEEE Int. Conf. Multimedia and Expo (ICME), pp. 1-6, San Jose, USA, Jul. 2013.

[14] D. Rusanovsky, K. Muller, and A. Vetro, “Common test conditions of 3DV core experiments,” ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, Doc. JCT3V-A1100, Stockholm, Sweden, Jul. 2014.

[15] G. Bjontegaard, “Calculation of average PSNR differences between RD curves,” 2001: ITU-T Video Coding Experts Group (VCEG).

[16] Jianjun Lei, Jinhui Duan, Feng Wu, et al, “Fast Mode Decision Based on Grayscale Similarity and Inter-View Correlation for Depth Map Coding in 3D-HEVC,” IEEE Trans. Circuits Syst. Video Technol., pp. 1–13, 2016. doi: 10.1109/TCSVT.2016.2617332.

[17] Qiuwen Zhang, Ming Chen , Xinpeng Huang, et al, “Low-complexity depth map compression in HEVC-based 3D video coding,” EURASIP Journal on Image and Video Processing, 2015.

[18] N. Zhang, D. Zhao, Y-W Chen, et al, “Fast encoder decision for texture coding in 3D-HEVC,” Signal Processing Image Communication., 2014, 29, (9), pp. 951–961, doi: 10.1016/j. image.2014.06.003