A fast CU Size Decision Algorithm for 3D-HEVC

Yubing Wang, Yongfang Wang[a] and Yawen Shi

Shanghai University, Shanghai, China

Abstract. The emerging 3D-HEVC has achieved the highest coding efficiency but requires a very high computational complexity. To speed up the encoding process for the dependent texture views, we propose a fast CU depth range selection algorithm by jointly making use of the inter-view and temporal-spatial correlations. Firstly, adaptive correlation weights are proposed to predict coding unit(CU) depth range and skip some specific depth levels rarely used in independent view, the previous frame and neighboring CUs. Besides, a new early termination algorithm is proposed to further reduce the coding time. Experimental results demonstrate that the proposed method saves about 56% coding time on average compared to HTM with maintaining the similar video quality.

1 Introduction

With the rapid development of multimedia technology, compared to text, voice, images, video is used more and more abroad. However, the traditional 2D video can't meet people's need, and 3D video is becoming more and more popular. It is well known that the data size of video is huge, so it asks for higher requirement for transferring, storing and playing. Furthermore, 3D video uses more than one camera to shoot the same scene, which leads to increasing a large amount of information.

To improve the coding efficiency of high-definition video, the Joint Collaborative Team on Video Coding (JCT-VC) designs the standard High Efficiency Video Coding (HEVC)[1]. Subsequently, the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) develops HEVC based 3D video coding standard (3D-HEVC), which is for the compression of multi view video plus depth (MVD) format[2]. In 3D-HEVC, similar to HEVC, the mode decision process in HTM is performed using all the possible CU sizes , prediction modes, and coding tools(disparity-compensated prediction (DCP), inter-view motion prediction, backward view synthesis prediction (BVSP)) to find the optimal one with the least rate distortion (RD) cost using Lagrange multiplier, which leads to high computational workloads. It obstructs wide application of 3D-HEVC.

In order to reduce the complexity of video coding, much work has been done to explore the fast algorithms for H.264/AVC. Liu et al.[3] propose block partition algorithm based on image features to reduce the coding complexity. The low complexity mode prediction in [4] is proposed based on the spatial-temporal correlation. The algorithm in [5] uses SKIP mode to early terminate mode decision. Fast algorithms for HEVC inter

prediction mainly aim at the process of coding unit (CU) selection which has high complexity. Method in [6] is proposed to terminate procedures of CU splitting by setting a threshold value based on the RD-cost of the CUs which have already been coded, if the RD-cost of the current CU is less than this threshold ,CU will stop splitting, and a fast CU size decision method based on coding tree pruning is proposed in [7]. Shen et al.[8] propose three early termination methods based on motion homogeneity checking, RD cost checking and SKIP mode checking to skip the procedure of motion estimation on unnecessary small CU sizes. In 3D-HEVC, the maximum depth of the related coding block in dependent view is first used to early terminate the CU splitting in [10]. A fast mode decision algorithm based on variable size CU and disparity estimation in [11] is proposed to reduce 3D-HEVC computational complexity. The texture quad-tree initialization (QTI) and depth quad-tree limitation (QTL) coding tools and their associated predictive coding (PC) algorithm by utilizing the correlations between the quad-tree of the texture and its associated depth is proposed in [12], which was adopted by the 3D-HEVC working draft. Zhang et al.[13] speed up the encoding process of dependent texture views based on inter-view correlation, which uses an early merge mode decision algorithm and an early CU splitting termination algorithm. CU depth correlation between the independent view and the dependent view is also studied to accelerate CU splitting process of dependent views in [14].

The aforementioned methods for 3D video coding only use view correlation or temporal-spatial correlation to predict the depth of current coding block, while not jointly considering these correlations. Consequently, in this paper, we comprehensively exploited the correlations between views , and spatial-temporal

---

[b] Corresponding author: yfw@shu.edu.cn

correlation, and then proposed a fast CU size decision algorithm based on adaptive weights called depth range selection algorithm and an early termination algorithm. Experimental results show that the proposed algorithm saves an average of 56% time compared to the original algorithm of HTM with negligible loss in compression efficiency.

The rest of the paper is organized as follows. Section 2 describes the proposed algorithm in detail. Experimental results are shown in Section 3. Conclusions are given in Section 4

2 Proposed fast coding algorithm

2.1 CU depth splitting analysis for texture views

3D-HEVC inheriting the same quad-tree structure as HEVC, which must traverse from depth level 0 to depth level 3, hugely increases the coding complexity of 3D-HEVC[15]. Zhang et.al[13] made a statistical analysis about the probability of all four splitting depth levels of dependent texture view for each test sequence. The result shows that 76.2%, 1%, 3.9% and 1.4% of treeblocks choose the depth level "0", "1", "2" and "3", respectively. It demonstrates that small depth levels are always selected, so selecting suitable CU depth range is an effective method to save coding time.

In HTM15.0, we make a statistic of the encoding time of texture video and depth video, the result of which is shown in Figure 1. Every test sequence contains three views, and every view includes texture video and depth video. Figure 1.(a) shows that the encoding time ratio is 75% for texture pictures and 25% for depth maps. Figure 1.(b) shows the encoding time ratio of three texture views, it can be seen that View0 accounts for about 17%, while View1 and View2 take up about 40% respectively, which demonstrates that the two dependent textures (View1 and View2) occupy most of the encoding time in 3D-HEVC. Accordingly, the proposed fast algorithm in this paper is only for the two dependent texture views.
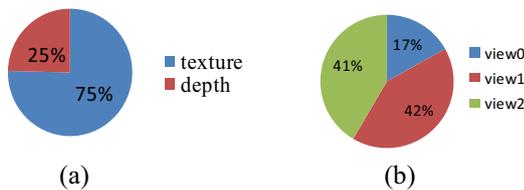
Figure 1. Analysis of coding complexity:(a) Encoding time comparison between texture and depth maps;(b) Proportion of encoding time of three views

2.2 The depth range selection algorithm

Many previous fast CU decision algorithms use spatial-temporal correlations to predict the depth range of current coding block, but most of them adopt fixed weights that can't adapt to the degree of correlation, which may lead to imprecise prediction. Based on the characteristics of 3D video, we propose a depth range

selection algorithm based on adaptive weights using split complexity(SC), which is defined as follows:

$$
SC\_i = \begin{cases} \dfrac{1}{256} \cdot \sum_{j=0}^{255} d_j & depth_{max}=0,1,2 \\[2mm] \dfrac{1}{128} \cdot \sum_{j=0}^{255} d_j & depth_{max}=3 \end{cases} \quad (1)
$$

Where d_j is the value of depth level, $depth_{max}$ is the maximum value of depth level. The corresponding blocks are shown in Figure 2. i indicates the type of correlated CU, for example, SC_Col is expressed as the split complexity of inter-view correlated CU. To avoid getting the same SC when depth levels are different, unlike the method used in [15], which directly averages the depth value, we adopt different definitions according to the maximum depth level.
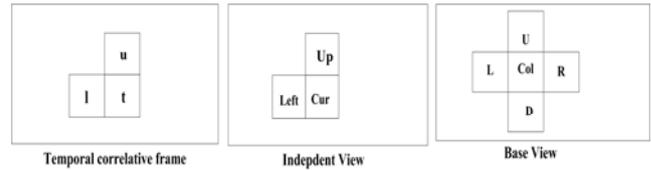
Figure 2. inter-view, spatial, and temporal correlated CUs. Cur: current CU; Left: Left CU; Up: upper CU; t: the co-located CU in the previously frame; Col: inter-view correlated CU.

Experimental result in [15] demonstrates that the spatial-temporal correlation is connected with temporal average depth error. The larger the value is, the stronger the temporal correlation is, and the weaker the spatial correlation is. We evaluate the degree of correlation by using temporal split complexity error (TSCE) and inter-view spatial split complexity error (ISCE) which are defined as follows:

$$
TSCE = 1/2 \cdot (|SC\_Left - SC\_l| + |SC\_Up - SC\_u|) \quad (2)
$$
$$
ISCE = 1/2 \cdot (|SC\_Col - SC\_L| + |SC\_Col - SC\_U|)
$$

The SC of current CU is predicted using spatial neighboring CU (Left, Up in Figure 2), the co-located CU (t in Figure 2) and inter-view correlated CU (Col in Figure 2) as follows.

$$
SC\_pre = w_{col} \cdot SC\_col + w_l \cdot SC\_Left + w_u \cdot SC\_Up + w_t \cdot SC\_t \quad (3)
$$

Where SC_pre is predicted SC of the current coding block, w_{col}, w_l, w_u, and w_t are weights, which are obtained in (4) from normalizing the W_{col}, W_l, W_u, W_t respectively in (5).

$$
w_{col} = \frac{W_l \cdot W_u \cdot W_t}{W_{col} \cdot W_l \cdot W_u + W_l \cdot W_u \cdot W_t + W_{col} \cdot W_u \cdot W_t + W_{col} \cdot W_l \cdot W_t}
$$

$$
w_l = \frac{W_{col} \cdot W_u \cdot W_t}{W_{col} \cdot W_l \cdot W_u + W_l \cdot W_u \cdot W_t + W_{col} \cdot W_u \cdot W_t + W_{col} \cdot W_l \cdot W_t} \quad (4)
$$

$$
w_u = \frac{W_{col} \cdot W_l \cdot W_t}{W_{col} \cdot W_l \cdot W_u + W_l \cdot W_u \cdot W_t + W_{col} \cdot W_u \cdot W_t + W_{col} \cdot W_l \cdot W_t}
$$

$$
w_t = \frac{W_{col} \cdot W_l \cdot W_u}{W_{col} \cdot W_l \cdot W_u + W_l \cdot W_u \cdot W_t + W_{col} \cdot W_u \cdot W_t + W_{col} \cdot W_l \cdot W_t}
$$

Where W_{col}, W_l, W_u, and W_t are calculated by jointly utilizing the temporal-spatial and inter-view correlations as follows:

$$W_{col} = 1.2$$
$$W_l = 1 + 0.2 \cdot ISCE$$
$$W_u = 1 + 0.2 \cdot ISCE \qquad (5)$$
$$W_t = 1 + 0.2 \cdot TSCE$$

In addition, in order to take full advantage of inter-view correlation, we define a Correlated Split Complexity (CSC) in (6).

$$CSC = \max\{SC\_col, SC\_L, SC\_U, SC\_R, SC\_D\} \qquad (6)$$

According to SC_pre and CSC, we can predict depth range of current LCU as follows:

If CSC<Th, DR is [0,1], otherwise do as follows:

1) When $SC\_pre \leqslant$ th1, DR is [0,1];

2) When th1<$SC\_pre \leqslant$ th2, DR is [0,2];

3) When th2<$SC\_pre \leqslant$ th3, DR is [1,3];

4) When SC_pre>th3, DR is [2,3].

Where Th,th1,th2,th3 are thresholds, and they are set as 3.0,1.0,2.0 ,5.0 according to extensive experiments.

To verify the effectiveness of the proposed adjusted depth range, four sequences with different activities are tested in Table 1 and Table 2. It is observed from Table I that when Th=3.5, the probability of selecting depth level "0" and "1" can be achieved about 98%. When th1, th2, th3, th4 are set as 3.0,1.0, 2.0, 5.5, we can see from Table II that the average prediction accuracy can be 92%-97%.

Table 1. Probability of selecting low depth level for small CSC.

Sequences	25	30	35	40
Ballons	98%	99%	99%	99%
Newspaper	97%	97%	98%	99%
Poznan_Street	97%	97%	98%	98%
GT_Fly	98%	99%	99%	99%
Average	97%	98%	98%	99%

Table 2. Accuracy of depth range selection

Sequences	25	30	35	40
Ballons	93%	95%	96%	98%
Newspaper	93%	94%	94%	96%
Poznan_Street	90%	93%	95%	96%
GT_Fly	93%	95%	96%	97%
Average	92%	94%	95%	97%

2.3 Early termination(ET) of CU

An early CU splitting termination is proposed to make a more accurate prediction about the splitting depth level of current CU by considering both inter-view correlation and spatial-temporal correlations.

For DR selection algorithm , the size of predicting CU is too large to predict accurately, so when early terminating the CU splitting, the size of referring CUs are 8x8, which locate in the left and top of current CU, and the maximum depth is named as Depth_neighbour. Besides, we also consider the depth level of corresponding block in forward reference frame, backward reference frame, and independent view of

current CU, which are named as Depth_t_0, Depth_t_1, Depth_col respectively.

If the following two conditions are true, early CU splitting termination is performed for the current CU.

1)
$$Depth\_\max = \max\{Depth\_neighbour, Depth\_col, Depth\_t_0, Depth\_t_1\}$$
$$Depth\_\max \leq uidepth$$

2) Skip mode is selected as the best prediction mode for the current CU.

Where $Depth\_max$ is the maximum value of all the depth levels, $uidepth$ is the splitting depth level of the current CU.

2.4 Overall algorithm

Based on the aforementioned analysis, the proposed overall algorithm incorporates the depth range selection algorithm and the early termination method. A flowchart of the proposed algorithm is shown in Figure 3. First, calculate SC_pre and CSC; Next, predict the depth range of current CU according to CSC and SC_pre; Last, check the early termination conditions of current CU, and if the conditions are met, CU splitting ends, otherwise, the depth level increases by 1.
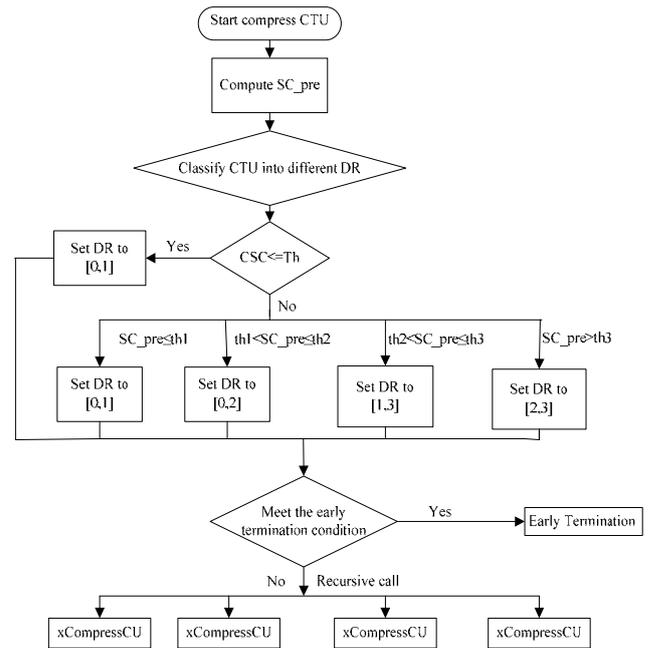
Figure 3. Flowchart of the proposed overall algorithm

3 Experimental results

To evaluate the performance of the proposed algorithm, it is implemented on HTM15.0 Test conditions are set as follows: 7 standard sequences (3 texture plus 3 depth views), four QPs (25/34, 30/39, 35/42, 40/45), QTLPC enabled, fast decision for Merge RD Cost (FDM) enabled, both fast merge mode decision and early CU determination for texture component of dependent view disabled.

Performance of the proposed algorithm is measured by BD-rate and ∆EncT. BD-rate evaluates the coding gains correspond to bitrate reductions, positive and

negative values represent increments and decrements respectively. ΔEncT shows the encoding time reduction compared with HTM15.0.

we compare the performance of the proposed algorithm called "DR&ET" and the fast encoder decision algorithm for texture coding called "EMD&ECUST"[13] with HTM15.0. The experimental results are presented in Tables III–IV. The "video1" and "video2" columns show the BD-rate of the two dependent texture views. The "Video PSNR/video bitrate" column shows the BD-rate performance considering Y-PSNR of the coded texture views over the bitrate of texture data.

As can be seen from Table 3, the proposed method can reduce the entire encoding time by 56% on average. Meanwhile, the loss of BD-rate in dependent views is 0.52% and 0.57%. Therefore, the proposed DR&ET can efficiently reduce encoding time with a little loss of RD performance. The performance of "EMD&ECUST" is so good that it has been adopted to 3D-HEVC reference software, but our proposed DR&ET can reduce more time than it, and the BD-rate seen in "Video PSNR/video bitrate" merely increases about 0.5%.

Table 3. Encoding performance compared with HTM15.0

Sequence	video 1	video 2	Video PSNR/ video bitrate	ΔEncT
Balloons	0.45%	0.80%	0.39%	58%
Kendo	0.97%	0.72%	0.73 %	52%
Newspaper	0.78%	0.68%	0.55%	58%
GhostTownFly	-0.15%	-1.40%	0.41%	51%
PoznanHall2	0.85%	1.41%	1.22%	63%
PoznanStreet	0.33%	1.00%	0.64%	55%
UndoDancer	0.40%	0.74%	0.88%	52%
average	0.52%	0.57%	0.69%	56%

Table 4. Encoding performance of EMD&ECUST compared with HTM15.0

Sequence	video1	video 2	Vide PSNR/ video bitrate	ΔEncT
Balloons	-0.67%	-0.29%	-0.10%	56%
Kendo	0.11%	-0.47%	0.15%	48%
Newspaper	-0.11%	-0.37%	0.22%	52%
GhostTownFly	-1.36%	-0.73%	0.05%	46%
PoznanHall2	-0.43%	0.52%	0.60%	52%
PoznanStreet	-0.62%	-0.57%	0.12%	48%
UndoDancer	-1.07%	-0.54%	0.28%	45%
average	-0.59%	-0.24%	0.19%	48%

Figure 4 (a) and (b) illustrates an example of RD and time saving curves under four different QPs compared to HTM15.0. As shown in Fig.4, the proposed method can reduce more coding time with maintaining the RD performance than that of EMD&ECUST. In addition, according to Fig.4(b), we can see that when QP is low, DR&ET is more effective than EMD&ECUST, while as the increase of QP, the performance of both algorithms will tend to be similar. In general, the proposed algorithm is better performance than EMD&ECUST.

(a)

(b)
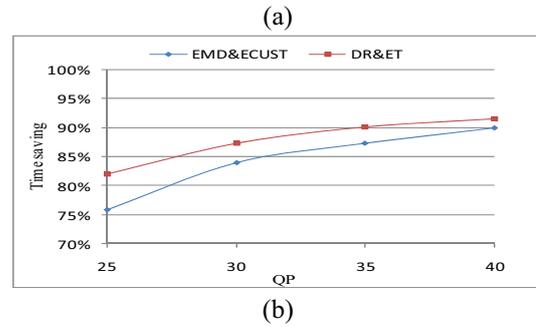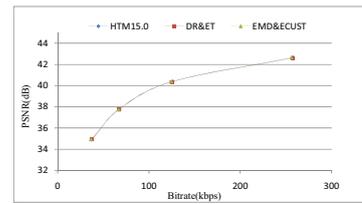
Figure 4. Experimental results of "Ballons" in video 1 under different QP settings. (a) RD curves of "Ballons" in video 1. (b) Time saving curves of "Ballons" in video 1.

4 Conclusions

In this paper, we propose a fast CU depth range selection algorithm and an early termination CU algorithm for 3D-HEVC. Experimental results show that the proposed algorithm can significantly reduce the computational complexity with negligible loss of BD-rate. And this algorithm will make 3D-HEVC more available for real-time applications.

Acknowledgments

References

1. Joint Collaborative Team–Video Coding, "Test Model under Consideration", JCTVC-A205 output document, Dresden, Germany, 15-23 April, 2010.
2. Ohm J R, Overview of 3D video coding standardization[C], International conference on 3D systems and applications, Osaka. 2013.
3. Liu Z, Shen L, Zhang Z, An efficient inter mode decision algorithm based on motion homogeneity for H. 264/AVC[J], Circuits and Systems for Video Technology, IEEE Transactions on, IEEE, 2009.
4. Zhao T, Wang H, et al, Fast mode decision based on mode adaptation[J], Circuits and Systems for Video Technology, IEEE Transactions on, IEEE ,2010.
5. Avishek Saha, Kallol Mallick, et al, SKIP prediction for fast rate distortion optimization in H.264, Consumer Electronics, IEEE Transactions on 53.3 (2007): 1153-1160.

6. Tan H L, Liu F, Tan Y H, et al, On fast coding tree block and mode decision for high-efficiency video coding (HEVC)[C], Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, IEEE.

7. Choi K, Jang E S, Fast coding unit decision method based on coding tree pruning for high efficiency video coding[J], Optical Engineering, 2012, 51(3): 030502-1-030502-3.

8. Shen L, Liu Z, Zhang X, et al, An effective CU size decision method for HEVC encoders[J]. Multimedia, IEEE Transactions on, IEEE, 2013.

9. Wu W, Song B, Just-noticeable-distortion-based fast coding unit size decision algorithm for high efficiency video coding[J], Electronics Letters, 2014, 50(6): 443-444.

10. Thaísa Leal da Silva, Luis A. da Silva Cruz, Inter-view prediction of coding tree depth for HEVC-based multiview video coding[C], Electronics, Circuits, and Systems (ICECS), 2013 IEEE 20th International Conference on, IEEE, 2013: 165-168.

11. Zhang Q, Li N, Gan Y, Low Complexity Mode Decision for 3D-HEVC[J], The Scientific World Journal, 2014.

12. Mora E G, Jung J, Cagnazzo M, et al, Initialization, limitation and predictive coding of the depth and texture quadtree in 3D-HEVC[J], IEEE Transactions on Circuits and Systems for Video Technology, 2014.

13. Zhang N, Zhao D, Chen Y W, et al, Fast encoder decision for texture coding in 3D-HEVC[J], Signal Processing: Image Communication, 2014, 29(9): 951-961.

14. Chi G, Jin X, Dai Q, A quad-tree and statistics based fast CU depth decision algorithm for 3D-HEVC[C], Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on. IEEE, 2014: 1-5.

15. Fang Shuqing, Yu Mei, Chen Fen, Jiang Gangyi, et al, Fast Adaptive Coding Unit Depth Range Selection Algorithm for High Efficiency Video Coding[J], Sensors & Transducers, 2014, 183(12): 1726-5479.