

# Effective Intra Mode Prediction of 3D-HEVC System based on Big Data Clustering and Data Mining

Jinchao Zhao, Shuaichao Wei, and Qiuwen Zhang<sup>\*</sup>

*College of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, 450002, China*

---

## Abstract

3D-high efficiency video coding (3D-HEVC) structure is a development of HEVC, but some new coding techniques are added on the basis of it to make it more conducive to encoding depth maps and multi-view. In 3D-HEVC, the intra prediction mode decision for depth level contains closed connections with coding unit (CU) partition. This process, quad-tree block splitting, gives the gorgeous coding efficiency to 3D-HEVC, but it incurs unacceptable computational burdens because each possible coding mode is tested to rank the most suitable one. According to previous works, whether the current CU will be divided into smaller sizes is dependent on encoding contexts. In view of that, this paper proposed a novel method to speed up intra coding unit splitting, relying on data clustering and data mining. The experimental results showed that our new approach can reach a satisfied balance between computational burdens and RD cost.

*Keywords:* 3D-HEVC system; data clustering; data mining

(Submitted on September 15, 2019; Revised on October 16, 2019; Accepted on November 25, 2019)

© 2019 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

3D-HEVC was developed by the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) in order to accelerate coding efficiency [1]. Its code structure is similar to MVC, and its structure is shown in Figure 1. In 3D-HEVC, all the input video images and depth images in the figure are scenes at the same time and at every position, and these images constitute an access layer. In the same access layer, the independent viewpoint and the reference viewpoint are first encoded, followed by the depth map of the viewpoint, and then the video image and the depth map of other viewpoints are encoded. In principle, the images of each viewpoint, including video images and depth images, can be encoded by using the HEVC coding framework, and all the input bit streams can be combined to form a 3D bit stream. The depth maps are the symbols of the distance between the camera and the object.

There exists a process called coding tree unit (CTU) in 3D-HEVC, which employs nimble dividing ways for coding units. In this stage, every CTU will experience rate-distortion optimization (RDO) so as to attain a fittable partition structure, where many combination possibilities will be considered. This greatly contributes to coding efficiency with almost the same increment of computation of other standards at the same time.

Zhang et al. studied 3D-HEVC depth encoding in order to reduce the time of choosing optimal CU splitting sizes by employing the good feature corner point [2]. Peng et al. set a special condition for omitting some useless processes of small sizes if the conditions are met [3]. In [4], Saldanha employed the edge detector to realize the shrinkage of the useless DMM process. Peng et al. offered a method that utilizes the prediction modes' RD cost to generate a threshold [5]. The algorithm attains the differences of the CU and the largest sub-blocks' deviation. The stage of partition will be executed when either of two conditions is met: the threshold is lower than the difference between the CUs, or the CU difference is lower than the largest sub-blocks' difference. In [6], Zhang et al. found the closed connections between a quadtree depth limit relying on QP and determined whether the process of partition for the current CU should be executed and whether the smaller blocks' information is involved.

---

<sup>\*</sup> Corresponding author.

E-mail address: [zhangqwen@126.com](mailto:zhangqwen@126.com)

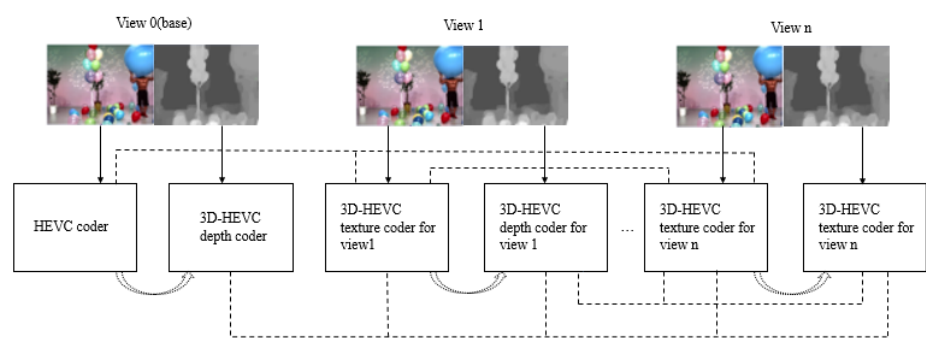


Figure 1. The structure of 3D-HEVC

In this paper, we employ the process of data mining to determine whether every CU needs to be partitioned into smaller sizes and whether the full RD cost calculation can be omitted. Then, we make use of tremendous data to generate a consolidated model decision based on AM-PCM. Finally, this model is used to speed up intra coding for texture intra coding.

The rest of the paper is arranged as follows. Section 2 describes the content of our algorithm. Experimental results and comparisons are detailed in Section 3. Then, Section 4 presents the conclusions.

2. Proposed Method

2.1. Data Mining for Depth Map Coding

To reduce the time of 3D encoding, 3D-HEVC has been developed by the Joint Collaborative Team. Multi-view video plus depth (MVD) [7] is an attached characteristic provided by 3D-HEVC, with each texture frame connected to the depth map. The depth map, which represents the distance from object to camera, is also employed to attain virtual views by using synthesis view techniques during decoding. Because depth map encoding occurs in the 3D encoder, the sum of computation is unacceptable.

In 3D-HEVC, there exists a huge difference in computation allocation between depth maps and texture maps. The calculation of depth level coding outdistances the texture. At the coding stage of the depth map, DMMs, DIS, and SDC assessments are employed. In the process of the intra prediction of depth maps, coding tree units (CTUs) are attained by splitting every frame, and every CTU is fragment into coding units (CUs). The range of the size is from  $4 \times 4$  to  $64 \times 64$ .

In the 3D-HEVC test model (3D-HTM), the partition structure of each CTU is selected by a rate-distortion optimization (RDO) method, which combines multiple coding structures (block partitioning and prediction mode). It can evaluate to find the best coding possibilities. Compared to previous standards, this achieves very high coding efficiency at the expense of greatly expanding the computational burden of the encoder.

In order to analyze 3D-HEVC encoder, these experiments use 3D-HTM version 16.0 and common test conditions (CTC) for All-Intra encoder configurations [8].

Table 1 shows the result of complexity allocation for texture and depth maps in four different quantization parameter-pairs' (QP-pair) values. We can see that the largest proportion of texture video is only 16.2%, which is far less than the complexity of depth map encoding.

Table 1. The complexity distribution for four QP-pair values in texture and depth maps

Maps	Four QP-pair Values			
	25/34	30/39	35/42	40/45
Depth	83.8%	84.3%	84.1%	84.2%
Texture	16.2%	15.7%	15.9%	15.8%

This phenomenon occurs because HEVC intra prediction is employed during the process of texture coding, while depth mapping coding is also calculated using DMMs, DIS, and SDC. This proves that the average time of depth map encoding is about 5.2 times that of texture map encoding. Thus, whether the depth encoding calculation complexity can be reduced is the bottleneck for 3D-HEVC.

Table 2 represents the allocation of CUs for four different QP-depth values and shows that changes in QP-depth result in different CU size distributions. The distribution of CU size of each code block is not uniform. It appears because the QP-depth value defines the compression ratio and acts on the quality of images directly. This phenomenon is caused by the QP-depth value, which plays a crucial role in image quality.

Table 2. The CU size for four QP-depth values for depth encoding

CU Size	Four QP-depth values			
	34	39	42	45
8×8	66%	43%	22%	5%
16×16	21%	25%	23%	16%
32×32	10%	19%	27%	28%
64×64	3%	13%	28%	51%

The relationship of image quality and QP value is an inverse ratio. The homogeneous region corresponds to a large QP value; in this condition, the image details are restrained, and employing a larger CU size will positively affect the efficiency. On the contrary, the complex region corresponds to a small QP value, and the efficiency can be sustained by choosing a smaller CU size.

In Table 2, 34 is the lowest QP-depth value. Under the condition of this QP depth value, 66% of the CUs sizes are encoded as 8×8, and only 3% are encoded as 64×64. For the largest QP value, which is 45, 5% of the CUs size are encoded as 8×8, and 51% are encoded as 64×64. In view of this, we can generate a way to determine whether the current CU needs to be divided again or not, and then the amount of RD cost calculation can be lessened.

During the encoding process, the encoder makes many other decisions to determine the coding efficiency. Therefore, it is necessary to consider the coding context, not just the RDO, and to statically set the solution of these decisions with little impact on the efficiency of the encoder.

Data mining can find the connections in dependent variable values so as to identify the regularity of the data set and build generalizations in the properties. Decision trees are usually employed when tiny complexity and pinpoint accuracy are needed. This paper introduces data mining so as to lessen the calculation complexity in depth map encoding.

During the process of data mining, the video sequence is encoded in all-intra configuration, taking into account all CTC QP values. At each evaluation, the size of the CTU is defined as 16×16, 32×32, and 64×64 pixels. It should be emphasized that the number of depth maps of 3D video sequences capable of 3D video coding experiments is limited. Therefore, in order to extract the required data from the offline training process, the Kendo video sequence is randomly selected from the CTC data set. In order to avoid overtraining, only one video sequence is used during the training.

## 2.2. AM-PCM Algorithm and Feature Selection for Texture Intra Coding

In 3D-HEVC, every coding tree unit (CTU) can be recursively split into a quadtree structure including prediction unit (PU) and coding unit (CU). The process of CU and PU is described in Figures 2 and 3, respectively.

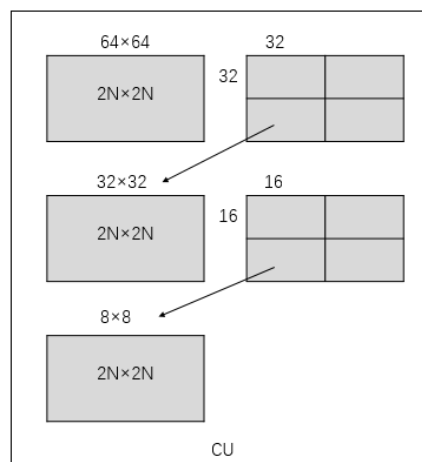


Figure 2. Process of coding unit

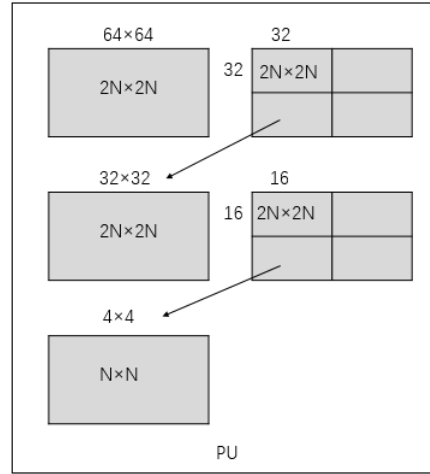


Figure 3. Process of prediction unit

Since the encoder evaluates the rate-distortion (RD) performance of each intra mode and recursively finds the optimal code rate in the process of encoding in the 3D-HEVC coding, the recursive partitioning process is too time-consuming. Although this evaluation has further improved compression efficiency and synthesis quality, it will lead to greater computational complexity in the encoding process.

In the encoding process, the selective training data is processed by automatic merging possibilistic clustering (AM-PCM), and then a unified decision model is generated on the massive data. Then, the obtained model is used to speed up the 3D-HEVC intra-frame coding.

Clustering analysis relies on similarities, and there exist more similarities between the patterns in a cluster than between the patterns in the alike cluster. AM-PCM belongs to clustering, and it was developed by possibilistic C-means (PCM) clustering [9].

The stages of again and again partition for CU and PU are used to select the optimal sizes, but this leads to unacceptable time consumption. Therefore, we introduce AM-PCM to generate a model to shrink the complexity in 3D-HEVC intra coding.

Let  $M = \{m_1, m_2, \dots, m_p\}$  be a set of  $M$  data.  $W = [w_1, w_2, \dots, w_Q]^T = [\mu_{ij}]_{P \times Q}$  is a matrix, and  $B = \{b_1, b_2, \dots, b_C\}$  is a cluster. We define the objective function as follows:

$$G_{AM-PCM}(W, B) = \sum_{i=1}^Q \sum_{j=1}^P \mu_{ij} [l_{ij}^2 - \lambda (1 - \frac{r}{1+r} \mu_{ij}^{\frac{1}{r}})], \quad r > 1 \quad (1)$$

From Equation (1), we obtain

$$\mu_{ij} = (\frac{\lambda - l_{ij}^2}{\lambda})^r, \quad b_i = \frac{\sum_{j=1}^P \mu_{ij} m_j}{\mu_{ij}} \quad (2)$$

We define  $d^2(y_i, y_j)$  as the Euclidian distance from  $y_i$  to  $y_j$ .

$$\begin{aligned} \Psi &= \{d^2(y_1, y_1), \dots, d^2(y_1, y_p), \dots, d^2(y_{p-1}, y_p)\} \\ \Delta &= \text{sorted}(\Psi) = (d_{(0)}, d_{(1)}, \dots, d_{(P(P-1)/2)}) \\ u &= \frac{1}{P-1} [P \cdot \sum_{i=1}^Q (\sum_{j=1}^P \mu_{i,j} / \sum_{i=1}^Q \sum_{j=1}^P \mu_{i,j})^2 - 1] \end{aligned} \quad (3)$$

There is a strong correlation between CU size and the homogeneity of screen videos. In the process of 3D-HEVC coding, the CUs allocated in simple regions usually choose a larger size to encode, and the CUs allocated in complex regions usually choose a smaller size to encode for efficiency. Therefore, we extract some features that can be used to find the CUs in simple regions so as to judge the size of predicting the CU.

We define the diff and amplitude of simplified mass center vector (ASMCV) from [10] and [11] as the extracted features.  $Y(i, j)$  is the luminance of the pixel located at  $(i, j)$  in a  $64 \times 64$  CU, and the diff, ASMCV, and amplitude are defined by Equations (4) to (6) as follows:

$$Diff = \frac{1}{64 \times 64} \sum_{i=0}^{63} \sum_{j=0}^{63} [Y(i, j)]^2 - \left[ \frac{1}{64 \times 64} \sum_{i=0}^{63} \sum_{j=0}^{63} Y(i, j) \right]^2 \quad (4)$$

$$\begin{aligned} l_{m_{i,j}} &= \sum_{j=0}^{64/2-1} \sum_{i=0}^{63} \left( \frac{64}{2} - j \right) \times |Y(i, j) - Y(i, 64 - j - 1)| \\ l_{n_{i,j}} &= \sum_{j=0}^{64/2-1} \sum_{i=0}^{63} \left( \frac{64}{2} - j \right) \times |Y(i, j) - Y(64 - i - 1, j)| \\ ASMCV &= l_{m_{i,j}} + l_{n_{i,j}} \end{aligned} \quad (5)$$

$$Amp = \sqrt{ASMCV^2 + Diff^2} \quad (6)$$

The range of CU size in 3D-HEVC depth map encoding for intra mode is varied from  $8 \times 8$  to  $64 \times 64$ . In this paper, we focus on three static decision trees in order to determine whether the  $16 \times 16$ ,  $32 \times 32$ , and  $64 \times 64$  CUs need to be divided again. The video sequence of Kendo in the stage of data mining is encoded using intra mode, relying on overall values of CTC QP. In the section of assessment for each one, the size of CTU is confined to three types of pixels:  $64 \times 64$ ,  $32 \times 32$ , and  $16 \times 16$ . We will save the encoded CU data, and the result of whether the CU should be partitioned again will be obtained from the information.

There is a pivotal truth that not all 3D video sequences and their depth maps are suitable for experiments. Therefore, we stochastically select Kendo sequences in all CTC data sets to abstract needed data for the stage of offline training. We employ one sequence. However, in order to prove that the trained process can attain high quality for variability encoding scenarios, we test whole video sequences that are defined in the CTCs.

At the stage of training for every decision tree, we employ the J48 algorithm with the Waikato environment for knowledge analysis (WEKA) [12]. The algorithm of J48 is an application of the C4.5 algorithm. With the purpose of choosing an optimal method for data balancing, we divide the input data into two kinds of data that have same sizes, including two types of inputs: leading to split and not split for CUs. In addition, in order to avoid overfitting of the training data, the process of reduced error pruning (REP) is executed. This not only improves the generalization, but also alleviates the depth of the decision tree.

For the  $32 \times 32$  CU size, the types of five decision levels are chosen. Meanwhile, eight decision levels will be selected for CUs of size  $16 \times 16$ . By calculating the information entropy and information gain, the branches of the decision tree are determined.

The introduction of the decision tree algorithm will not generate additional computational pressure on 3D-HEVC coding [13-15], because the process of training data is performed offline and only executed once when the static tree is defined. In addition, the analysis of the decision tree is also easily obtained through calculation.

For the intra mode decision for texture video, the analysis of big data is introduced. We choose video sequences including "GT-Fly," "Undodancer," "Newspaper," "Ponznan-Hall2," "PonznanStreet," "Balloons," "Shark," and "Kendo" as the training data. All these video sequences are composed in JCT-3V. For each period intra coding, there are 24 frames, two features are composed of diff (from Equation (3)), and the ASMCV (from Equation (4)) of the first texture frame is calculated for all CUs. Then, we sort them, where the total number is 100,520 based on Amp (from Equation (5)). The feature vectors whose frequency is 25 are selected to be used for the input of AM-PCM training from the ordered data. The sum of these vectors is 4,000, and all possible contexts are included.

The purpose of this process is to reduce the computational burdens of texture intra coding relying on the analysis of big data. In order to minimize the Euclidean distance, every factor of big data is assigned to its own central cluster, in which large data sets and central clusters are combined. The main size of modes in every center cluster will be decided with the occupancy ratio of size modes, where the ratios of size modes higher than 10% will be regarded as main size modes, and the mode sizes lower than 10% will be ignored. Clusters with the alike main size pattern are regrouped into five regions. From Table 3, we can see the results during the stage of regrouping.

Table 3. The distribution for different QPs in every cluster

Index	QP	Size of coding unit				
		64x64	32x32	16x16	8x8	4x4
1	25	96.50%	3.50%	0.00%	0.00%	0.00%
	30	97.60%	2.37%	0.03%	0.00%	0.00%
	35	96.26%	3.74%	0.00%	0.00%	0.00%
	40	93.50%	6.12%	0.35%	0.03%	0.00%
Average		95.96%	3.93%	0.10%	0.01%	0.00%
2	25	52.29%	41.07%	6.04%	0.50%	0.10%
	30	67.44%	28.21%	3.82%	0.07%	0.10%
	35	72.19%	23.45%	3.57%	0.69%	0.10%
	40	70.39%	24.85%	4.08%	0.61%	0.08%
Average		65.58%	29.40%	4.38%	0.47%	0.10%

### 3. Experimental Results

For the purpose of evaluating the performance of our method, we introduce three decision trees to 3D-HTM (version 16.0). The CTC is also assessed in all-intra encoder. Considering the comprehensive quality of views, we reduce the complexity of the entire 3D-HEVC encoder (texture and depth mapping) and the percentage of unregimented decisions for every decision tree and the solution Bjontegaard Delta-rate (BD-rate). The detailed results are illustrated in Table 4.

The proposed method can reduce the computational complexity by about 52.3% (from 38.1% to 65.2%) on average. For the encoding of depth mapping, the calculation complexity can be decreased by about 59.0%. The increment of BD-rate is very small, with an average of 0.18 % (from 0.03% to 0.64%). From the results of the experiment, we can see that the larger the QP-depth value, the higher the probability of not splitting. This information proves that our method attains good effects on the reduction of calculation complexity by omitting the stage of processing small size CUs with tiny obstruction for the efficiency of RD cost.

Table 4. The results of CTC assessment

Video	Our method					Literature [5]		Literature [6]	
	CUs not splitting			BD-rate	Complexity reduction	BD-rate	Complexity reduction	BD-rate	Complexity reduction
	16x16	32x32	64x64						
Balloons	49.6%	47.9%	52.3%	0.15%	46.6%	1.11%	26.3%	0.28%	45.1%
Kendo	46.5%	56.2%	56.8%	0.18%	47.3%	0.92%	27.5%	0.29%	38.0%
Newspaper_CC	27.4%	42.3%	29.6%	0.13%	38.1%	1.32%	28.9%	0.22%	37.5%
GT_Fly	31.2%	47.1%	71.9%	0.09%	57.9%	0.44%	44.3%	0.21%	46.2%
Poznan_Hall2	36.7%	63.5%	86.3%	0.64%	65.2%	1.26%	45.8%	0.35%	47.3%
Poznan_Street	27.1%	47.2%	55.9%	0.11%	54.9%	1.33%	47.9%	1.14%	41.2%
Undo_Dancer	39.6%	54.4%	68.2%	0.16%	58.1%	0.55%	48.2%	1.02%	37.4%
Shark	30.7%	46.2%	63.8%	0.03%	50.3%	0.26%	34.1%	0.50%	42.7%
Average	36.1%	50.6%	60.6%	0.18%	52.3%	0.90%	37.9%	0.50%	41.9%

It is worth emphasizing that the stage of offline training and the creation of decision trees use only the video sequence of Kendo. In addition to the video sequence described above, the residual test sequences are still tested, and the results obtained are hopeful in terms of reduced complexity and BD-rate degradation. Overall, the information provides evidence that the decision tree can positively impact experimental analysis.

In addition, Figure 4 reveals a negligible PSNR loss and bit improvement compared with the original encoder system. We can observe that the proposed overall method can achieve consistent runtime saving from low to high bitrate ranges with similar RD performance compared with the original 3D-HEVC system.

We can see that the algorithm achieves a decrease in complexity of 37.9%, and the augmentation of the BD-rate is 0.90% in [5] from Table 4. For the method in [6], the reduction of calculation is about 41.9%, with a 0.50% BD-rate increment. Therefore, our method has a better effect on depth map encoding than related works.

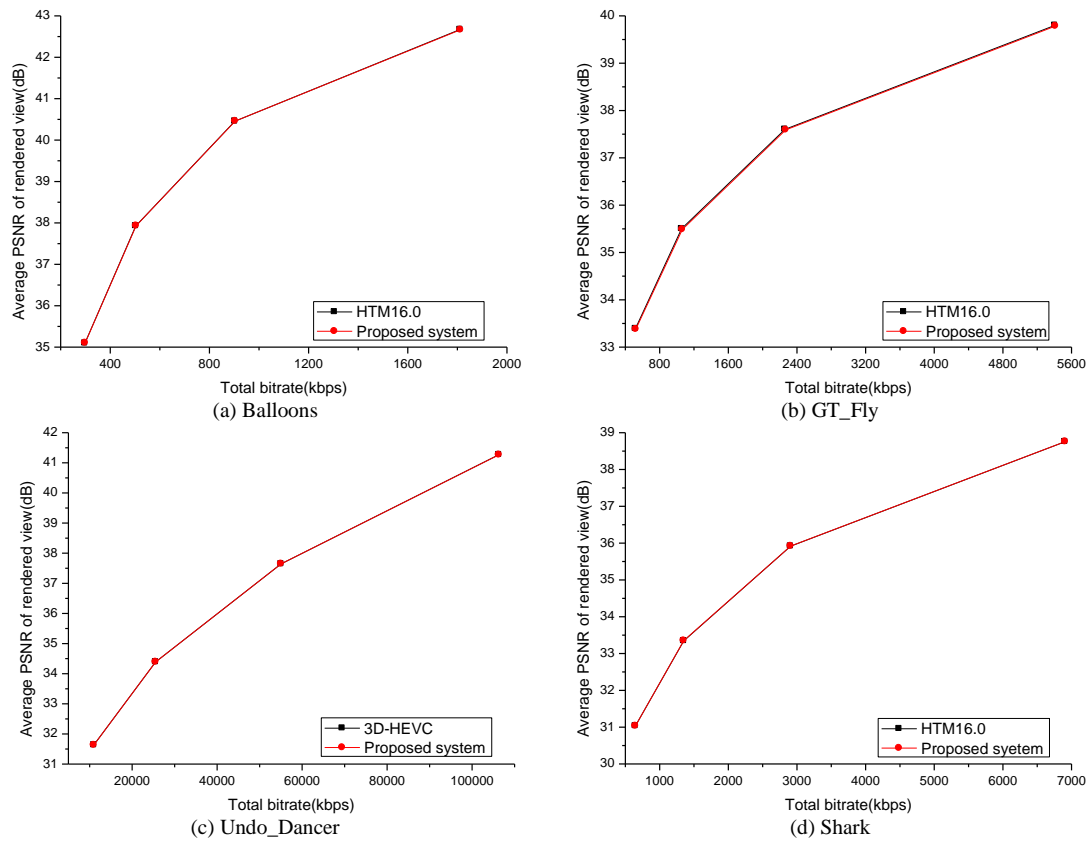


Figure 4. System coding performances of the test videos

#### 4. Conclusions

In our paper, we offered fast 3D-HEVC intra mode prediction that relies on big data clustering and data mining. At the stage of depth map encoding, we employed static CU splitting decision trees in data mining. WEKA software was selected to train three static decision trees to decide whether the current encoding CU should be divided into a smaller size. We also evaluated the most involved encoder properties, some of which were selected for offline training. We introduced a static decision tree to 3D-HTM 16.0, which was assessed with CTCs in the experiment. The experimental results showed that our scheme realized about 52.3% calculation complexity shrinkage in view of the texture and depth maps. The complexity shrinkage reached 59.0% if only the depth map was considered. At the same time, a BD rate loss of about 0.18% was determined to be acceptable.

#### Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. 61771432, 61302118, and 61702464), Scientific Project (No. 182102210156), and Young Key Teacher (No. 2016GGJS-087).

#### References

1. K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, et al., "3D High-Efficiency Video Coding for Multi-View Video and Depth Data," *IEEE Transactions on Image Processing (TIP)*, Vol. 22, No. 9, pp. 3366-3378, September 2013
2. H. B. Zhang, Y. L. Chan, C. H. Fu, S. H. Tsang, and W. C. Siu, "Quadtree Decision for Depth Intra Coding in 3D-HEVC by Good Feature," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1481-1485, Shanghai, China, March 2016
3. K. Peng, J. C. Chiang, and W. N. Lie, "Low Complexity Depth Intra Coding Combining Fast Intra Mode and Fast CU Size Decision in 3DHEVC," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, pp. 1126-1130, September 2016
4. M. Saldanha, G. Sanchez, C. Marcon, and L. V. Agostini, "Fast 3D-Hevc Depth Maps Intra-Frame Prediction using Data Mining," in *Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 1738-1742, April 2018
5. K. Peng, J. Chiang, and W. Lie, "Low Complexity Depth Intra Coding Combining Fast Intra Mode and Fast CU Size Decision

- in 3D-HEVC,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 1126-1130, 2016
6. H. Zhang, Y. Chan, C. Fu, S. Tsang, and W. Siu, “Quadtree Decision for Depth Intra Coding in 3D-HEVC by Good Feature,” in *Proceedings of the Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1481-1485, 2016
7. P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, et al., “Depth Map Creation and Image-based Rendering for Advanced 3DTV Services Providing Interoperability and Scalability,” *Signal Processing: Image Communication*, Vol. 22, No. 2, pp. 217-234, February 2007
8. F. Bossen, “Common Test Conditions and Software Reference Configurations,” JCT-VC Document L1100, January 2013
9. R. Krishnapuram and J. M. Keller, “A Possibilistic Approach to Clustering,” *IEEE Trans. Fuzzy Syst.*, Vol. 1, No. 2, pp. 98110, May 1993
10. Y. H. Huang, T. S. Ou, and H. H. Chen, “Fast Decision of Block Size, Prediction Mode, and Intra Block for H.264 Intra Prediction,” *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 20, No. 8, pp. 1122-1132, August 2010
11. A. Elyousfi, H. Hamout, and A. E. Hachimi, “An Efficient Intra Block Size Decision for H.264/AVC Encoding Optimization,” in *Proceedings of 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, Morocco, pp. 1-5, Agadir, November 2016
12. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, “The WEKA Data Mining Software: An Update,” *SIGKDD Explorations*, Vol. 11, No. 1, pp. 10-18, May 2009
13. Q. Zhang, K. Huang, X. Wang, B. Jiang, and Y. Gan, “Efficient Multiview Video Plus Depth Coding for 3D-HEVC based on Complexity Classification of the Treeblock,” *Journal of Real-Time Image Processing*, 10.1007/s11554-017-0692-5, May 2017
14. Q. W. Zhang and K. Q. Huang, “Coding Method for HEVC System based on Homogeneity Region Classification Analysis,” *International Journal of Performability Engineering*, Vol. 14, No. 9, pp. 1937-1946, September 2018
15. Q. W. Zhang, Y. H. Wang, T. Wei, B. Jiang, and Y. Gan, “A Complexity Reduction Scheme for Depth Coding in 3D-HEVC,” *Information*, Vol. 10, No. 5, pp. 164, May 2019