Fast Intra Mode Decision for 3D-HEVC Depth Map Coding using Decision Trees

Christopher Moura[1], Mário Saldanha[1], Gustavo Sanchez[2], César Marcon[3], Marcelo Porto[1], Luciano Agostini[1]
[1]Federal University of Pelotas – Pelotas, Brazil
[2]IF Farroupilha – Pelotas, Brazil
[3]Pontifical Catholic University of Rio Grande do Sul – Porto Alegre, Brazil
{ccmoura, mrdfsaldanha, porto, agostini}@inf.ufpel.edu.br, gustavo.sanchez@iffarroupilha.edu.br, cesar.marcon@pucrs.br

***Abstract*—This paper presents a fast intra mode decision for depth map coding on 3D-High Efficiency Video Coding (3D-HEVC) based on decision trees. The proposed solution uses data mining and machine learning to correlate the encoder context attributes and build a set of decision trees. Each decision tree defines if a depth map block must be or not be evaluated by the Depth Modeling Modes (DMMs), considering the encoding context. The decision trees were trained using data extracted from the 3D-HEVC Test Model (3D-HTM) under all-intra encoder configuration. The proposed solution was evaluated according to the Common Test Conditions (CTC), reducing 50.2% the execution time of the depth map coding, and impacting only 0.07% in the Bjontegaard Delta BitRate (BDBR) of the synthesized views.**

Keywords— 3D-HEVC, depth maps, intra coding, decision trees, encoding time reduction

I. Introduction

Many technologies of 3D video compression have been studied for supplying the demand of cinema and 3D videos; these technologies aim to reach high compression rates for video storage and transmission without notable quality loss. The 3D-High Efficiency Video Coding (3D-HEVC) [1] is the state-of-the-art video encoder for 3D videos, which was developed by the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) to increase the compression rate of 3D videos significantly compared to the previous coding standards.

3D-HEVC employs Multiview Video plus Depth (MVD) [2] that associates a depth map to each texture frame. A depth map indicates the distances from objects to the camera in grayscale images, which is characterized by containing large homogeneous regions and sharp edges. Through view synthesis techniques, virtual views can be generated using the combined information of texture frames and depth maps, therefore reducing the number of views to be encoded and transmitted.

The 3D-HEVC encoder inherited several tools from HEVC [3], especially for texture coding. However, new tools were introduced to explore 3D video characteristics efficiently. As 3D-HEVC uses the MVD data format, several of these innovations encompass the depth map coding.

Since depth maps have different characteristics in relation to texture frames, the depth map encoding flow should be modified to achieve high coding efficiency. The conventional HEVC intra-frame prediction is more suitable to encode depth maps with homogeneous regions or regions with smooth transitions in the sample values; however, regions with sudden sample transitions such as sharp edges, those prediction modes can generate artifacts that are noticed in the synthesized views. Consequently, 3D-HEVC inserted the Depth Modeling Modes (DMM), composed of DMM-1 and DMM-4, as an alternative to the conventional HEVC intra prediction modes. DMMs code depth maps dividing the block into two well-defined regions, improving the coding efficiency in edge regions. However, the encoding computational burden is significantly increased, especially when using DMM-1. As demonstrated in [4], for certain values of QP (Quantization Parameter) and block size, DMM-1 can represent up to 56% of the total execution time. Clearly, the DMM evaluation is a costly task; thus, avoiding unnecessary evaluations can save the 3D-HEVC encoding time.

Some works already have proposed solutions to reduce the intra coding time of depth maps. Fu et al. [5] propose an algorithm that uses corner detection for skipping unnecessary DMM evaluations on blocks where the conventional HEVC intra prediction modes can perform a good prediction. Wang et al. [6] propose an algorithm that uses the Roberts operator, which uses local differences to detect edges in the image for skipping DMMs evaluation. Despite the gains in complexity reduction, none of these works explore the correlation between the attributes of the encoder, both are limited to local information in the block.

This paper proposes a decision-trees fast intra-mode decision which reduces the intra encoding time of the depth maps with negligible impact on the coding efficiency. The proposed solution is composed of four decision trees specialized for each block size which can decide when the evaluation of DMMs could be avoided. Experimental results demonstrated that our solution reduces 50.2% of the depth map encoding time, impacting only 0.07% in the Bjontegaard Delta BitRate (BDBR) [7] of the synthesized views.

II. 3D-HEVC Depth Map Intra Coding

The 3D-HEVC depth map intra coding flow is similar to the one in HEVC with its 35 traditional (texture) intra modes and applied to five block sizes (64×64, 32×32, 16×16, 8×8, and 4×4), but with new specific tools for depth map coding. As the Rate Distortion Optimization (RDO) evaluation of all 35 HEVC modes has a high computational cost to the encoder, the Rough Mode Decision (RMD) [8] technique is applied and uses Sum of Absolute Transformed Differences (SATD) to select a subset of modes that will be sent for RDO evaluation, inserting these modes into the RD-list along with DMM-1 and DMM-4.

Fig. 1 shows the intra coding dataflow for depth maps, covering prediction modes of the HEVC intra prediction, DMM-

978-1-7281-6044-3/20/$31.00 ©2020 IEEE

1, DMM-4, and Depth Intra Skip (DIS). The modes inside RD-list are evaluated by Transform-Quantization (TQ) and Segment-wise DC (SDC) and then sent to entropy coding. DIS mode does not perform residual coding and sends the information directly to entropy coding. RD-cost is obtained after the entropy coding, and it is based on the distortion between the original and predicted block and the number of bits needed to represent the block. The mode with the lowest RD-cost in the list is selected as the best encoding mode for the depth block.

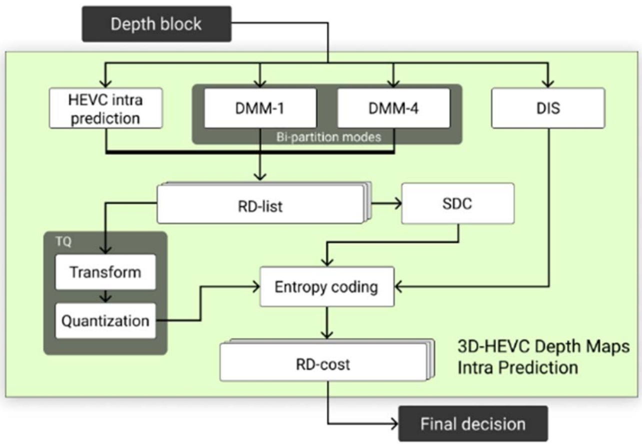

Fig. 1. Dataflow model used for the 3D-HEVC depth map intra prediction.

A. HEVC Intra Prediction

Fig. 2 shows that the HEVC intra prediction encompasses 35 prediction modes - 33 directional modes and two non-directional modes: DC and planar. The HEVC intra prediction applies RMD to select the three best modes for 64×64, 32×32 and 16×16 blocks and eight for 8×8 and 4×4 blocks. RMD evaluates all the mode possibilities locally using the SATD between the original and predicted signal. The Most Probable Modes (MPM) technique is also applied, where a maximum of two coding modes are inserted in the RD-list, these modes are chosen based on information from neighboring blocks previously coded.

B. Depth Intra Skip (DIS)

DIS is a tool to explore the homogenous regions of depth maps, where the sample values have slight variations. Since homogeneous regions do not tend to significantly impact the visual quality of synthesized views, the DIS mode skips the residual coding and does not transmit residues information to the bitstream.

The DIS uses four prediction modes, two vertical and two horizontal modes. The first vertical mode is the same as the HEVC intra-frame prediction, and the second vertical mode encodes the block using the central sample of the block as reference. The two horizontal modes are similar to the vertical modes, but use samples from the left previously encoded block as reference.

C. Depth Modeling Modes (DMMs)

3D-HEVC adds DMM-1 and DMM-4 for reaching high efficiency on edges evaluation in the depth map intra-frame coding, for block sizes 32×32, 16×16, 8×8, and 4×4.

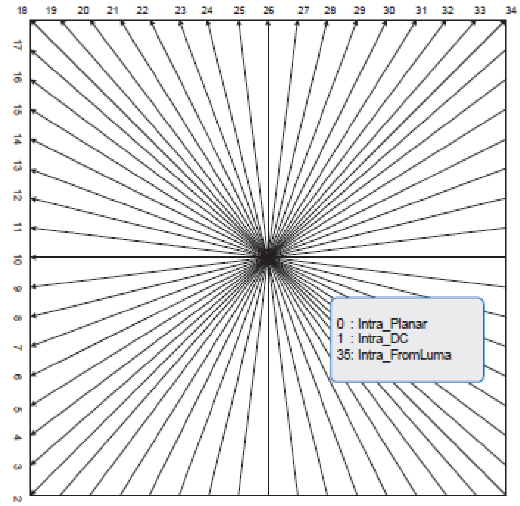

Fig. 2. HEVC intra prediction modes.

The DMMs segment the encoding block into two regions, and each region is represented by a constant value called Constant Partition Value (CPV), this value is the average of all values in the region. The DMM-1 mode divides the block into two regions by a *wedgelet*, where wedgelet is a straight line that divides the block at the edge region, as presented in Fig. 3(a). The DMM-4 mode divides the block by *contours* that are created using the texture view as reference, creating two regions consisting of non-linear edges or disconnected regions, as demonstrated in Fig. 3(b).

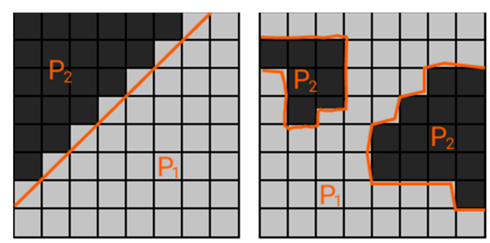

Fig. 3. Wedgelet and contour partitions of DMM-1 and DMM-4, respectively.

The evaluation of DMMs is a complex computational task for the 3D-HEVC encoder [4], especially DMM-1, where the best angle of the line that will partition the block must be chosen because there are several possible wedgelets for a depth block.

DMMs tend to be chosen as the best coding mode in regions with well-defined edges, in homogeneous regions of depth maps the DMMs evaluation can be unnecessary. Generally, depth maps have more homogeneous regions, where the evaluation of DMMs can be avoided resulting in a significant execution time decrease for the 3D-HEVC encoder.

III. Attributes Analysis

We selected some encoder attributes to analyze their efficiency in the decision tree training regarding each block size. The data mining process consists of extracting and storing data from the encoder context and video sequences. In the context of the intra coding, the blocks are called Prediction Units (PUs), which are responsible for storing relevant prediction information. Some used attributes are based on information extracted from PU parents. Parent PU is the one that is up to one level in the size of the current PU, for example, 8×8 is the parent of a 4×4 PU, 16×16 is the parent of an 8×8 PU, and so on.

We assessed the following attributes and stored their values according to each block size, along with the decision of selecting DMM as the best mode in RDO; the values were used to train the decision trees to decide if the DMMs evaluation should be performed.

A. *Attributes*

- The *QP-depth* value of the current block. This value defines the compression rate for the block. Higher QPs indicate that compression will retain less image detail than compression with smaller QPs, influencing the number of DMM evaluations.
- The *average* of all samples in the current block. This value indicates if the block is distant or close to the camera. Closer blocks tend to preserve more details in the image compared to more distant blocks, so the amount of DMM ratings also varies with the distance.
- The maximum difference (*MaxDiff*) between the samples in the current block. This value informs the difference between two samples; a high variation indicates that the block may contain an edge.
- The first two modes stored in the RD-list (*FirstRDListMode* and *SecondRDListMode).* The indication of a homogeneous zone occurs when the first modes are dedicated to encoding homogeneous regions.
- The RD-cost of the first mode stored in the RD-list (*RDCostFirst*). If this value is small, the block tends to be well encoded with the traditional HEVC intra prediction modes, and the DMMs evaluation can be skipped.
- The maximum variance (*MaxVar_size*) of the samples of smaller blocks inside the current block. *MaxVar_size* is used as *MaxVar_16*, *MaxVar_8*, *MaxVar_4*, indicating the maxima variances of the 16×16, 8×8, and 4×4 sub-blocks inside the current block. Low values of *MaxVar_size* can indicate a homogeneous zone in the current block.
- The best mode selected for encoding the parent PU (*ParentMode*), if it was a DMM or an intra HEVC mode. Parent block is the one that is up one level when partitioning the current block; e.g., a 16×16 is parent of an 8×8 block.
- If the parent block used SDC or TQ (*ParentSDC*). SDC was designed to explore the proximity characteristic in the block sample values. If the parent block was encoded using SDC, then there is evidence that the current block is homogeneous.
- The RD-cost of the parent block (*ParentRDCost*).

B. *Probability Density Analysis*

Fig. 4 exemplifies the effectiveness of some attributes with the probability density function choosing one of the DMMs as the best mode for 32×32 block size encoded with QP-depth 40. The blue line indicates the probability density of the block not performing DMM encoding, and the orange dotted line indicates when the block tends to perform DMMs coding. Similar behavior was obtained in lower block sizes; however, they were omitted in the paper due to space limitation.

Fig. 4(a) shows that the encoder tends to select DMM as the best mode for high *MaxDiff* values. As previously stated, this attribute can signal if the block has characteristics of a homogeneous or heterogeneous zone. Fig. 4(b) displays that low *RDCostFirst* values imply a low probability of evaluating the DMMs. Fig. 4(c) illustrates that when a block has low values of maximum variance of 4×4 sub-blocks, there is a low probability of this block choose the DMMs as best coding mode. Then, *MaxVar_4* is useful information to decide the DMM evaluation in blocks of size 32×32, as it indicates that there may be borders in the image. Fig. 4(d) shows that low values of *ParentRDCost* indicate that the encoder tends not to evaluate DMMs.

The probability density analysis shows that these attributes are relevant to the decision trees - the other attributes not mentioned here present similar behavior.

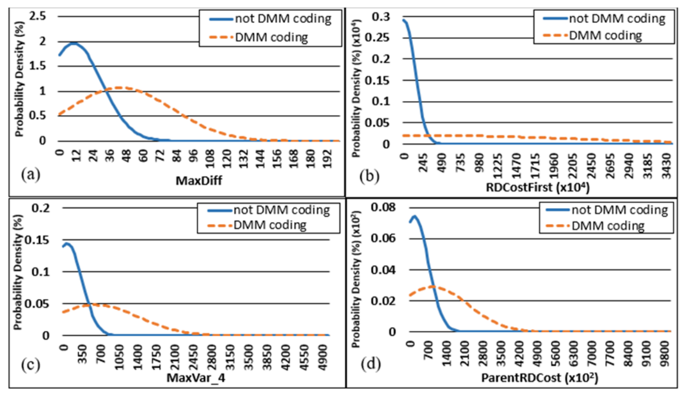

Fig. 4. Probability density function for 32×32 block sizes encoded with QP-depth value 40.

IV. Developed Decision Trees

The decision trees generated by the data mining process decide if the DMMs should be evaluated in PUs of sizes 4×4, 8×8, 16×16, and 32×32. We encoded 50 frames of *Balloons*, *Champagne_tower*, and *Pantomine* test video sequences to generate the decision trees. The extracted data refer to blocks of 4×4, 8×8, 16×16, and 32×32 sizes, the amount of extracted data for each block size is different, given that there is a larger number of small blocks. The encoding was done in the 3D-HEVC Test Model (3D-HTM) (version 16.3) [9] under all-intra configuration according to the QP-pair values (25, 34), (30, 39), (35, 42), and (40, 45) as indicated in the Common Test Conditions (CTC) [10]. After collecting the attributes of the three sequences, the data for each decision tree was trained in the Waikato Environment for Knowledge Analysis (WEKA) [11], version 3.8.3. The WEKA Spread Subsample filter is used to create two groups with equal size, where the first group contain only blocks encoded by DMMs and the second group contain only blocks not encoded by DMMs. Both groups contain the same number of samples in order to avoid training specialized trees in only one class.

We chose the *Reduced Error Pruning Tree* (REPTree) algorithm for decision trees training; this algorithm uses the *Reduced Error Pruning* (REP) [12] to reduce the risk of sample overfitting. Table I shows the attributes, accuracy, and size in nodes of each decision tree.

Table I. Attributes, accuracy, and size of each decision tree.

Attribute	Decision tree			
	32×32	**16×16**	**8×8**	**4×4**
QP-depth	×	×	×	×
Average	×	×	×	×
MaxDiff	×	×	×	×
MaxVar_4	×	×	–	*
MaxVar_8	×	–	*	*
MaxVar_16	×	*	*	*
ParentMode	–	×	×	×
ParentSDC	×	×	×	–
ParentRDCost	–	×	×	×
FirstRDListMode	×	×	×	×
SecondRDListMode	×	×	×	×
RDCostFirst	×	×	×	×
Accuracy	**82.16%**	**86.02%**	**88.93%**	**93.72%**
Size	**49**	**55**	**45**	**75**

$^{\times}$Selected attribute $^{-}$Unused attribute *Unavailable attribute

V. Experimental Results

The performance of the proposed decision-trees based fast depth maps intra encoding was evaluated implementing the four decision trees (32×32, 16×16, 8×8, and 4×4) in the 3D-HTM intra-frame coding module, and employing seven video sequences (*Kendo*, *Newspaper_CC*, *GT_Fly*, *Poznan_Hall2*, *Poznan_Street*, *Undo_Dancer*, and *Shark*) specified in the CTC and encoded with the four QP-pair values ((25, 34), (30, 39), (35, 42), and (40, 45)) under all-intra encoder configuration.

Table II presents the coding efficiency of the decision trees considering the BDBR of synthesized views, and the Execution Time Reduction (ETR) concerning the original encoder without applying the complexity reduction technique presented in [13]. The proposed solution can reduce 47.7% of the execution time of the entire 3D-HEVC encoding, on average (ranging from 33.5% to 58.2%), with an increase of 0.07% in BDBR, on average (ranging from 0.02% to 0.15%). Considering only depth map coding, the execution time is reduced by 50.2%, on average. The average time reductions were 37.4% and 51.8% for 1024×768 and 1920×1088 resolution videos, respectively. Activating the technique of [13], on average, ETR is 12.5% and the BDBR of the synthesized views increases by 0.05%.

Table II. Results of the proposed solution and related works.

Video	This work		Fu [5]		Wang [6]	
	BDBR	ETR	BDBR	ETR	BDBR	ETR
Balloons	-	-	0.16%	15.3%	1.21%	36.7%
Kendo	0.12%	41.3%	0.12%	17.4%	1.02%	34.3%
Newspaper_CC	0.09%	33.5%	0.12%	11.7%	0.86%	34.8%
GT_Fly	0.04%	51.8%	0.03%	20.6%	0.86%	38.2%
Poznan_Hall2	0.15%	58.2%	0.52%	23.0%	0.98%	33.1%
Poznan_Street	0.01%	47.1%	0.00%	9.1%	0.66%	36.6%
Undo_Dancer	0.07%	54.8%	0.37%	18.8%	1.35%	34.8%
Shark	0.02%	47.0%	-	-	1.57%	34.3%
Average	0.07%	47.7%	0.19%	16.5%	1.06%	35.4%

Additionally, Table II presents the results of two related works. The solution proposed in [5] obtained average BDBR increase of 0.19% and 16.5% of ETR. The work [6] attained average ETR increase of 35.4%, with a 1.06% increase in the BDBR. Although these works have a minimal impact on the coding efficiency, the solution proposed in this work achieved a more significant ETR in the 3D-HEVC encoder.

VI. Conclusions

This paper presented a decision-trees based fast depth maps intra-frame coding for the 3D-HEVC. Four decision trees were trained, one for each block size supported in the intra depth maps coding, to decide when DMMs evaluation can be avoided. The decision trees were implemented in 3D-HEVC intra coding module and evaluated according to the CTC. Experimental results showed that the proposed solution can reduce the execution time of the 3D-HEVC encoder by 47.7% with a negligible increase of 0.07% in BDBR of synthesized views, achieving higher execution time reduction and lower BDBR increase compared to the related works.

Acknowledgment

The authors thank FAPERGS, CNPq and CAPES (Finance Code 001) Brazilian research support agencies that financed this investigation.

References

[1] G. Tech, Y. Chen, K. Muller, J. Ohm, A. Vetro, Y. Wang, “Overview of the Multiview and 3D extensions of High Efficiency Video Coding,” IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), v. 26, n. 1, pp. 35-49, Jan. 2016.

[2] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Smolic, R. Tanger, “Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability,” Signal Processing: Image Communication, v. 22, n. 2, pp. 217-234, Feb. 2007.

[3] G. Sullivan, J. Ohm, W. Han, T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” Transactions on Circuits and Systems for Video Technology (TCSVT), v. 22, n. 12, pp. 1649-1668, Dec. 2012.

[4] G. Sanchez, R. Cataldo, R. Fernandes, L. Agostini, C. Marcon, “3D-HEVC depth maps intra prediction complexity analysis,” Proceedings of the IEEE International Conference on Electronics, Circuits and Systems (ICECS), pp. 348-351, 2016.

[5] C. Fu, Y. Zhao, H. Zhang, Y. Chan and W. Siu, "Depth modelling mode decision for depth intra coding via good feature," 2017 IEEE International Conference on Image Processing (ICIP), pp. 4018-4022, Beijing, 2017.

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

[7] G. Bjontegaard, “Calculation of Average PSNR Differences between RD-Curves”, Proceedings of the ITU-T Video Coding Experts Group (VCEG) Thirteenth Meeting, pp. 1-5, 2001.

[8] L. Zhao, L. Zhang, S. Ma, D. Zhao, “Fast Mode Decision Algorithm for Intra Prediction in HEVC,” Proceedings of the IEEE Visual Communications and Image Processing (VCIP), pp. 1-4, 2011.

[9] 3D-HEVC Test Model, Available at: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware /tags/HTM-16.3/, access in Jun. 2020.

[10] D. Rusanovskyy, K. Muller, A. Vetro. “Common Test Conditions of 3DV Core Experiments,” ISO/IEC JTC1/SC29/WG11 MPEG2011/N12745, Geneva, Jan. 2013.

[11] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. Witten. “The WEKA data mining software: an update,” SIGKDD Explorations, v. 11, n. 1, pp. 10-18, Nov. 2009.

[12] C. Brunk, M. Pazzani, “An investigation of noise-tolerant relational concept learning algorithms,” Proceedings of the International Workshop on Machine Learning. pp. 389-393, 1991.

[13] Z. Gu, J. Zheng, N. Ling, P. Zhang, “3D-CE5.h related: Fast Intra Prediction Mode Selection for Intra Depth Map Coding”, Document: JCT3V-E0238, 5th meeting, Vienna, Aug. 2013.