

Time Reduction on 3D-HEVC Depth Maps Coding using Static Decision Trees Built Through Data Mining

Mário Saldanha¹, Marcelo Porto¹, César Marcon², Luciano Agostini¹

¹Video Technology Research Group (ViTech) – Federal University of Pelotas (UFPEL) – PPGC – Pelotas, Brazil

²Pontifical Catholic University of Rio Grande do Sul (PUCRS) – Porto Alegre, Brazil

{mrdfsaldanha, porto, agostini}@inf.ufpel.edu.br, cesar.marcon@pucrs.br

ABSTRACT

This dissertation presents a fast depth map coding for 3D-High Efficiency Video Coding (3D-HEVC) based on static Coding Unit (CU) splitting decision trees. The proposed solution is based on our previous works and avoids the costly Rate-Distortion Optimization (RDO) process for depth maps coding, which evaluates several possibilities of block partitioning and encoding modes for choosing the best one. This coding approach uses data mining and machine learning to extract the correlation among the encoder context attributes and to build the static decision trees. Each decision tree defines if a depth map CU must be split into smaller blocks, considering the encoding context through the evaluation of the CU features and encoder attributes. The results demonstrated that this approach can halve the 3D-HEVC encoder processing time with negligible coding efficiency loss. Besides, the obtained results surpass all related works regarding processing time and coding efficiency. The results reported in this dissertation were published in three journals and two events, besides generate a patent deposit. These products have the master student as the first author.

KEYWORDS

3D video coding, 3D-HEVC, depth maps, computational effort reduction, machine learning.

1 Introduction

This paper summarizes the obtained results in the dissertation defended in the Graduate Program in Computer Science on March 1, 2018. The work was developed by Mário Saldanha, in 23 months, advised by Marcelo Porto, César Marcon, and Luciano Agostini.

3D video systems have evolved considerably in the last few years due to their real-world visual experience that goes beyond 2D videos. Focusing on increasing the 3D coding efficiency, the experts of Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) has developed the 3D-High Efficiency Video Coding (3D-HEVC) [1] as an extension of High Efficiency Video Coding (HEVC) standard used on 2D videos.

3D-HEVC uses the Multi-View plus Depth (MVD) [2] data format, where each texture frame is associated with a depth map. The depth maps provide geometric information of the scene and allow the generation of virtual views after the decoding process through lightweight synthesis views techniques. Since only a subset of texture views and associated depth maps are encoded and transmitted, a significant bitrate saving is achieved. However, the encoder and decoder systems have the computational effort increased due to the insertion of the depth maps coding.

The depth maps coding process provides a flexible quadtree-based structure, where each frame is divided into Coding Tree Units (CTUs), and each CTU can be recursively divided into Coding Units (CUs) [3]. The maximum and minimum sizes of a CU are 64×64 and 8×8, respectively. Additionally, each CU may be divided into one, two or four Prediction Units (PUs) with sizes varying from 4×4 to 64×64 [3]. Besides, for each variation of CU and PU sizes several prediction modes are evaluated.

In 3D-HEVC Test Model (3D-HEVC) [4] the partitioning structure for each CTU is chosen through Rate-Distortion Optimization (RDO), which evaluates several combinations of block partitioning and prediction modes seeking for the best encoding possibility. This process reaches a very high coding efficiency at the cost of a significant increase in the encoder computational effort.

Some works in the literature proposed solutions to decrease the encoding computational effort (the encoding time was used to measure the computational effort) of depth maps coding, such as [5]-[7]. Our previous works [8]-[10], which served as basis for this dissertation, also focus on reducing the computational effort of depth maps coding using a variety of solutions. However, all these works considered limited features of the encoding process and do not achieved satisfactory tradeoff between encoding time reduction and coding efficiency losses. Then, this dissertation presents a solution based on data mining and machine learning to extract useful attributes of the encoding process and to build static decision trees to define if each CU should be split into smaller CUs and avoid the complex RDO process. Experimental results demonstrate that the encoding time is reduced about 50% maintaining the encoding efficiency. Moreover, the proposed solution surpasses related works methods regarding both axes. A preliminary version of this work, which considers only intra-frame prediction, was published in [11]. The complete solution considering intra-frame and inter-frame/view predictions was published in [12] and the

In: I Concurso de Teses e Dissertações (CTD 2019), Rio de Janeiro, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2019.
©2019 SBC – Sociedade Brasileira de Computação.
ISSN: 2596-1683

current paper is based on that publication. Finally, this dissertation also generated a patent deposit [13].

2 Analysis and Motivation

Experimental analyzes were performed to evaluate the impact of depth map computational effort inside the 3D-HEVC encoder in both All-Intra (AI) and Random-Access (RA) encoder configurations [14] using the reference software 3D-HTM version 16.0 [15]. Besides, the experiments considered the Common Test Conditions CTC [14] using eight 3D video sequences and four pairs of Quantization Parameters (QP-pair).

Fig. 1 presents the distribution of the 3D-HEVC encoding effort considering the average results for CTC sequences and QP-pairs. For both AI and RA, the depth maps require higher encoding effort than texture views. In AI configuration, the encoder spends an average of 83% and 17% of the time to encode depth maps and texture, respectively. Considering AI, the depth map coding is 4.8 times (on average) more time demanding than texture. In RA, the depth maps also require higher encoding effort than texture and the QP-pair has a higher influence. For QP-pair (40, 45), 80% of the encoding time is spent to encode depth maps and 20% to encode texture. In this case, the depth map coding is 2.9 times more time demanding than texture coding, on average.

This evaluation demonstrates that the depth maps coding is a bottleneck in the 3D-HEVC encoder and solutions for reducing the computational effort are needed.

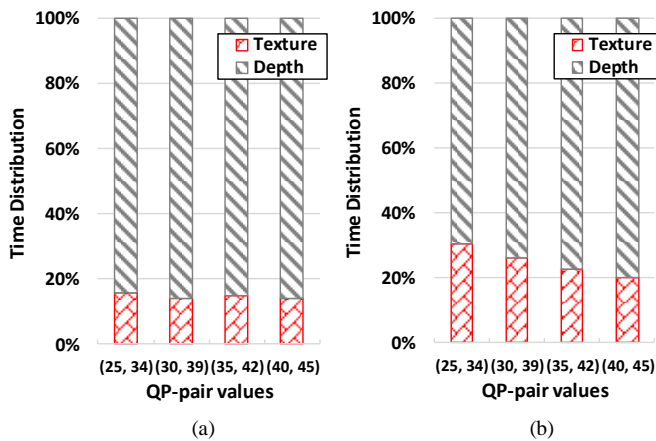


Fig. 1. 3D-HEVC computational effort distribution in (a) AI and (b) RA configurations [12].

Fig. 2 presents an evaluation of the distribution of CU sizes for depth maps coding considering CTC QP-pairs in AI and RA configurations. One can notice that the QP value directly affects the CU size distribution for both configurations. It occurs because QP defines the compression rate, influencing the image quality. High QPs generate more homogenous areas in the coded image that are efficiently encoded using larger CUs sizes. However, for lower QPs the encoder maintains several image details, requiring lower CU sizes to achieve higher coding efficiency.

In this analysis, RA configuration uses less small CUs compared to the AI configuration, because RA allows the inter-frames and inter-view redundancies exploration using motion and

disparity estimation. The exploration of these redundancies in P- and B-frames allows higher encoding efficiency reusing larger blocks from reference frames.

Fig. 3 illustrates the computational effort to evaluate each CU size, considering only depth map encoding time and AI configuration. One can observe there is a high computational cost associated with the evaluation of smaller CU sizes, such as 8×8. Also, one can conclude the smallest encoding effort is spent in 64×64 CUs for all QP-pair values.

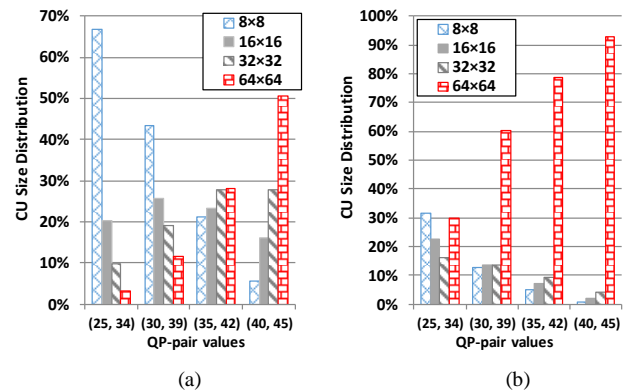


Fig. 2. Distributions of CU sizes for 3D-HEVC depth maps in (a) AI and (b) RA configurations [12].

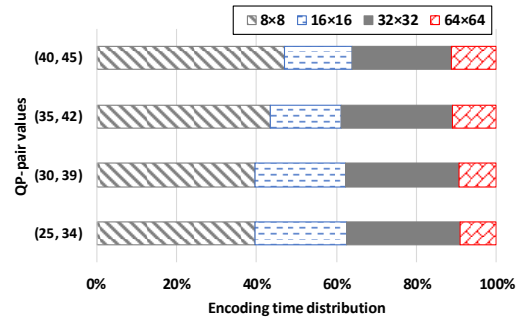


Fig. 3. Encoding time distribution of each CU size for four QP-pairs considering AI configuration [12].

From the presented results, one can conclude that an early decision algorithm limiting the evaluation of smaller CUs for some cases of the depth map coding could provide a good encoder decision accuracy and high encoding time reduction. Besides, a specialized solution for frames that consider only intra-frame coding and a specialized solution for frames that use inter-frame and inter-view coding can lead with higher accuracy due to the different distributions of CU sizes in AI and RA configurations.

Although the QP values and the encoder configurations have a substantial influence on the quadtree definition, the simple removal of some quadtree level is not an appropriate solution for reducing the encoding effort, since this decision can significantly degrade the encoding efficiency (bitrate versus quality). However, a solution able to decide when a current CU should not be split into smaller CUs considering the QP value, the encoder configurations and other encoder attributes, can avoid the high cost of evaluating the full RDO process. Then, a most robust solution should be

provided to reduce the computational effort with minor impact on the coding efficiency.

3 Proposed Solution

This dissertation explores the use of data mining and machine learning to identify attributes with high correlation with the CU split decision to build static decision trees. The decision trees determine when a CU should be split into smaller sizes. The split choice occurs when the decision trees classifies the encoding CU as requiring further evaluations; otherwise, the computation of that CU is finalized. The use of data mining and machine learning allows discovering correlations between the encoding context and its attributes. If these correlations are strong enough, it is possible to define static decisions trees to reduce the encoding computational effort with negligible coding efficiency loss.

This work proposes one tree for each splitting decision of CU sizes 16×16, 32×32, and 64×64 instead of the full RDO process. Since I-frames uses different encoding tools than P- and B-frames, three trees are employed for I-frames and three trees are defined for P- and B-frames.

A large amount of data from the depth video sequences and internal encoding variables were collected to find features that could lead to effective decisions of CU splitting. The attributes evaluated to define the static decision trees for I-frames are presented in Table I. For building the static decision trees of P- and B-frames the same attributes of Table I were considered and new attributes were evaluated, which are demonstrated in Table II. The all-intra configuration was considered to define the I-frames decision trees and the random-access configuration was used to define the decision trees for P- and B-frames.

Table I. Attributes evaluated in I-frames [12].

Attribute	Description
<i>Average</i>	Average value of original samples of the current CU
<i>VAR</i>	Variance of the CU original samples
<i>VAR_size</i>	Maximum <i>VAR</i> of smaller blocks inside the current CU (4×4 up to the current CU size, exclusive). <i>VAR_size</i> is used as <i>VAR_32</i> , <i>VAR_16</i> , <i>VAR_8</i> , and <i>VAR_4</i> - the maxima variances of the 32×32, 16×16, 8×8, and 4×4 sub-blocks inside the current CU, respectively
<i>Grad</i>	Gradient of the 4 corners of the original CU (i.e., the maximum absolute difference among the 4 samples in the corners)
<i>Grad_size</i>	Maximum <i>Grad</i> of smaller blocks inside of the current CU. It can assume sizes similar to <i>Var_size</i>
<i>MaxDiff</i>	Maximum difference among samples of the current CU
<i>RD-cost</i>	RD-cost when encoding the current CU
<i>QP</i>	The current QP-depth value

For the data mining process, Kendo video sequence was coded in both AI and RA configuration, considering all CTC QP values. For each coded CU, were stored all information in Table I and Table II, and the information indicating if the CU has been split into smaller sizes or not. More details about the used attributes and their evaluations can be found in [12].

Since there are a limited number of 3D video sequences with their depth maps available to make 3D video coding experiments,

Kendo video sequence was randomly selected from the CTC dataset, and it was used to extract the data necessary to the offline training process. However, all video sequences defined in the CTCs were evaluated to validate the proposed solution and to demonstrate that the trained solution can achieve a high performance in different encoding scenarios.

Table II. New attributes evaluated in P- and B-frames [12].

Attribute	Description
<i>RD_MSM</i>	RD-cost of <i>Merge/Skip</i> mode
<i>RD_DIS</i>	RD-cost obtained in <i>DIS</i> mode
<i>Ratio</i>	RD-cost obtained in intra-frame prediction divided by the <i>RD_DIS</i>
<i>RatioInter</i>	RD-cost of inter 2N×2N evaluation divided by <i>RD_MSM</i>
<i>RelRatio</i>	The normalized difference between the RD-cost of inter 2N×2N and <i>RD_MSM</i>
<i>Neigh_depth</i>	The average quadtree depth of the top, left, top-left, top-right, and co-located CTUs of both reference lists
<i>SKIP_flag</i>	Notify if the CU has been encoded using the <i>Skip</i> mode
<i>DIS_flag</i>	Notify if the CU has been encoded using the <i>DIS</i> mode
<i>MAD_size</i>	The maximum Mean Absolute Deviation of the smaller blocks inside of the current CU (4×4 up to current CU size, exclusive)

The Waikato Environment for Knowledge Analysis (WEKA) [16], version 3.8, was used to train the decision trees using the J48 algorithm. Seeking for a better solution for data balancing the input files were organized in two sets of data with equal sizes containing inputs that result in splitting and not splitting of CUs. Besides, to avoid the overfitting problem, the Reduced Error Pruning (REP) was performed in each tree, allowing a better generalization.

Fig. 4 illustrates the static decision tree generated for 64×64 CUs of I-frames, where the leaves “N” and “S” correspond to the not split and split decisions, respectively. The remaining decision trees for 32×32 and 16×16 CUs and for P- and B-frames have similar structure.

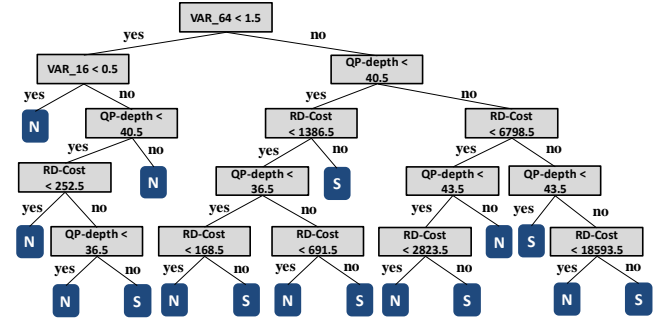


Fig. 4. Decision tree for splitting decision in 64×64 CUs [12].

4 Experimental Results

The proposed static decision trees were implemented into the 3D-HEVC Test Model (3D-HTM) (version 16.0) and evaluated following the CTC for 3D experiments using RA encoder configuration. Table III presents the results of this solution regarding Bjontegaard Delta-rate (BD-rate) [17] considering the synthesized views quality, the computational effort reduction

(regarding processing time) for the whole 3D-HEVC encoder (texture and depth maps). The average result of test sequences does not consider the training sequence Kendo, but the results for this sequence also are presented in Table III. These results demonstrated that this solution is not overfitted in the training data and can achieve good results in different scenarios.

The proposed solution achieved an average computational effort reduction of 52.7% (from 47.7% to 59.4%) with an average BD-rate degradation of 0.18% (from 0.01% to 0.55%). Besides, the proposed solution was compared with the works [5] and [6], which also focused in the computational effort reduction for 3D-HEVC depth maps encoding. The method proposed by [5] achieves a computational effort reduction of 52.0% with an average BD-rate increase of 1.02%. In [6], a computational effort reduction of 41.5% was obtained with an impact in BD-rate of 2.06%. Therefore, the solution proposed in this work reached the highest computational effort reduction and the lowest encoding efficiency degradation when compared with the related works.

5 Conclusions

This dissertation presented a fast 3D-HEVC depth maps coding solution, which uses data mining and machine learning to extract correlations among the encoder context attributes to build static decision trees and define the partitioning of CUs instead of the complete and complex RDO process. After encoding a given CU size, a decision tree decides if the CU must be split into smaller sizes. Two types of trees were trained using J48 algorithm, being one for I-frames and other for P- and -B-frames since the encoder behavior is significantly different for these types of frames, requiring specialized trees. The static decision trees were implemented in the 3D-HTM and evaluated under the CTC. Experimental results demonstrated that the proposed solution can halve the 3D-HEVC encoder processing time, with a negligible impact on BD-rate, reaching the best results in both axes when compared with the related works. Considering the relevance of the reached results, this dissertation generated three articles published in journals, two works published in international events and one patent deposit, as presented in this text.

ACKNOWLEDGMENTS

The authors acknowledge FAPERGS, CNPq and CAPES (Finance Code 001) Brazilian research support agencies to support this work.

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] K. Muller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. Rhee, G. Tech, M. Winken, T. Wiegand, "3D High-Efficiency Video Coding for Multi-View Video and Depth Data," *IEEE Transactions on Image Processing (TIP)*, v. 22, n. 9, pp. 3366-3378, Sep. 2013.
- [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] Y. Chen, G. Tech, K. Wegner, S. Yea, "Test Model 11 of 3D-HEVC and MV-HEVC," document JCT3VK1003 of JCT-3V, Geneva, CH, Feb. 2015.
- [5] E. Mora, M. Cagnazzo, B. Pesquet-Popescu, "Initialization, Limitation, and Predictive Coding of the Depth and Texture Quadtree in 3D-HEVC," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, v. 24, n. 9, pp. 1554-1565, Sep. 2014.
- [6] J. Lei, J. Duan, F. Wu, "Fast Mode Decision Based on Grayscale Similarity and Inter-View Correlation for Depth Map Coding in 3D-HEVC," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, v. 28, n. 3, pp. 706-718, Mar. 2016.
- [7] H. Zhang, Y. Chan, C. Fu, S. Tsang, W. Siu, "Quadtree Decision for Depth Intra Coding in 3D-HEVC by Good Feature," *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1481-1485, 2016.
- [8] M. Saldanha, G. Sanchez, C. Marcon, L. Agostini, "Block-level fast coding scheme for depth maps in three-dimensional high efficiency video coding," *Journal of Electronic Imaging (JEI)*, v. 27, p. 1-4, 2018.
- [9] M. Saldanha, G. Sanchez, B. Zatt, M. Porto, L. Agostini, "Energy-aware scheme for the 3D-HEVC depth maps prediction," *Journal of Real-Time Image Processing*, v. 13, p. 55-69, 2017.
- [10] M. Saldanha, B. Zatt, M. Porto, L. Agostini, G. Sanchez, "Solutions for DMM-1 complexity reduction in 3D-HEVC based on gradient calculation," *Latin American Symposium on Circuits & Systems (LASCAS)*, pp. 211-214, 2016.
- [11] M. Saldanha, G. Sanchez, C. Marcon, L. Agostini, "Fast 3D-HEVC Depth Maps Intra-Frame Prediction Using Data Mining," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1738-1742, 2018.
- [12] M. Saldanha, G. Sanchez, C. Marcon, L. Agostini, "Fast 3D-HEVC Depth Map Encoding Using Machine Learning," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, v. pp, p. 1-1, 2019.
- [13] M. Saldanha, G. Sanchez, C. Marcon, L. Agostini, "Método de Otimização em Codificação de Imagens Tridimensionais e Codificador para Imagens Tridimensionais. 2018, Brasil. Patente: Privilégio de Inovação. Número do registro: BR10201800597. Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial. Depósito: 26/03/2018.
- [14] K. Muller, A. Vetro, "Common Test Conditions of 3DV Core Experiments," *ISO/IEC JTC1/SC29/WG11 MPEG2011/N12745*, Jan. 2014.
- [15] 3D-HEVC Test Model, Available at: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-16.0/, access in Jul. 2018.
- [16] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. Witten, "The WEKA Data Mining Software: An Update," *ACM SIGKDD Explorations Newsletter*, v. 11, n. 1, pp. 10-18, Jan. 2009.
- [17] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-Curves," *VCEG Meeting*, pp. 1-5, 2001.

Table III. Proposed solution results for CTC evaluation in random-access configuration.

Video	This work		Mora [5]		Lei [6]	
	BD-rate	Computational Effort reduction	BD-rate	Computational Effort reduction	BD-rate	Computational Effort reduction
Balloons	0.06%	47.7%	2.03%	48.2%	4.32%	39.7%
Kendo	-	-	1.65%	48.5%	1.20%	33.9%
Newspaper_CC	0.05%	48.1%	1.40%	52.3%	3.43%	43.0%
GT_Fly	0.07%	52.2%	0.33%	50.3%	0.87%	45.1%
Poznan_Hall2	0.39%	59.4%	2.01%	58.6%	3.31%	43.3%
Poznan_Street	0.10%	57.0%	0.05%	56.5%	1.22%	45.4%
Undo_Dancer	0.55%	53.3%	0.37%	50.7%	1.67%	48.3%
Shark	0.01%	51.4%	0.35%	50.8%	0.48%	33.3%
Average	0.18%	52.7%	1.02%	52.0%	2.06%	41.5%
Kendo	0.03%	47.1%	-	-	-	-