Journal of
Electronic Imaging

JElectronicImaging.org

Block-level fast coding scheme for depth maps in three-dimensional high efficiency video coding

Mário Saldanha
Gustavo Sanchez
César Marcon
Luciano Agostini

Mário Saldanha, Gustavo Sanchez, César Marcon, Luciano Agostini, "Block-level fast coding scheme for depth maps in three-dimensional high efficiency video coding," *J. Electron. Imaging* **27**(1), 010502 (2018), doi: 10.1117/1.JEI.27.1.010502.

Block-level fast coding scheme for depth maps in three-dimensional high efficiency video coding

Mário Saldanha,[a,*] Gustavo Sanchez,[b,c] César Marcon,[b] and Luciano Agostini[a]
[a]Federal University of Pelotas, Video Technology Research Group, Pelotas, Brazil
[b]Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil
[c]IF Farroupilha, Alegrete, Brazil

Abstract. A block-level fast coding scheme for depth maps coding in 3D-high efficiency video coding (3D-HEVC) is presented. The proposed scheme uses the information of the correlated texture coding unit (CU) to accelerate the encoding depth map CU. The experimental analysis demonstrates that when the SKIP mode encodes the texture CU, the current depth map CU has a high probability of being encoded by SKIP or depth intraskip mode. Therefore, the evaluation of the remaining encoding modes can be skipped when any of these modes obtained a low rate-distortion cost. Experimental results obtained through 3D-HEVC test model 16.0 report that our scheme reduces the encoding time by 26.9% with encoding efficiency losses lower than 0.3%. [DOI: 10.1117/1.JEI.27.1.010502]

Keywords: 3D-high efficiency video coding; depth maps; time saving; intercomponent; early decision.

Paper 170789L received Sep. 13, 2017; accepted for publication Jan. 19, 2018; published online Feb. 7, 2018.

1 Introduction

In response to the need for higher compression rates for three-dimensional (3-D) video applications, 3D-high efficiency video coding (3D-HEVC)[1] was developed as an extension of the high efficiency video coding (HEVC) standard. The usage of the multiview video plus depth (MVD)[2] representation format is a crucial factor for 3D-HEVC efficiency. In MVD representation, each texture view is associated with a depth map, which provides the geometrical information of the captured texture view according to the distance between the objects and the camera.

The texture views and their associated depth maps can be used to synthesize intermediate virtual viewpoints using techniques such as depth image-based rendering,[2] where the virtual views are required only at decoder side and can be displayed using autostereoscopic displays. The 3D-HEVC depth map coding inherits tools existing for texture views in HEVC (such as SKIP, MERGE, and intraframe encoding modes).[3] Furthermore, specific tools for depth maps were inserted during the standardization since depth maps have distinct characteristics from texture views. While texture views have a complex color behavior with a multicolor view, depth maps are composed of large homogeneous regions (background and objects bodies) and sharp edges (borders of objects). Therefore, tools such as depth modeling modes (DMMs),[4] segment-wise direct component (SDC),[5] and depth intraskip (DIS)[6] were developed to encode depth maps efficiently. Since the texture and its associated depth map represent the same scene at the same time and viewpoint, their encoding information is substantially correlated, and significant gains may be achieved by efficient predictions or inheritances of encoding information.[7,8]

3D-HEVC utilizes the same quadtree-based encoding structure introduced in HEVC for both texture and depth components. In the current 3D-HEVC reference software implementation, called 3D-HEVC test model (3D-HTM),[9] a complex rate-distortion optimization (RDO) process is performed at each level of the quadtree to determine the best encoding mode and partition size for a coding unit (CU). At this RDO analysis, all above-mentioned tools, along with the original texture algorithms, are evaluated to find the best encoding possibility.

One of the main challenges of the 3-D video coding is the high encoding time associated with a large number of modes tested during the encoding process. Some techniques aimed at accelerating the 3D-HEVC depth maps encoding process were proposed in Refs. 10–12.

Zhang et al.[10] proposed a fast depth intramode decision, which is responsible for exploiting the characteristics of depth map block and features of the reference pixels. This mode decision avoids unnecessary intraprediction mode evaluations in most cases, saving the depth coding time. Conceição et al.[11] proposed a block-level decision scheme for 3D-HEVC depth maps encoding that exploits the occurrence of the encoding modes. This scheme avoids mode evaluations with low probability of being chosen. Sanchez et al.[12] proposed a fast pattern selector for DMM to accelerate the depth maps encoding. It uses neighboring information of blocks previously encoded to avoid evaluating all DMM-1 patterns. However, when this technique presents a high time saving, it is accompanied by significant encoding efficiency losses.

In this work, we propose an intercomponent block-level fast coding scheme that uses information of the encoded texture view to reduce the depth maps encoding time. Our scheme decides early in the process if depth maps are encoded using only SKIP or DIS mode or following the traditional flow of 3D-HTM implementation without any simplification. This scheme can reach a high reduction in encoding time with negligible impact on encoding efficiency.

2 3D-HEVC Encoding Structure and Analysis

3D-HEVC provides a flexible encoding quadtree-based structure. To select the best encoding mode and block partition, the latest 3D-HTM encoder implementation performs a complex RDO process to evaluate the possible combinations and choose the one with lowest rate-distortion cost (RD-cost). Similar to the HEVC, a CU can be predicted with a intra- or interframe prediction unit (PU), for both texture and depth information.

The encoder uses the following for every interframe PU: (i) explicit encoding of motion parameters (referenced in this

*Address all correspondence to: Mário Saldanha, E-mail: mrdfsaldanha@inf.ufpel.edu.br

1017-9909/2018/$25.00 © 2018 SPIE and IS&T

work as INTER), (ii) motion merge mode (referenced in this work as MERGE), or (iii) SKIP mode. For an intraframe PU with depth map information, the encoder can select (i) HEVC intraframe prediction modes or DMMs (referenced in this work as INTRA) or (ii) DIS mode. A texture intraframe PU is predicted using only HEVC intraframe prediction modes.

Although this process achieves a high encoding efficiency, a significant increase in the encoder computational effort is noticed. Consequently, one of the main challenges of a 3D-HEVC encoder implementation is to reduce the encoding time associated with a large number of modes tested during the encoding process.

The modes SKIP and DIS were designed to encode regions efficiently with low movement intensity and areas sharing similar pixels values in depth maps, respectively. Therefore, all CUs processed with SKIP and DIS modes are encoded using few bits since such modes do not transmit residual information. Consequently, when the distortion achieved by encoding a CU with SKIP or DIS is small, the RD-cost is also small, increasing the probability of the 3D-HEVC encoder choosing SKIP or DIS to encode the current CU.

2.1 *Statistical Analysis and Discussion*

Figure 1 presents an analysis of the 3D-HEVC encoder behavior. These experiments were performed following the common test conditions (CTCs)[13] for 3-D videos using the 3D-HTM reference software, version 16.0, considering random access (RA) encoder configuration.[13] First, we performed an analysis to evaluate the mode SKIP occurrence in texture encoding considering different quantization parameters (QPs). Figure 1(a) shows that SKIP mode has a high probability of occurrence in texture encoding for all cases evaluated, ranging between 27% (8×8 CUs and $QP = 40$) and 95% (64×64 CUs and $QP = 40$). In addition, for all QPs evaluated, more than 90% of the 64×64 CUs were encoded using SKIP mode.

A second analysis was done aiming to verify the depth maps coding behavior when the correlated reference block in texture view is encoded with SKIP mode. For convenience, let M_{text} be the selected mode by the correlated reference block in texture view. Figure 1(b) shows the modes distribution for depth maps encoding when $M_{\text{text}} = \text{SKIP}$, showing a low probability, <4% (on average), of the modes INTER, MERGE, or INTRA being chosen for encoding a current depth map CU. On the other hand, 93% of 64×64 CUs and more than 80% (on average) of depth maps CUs are encoded using SKIP mode. In addition, the DIS mode is selected more than 15% (on average) for encoding the depth maps CUs. Thus, when $M_{\text{text}} = \text{SKIP}$, the probability of the encoding depth map CU being SKIP or DIS is more than 95%, on average.

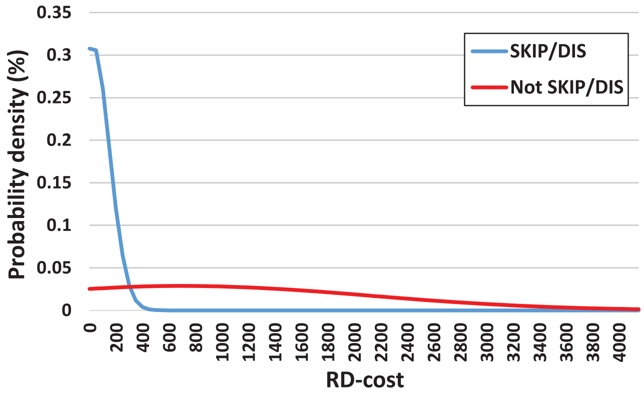

Fig. 2 PDF of encoding depth map CU with SKIP or DIS according to the RD-cost divided by CU width size.

Figure 2 presents the probability density function (PDF) of SKIP or DIS mode (SKIP/DIS) being selected as the best encoding mode for the current CU, when $M_{\text{text}} = \text{SKIP}$, according to the RD-cost (divided by CU width size) obtained by evaluating only these two modes. The division of RD-cost by the CU width size has been selected as an evaluation criterion since larger CU sizes tend to have larger RD-cost values to compute. The analysis presented in Fig. 2 considers 64×64 CUs with the Undo_Dancer video sequence encoding under RA encoder configuration with $QP = 30/39$ ($QP_{\text{texture}}/QP_{\text{depth}}$). Similar results were obtained for the other CU sizes and QPs.

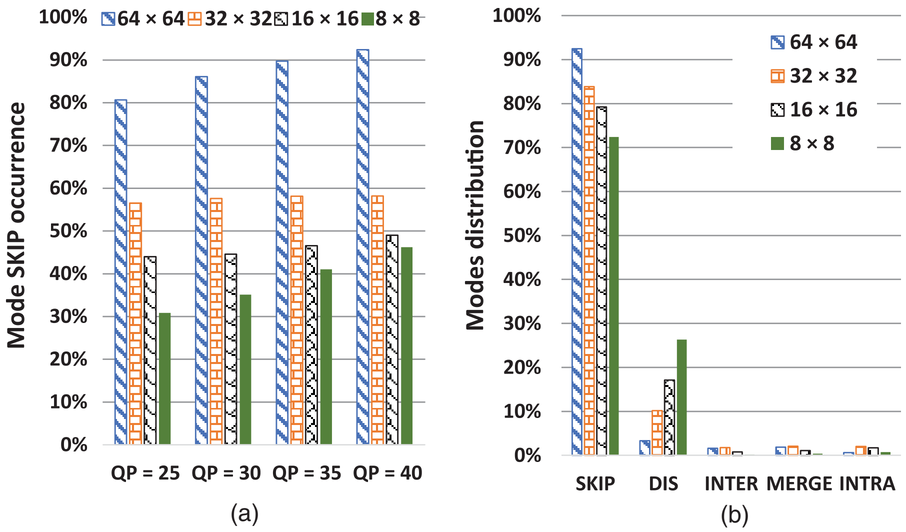

Fig. 1 (a) Occurrence of SKIP mode for texture encoding and (b) modes distribution for depth maps encoding when $M_{\text{text}} = \text{SKIP}$.

These results show a high probability of the current CU being encoded using SKIP or DIS mode for small values of RD-cost when $M_{\text{text}} = \text{SKIP}$. On the other hand, for higher values of RD-cost, SKIP and DIS have no chance of being selected, since large values of RD-cost mean a different encoding mode should encode the CU. Thus, when the correlated block in texture view is encoded with SKIP mode ($M_{\text{text}} = \text{SKIP}$) and the RD-cost obtained by SKIP and DIS is a low value, the probability of it being selected as the best encoding mode is considerably high. Hence, it can be used to perform an early termination decision according to a threshold (TH) criterion, avoiding the remaining encoding mode evaluations.

3 Proposed Solution

This work proposes an early termination block-level decision scheme for the 3D-HEVC depth maps encoding based on the analysis presented in Sec. 2. The proposed early termination scheme is presented in Sec. 3.1, and an analysis of TH is discussed in Sec. 3.2.

3.1 *Early Termination Scheme*

Looking for a light-weight solution for avoiding the excessive encoding mode evaluations in traditional 3D-HTM encoding flow, the early termination block-level decision scheme privileges the modes with a high probability of being selected, with low computational effort and using fewer bits to encode blocks. The flowchart of the proposed scheme is presented in Fig. 3.

Based on our previous analysis, the scheme starts computing the RD-cost for SKIP and DIS modes for the given depth map CU. If the M_{text} (selected mode by the correlated texture block) is SKIP mode, the minimum RD-cost between SKIP and DIS modes is computed. If the minimum RD-cost divided by CU width size is lower than the TH defined by offline analysis (described in Sec. 3.2), then the previous evaluation mode that obtained the lowest RD-cost is selected. Otherwise, further mode evaluations are required, and the encoding process follows without simplification, requiring evaluation of INTER, MERGE, and INTRA encoding modes seeking better encoding results. The decision process is summarized as

$$\text{Decision} = \begin{cases} \text{lowest RD(DIS, SKIP)} & \text{if } \frac{\text{MinRD}}{\text{Wsize}} < TH \\ \text{lowest RD(DIS, SKIP, INTER, MERGE, INTRA)} & \text{otherwise} \end{cases}. \quad (1)$$

In the best case, only SKIP and DIS are required to be evaluated in our scheme. However, in the worst case, the RD-cost calculated by our solution for that depth map CU is the same as the conventional 3D-HTM encoding flow, without inserting additional computational effort than the default approach would require.

3.2 *Threshold Definition*

The TH values lead to light or aggressive solutions regarding both encoding time and video quality. Therefore, we employed an experimental analysis that evaluates some scenarios to explain the impact of the TH variation.

We selected seven THs to be the target of our evaluation, ranging from 100 to 400, with step 50. The corner values were selected by analyzing the PDF presented in Fig. 2, while the step 50 was empirically selected to refine its results because smaller values would lead to a minimal variation. Ten frames of the Undo_Dancer and GT_Fly 3-D video sequences, which were selected randomly among the available sequences, were evaluated using these THs. Only two 3-D video sequences were used in this analysis in an aim to avoid overfitting the designed scheme.

Figure 4 shows the results of the TH scenario evaluations, highlighting the percentage of early termination of encoding mode evaluation according to the well-accepted Bjontegaard Delta rate (BD-rate)[14] criterion. The percentage of early termination was computed as a division of the total cases that our scheme skips of the remaining evaluations by the total cases of $M_{\text{text}} = \text{SKIP}$.

Figure 4 shows that the evaluated THs provide different operation points considering encoding efficiency and with a high percentage of early termination in traditional 3D-HTM encoding flow (more than 96% of all cases evaluated). Based on this analysis, we considered TH = 300 as the best operating point for our early termination scheme presented in Sec. 3.1.

4 Experimental Results and Comparisons

The block-level scheme has been implemented in the 3D-HTM 16.0 and evaluated under the entire CTC at RA

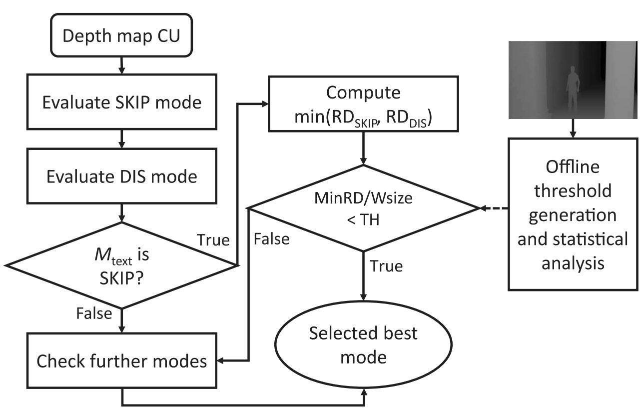

Fig. 3 Dataflow model for early termination scheme.

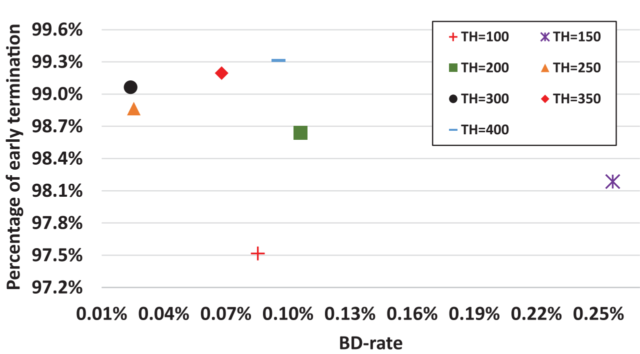

Fig. 4 Percentage of early termination of the encoding mode evaluation according to the BD-rate impact.

Table 1 Results of early termination scheme for CTC under RA encoder configuration.

Video		Skipped evaluation (%)	Synthesized views BD-rate (%)	Depth time saving (%)
TS	Undo_Dancer	96.71	0.133	30.6
	GT_Fly	98.26	0.027	30.5
	Average	97.49	0.080	30.6
Evaluation set	Balloons	97.38	0.208	26.5
	Newspaper_CC	91.53	0.262	23.2
	Kendo	95.09	0.295	23.2
	Poznan_Hall2	99.25	0.386	30.2
	Poznan_Street	98.38	0.145	30.0
	Shark	96.36	0.125	28.4
	Average	96.33	0.237	26.9

Note: TS, training sequences.

encoder configuration, using $\mathrm{TH} = 300$. The obtained results are presented in Table 1, showing separately the results obtained in the video sequences used for training and evaluation. From now, only the results obtained in the video sequences for evaluation will be discussed.

Our scheme is capable of reaching an average time saving of 26.9% in depth maps coding, which varies from 23.2% to 30.2% according to the encoded video sequence. As a drawback, a small increase in BD-rate occurs, varying from 0.125% to 0.386% with an average result of 0.237%. These variations happen mainly because there are differences among the encoding sequences, and higher resolution video sequences tend to select more SKIP than other modes in the texture coding. The significant time saving results are obtained because our scheme is capable of evaluating only SKIP and DIS in 96.33% (on average) of the cases when $M_{\text{text}} = \text{SKIP}$.

Table 2 compares the related works previously presented with our work. The fast intramode decision proposed in Ref. 10 achieves 27.9% of depth time saving with a BD-rate increase of 1.030%. The work in Ref. 11 reaches a depth time saving of 33.7% with 0.409% of BD-rate increase. In Ref. 12, the fast pattern selector for DMM achieves 1.3% of time saving considering texture and depth information with 0.072% of BD-rate increase. Therefore, our work provides time saving results as good as the related work with only a small impact on the encoding efficiency.

5 Conclusions

This letter presented a block-level fast coding scheme for depth maps coding. Through our analysis, most of the selection in texture CU is SKIP mode; moreover, when it happens, there is a high probability (over 95%) of the correlated depth map CU selecting SKIP or DIS as the best encoding mode. Consequently, we designed our block-level scheme that is capable of avoiding the remaining depth maps coding tools, when SKIP and DIS modes have a high chance of being selected, and obtained a low RD-cost result. Experimental results demonstrated that our scheme was capable of achieving a time saving of 26.9% at depth maps coding with a drawback in BD-rate of only 0.237%, surpassing related works' results.

Table 2 Comparison with related works.

Work	BD-rate (%)	Time saving	
		Texture and depth (%)	Depth only (%)
Ref. 10	1.030	—	27.9
Ref. 11	0.409	13.8	33.7
Ref. 12	0.072	1.3	—
This	0.237	13.4	26.9

Acknowledgments

This letter was performed in cooperation with Hewlett-Packard Brazil Ltd. using incentives of Brazilian Informatics Law (Law No. 8.248 of 1991). Authors also would like to thank CAPES, CNPq, and FAPERGS Brazilian research agencies for supporting the development of this work.

References

1. K. Muller et al., "3D high-efficiency video coding for multi-view video and depth data," *IEEE Trans. Image Process.* **22**(9), 3366–3378 (2013).
2. P. Kauff et al., "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Process. Image Commun.* **22**(2), 217–234 (2007).
3. G. Sullivan et al., "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1649–1668 (2012).
4. P. Merkle et al., "Depth intra coding for 3D video based on geometric primitives," *IEEE Trans. Circuits Syst. Video Technol.* **26**(3), 570–582 (2015).
5. H. Liu and Y. Chen, "Generic segment-wise DC for 3D-HEVC depth intra coding," in *Proc. of the IEEE Int. Conf. on Image Processing (ICIP)*, pp. 3219–3222 (2014).
6. J. Y. Lee, M. W. Park, and C. Kim, "3D-CE1: depth intra skip (DIS) mode," Document JCT3V-K0033, Geneva, Switzerland (2015).
7. E. Mora, M. Cagnazzo, and B. Pesquet-Popescu, "Initialization, limitation, and predictive coding of the depth and texture quadtree in 3D-HEVC," *IEEE Trans. Circuits Syst.* **24**(9), 1554–1565 (2014).
8. L. Shen, Z. Zhang, and Z. Liu, "Effective CU size decision for HEVC intracoding," *IEEE Trans. Image Process.* **23**(10), 4232–4241 (2014).
9. JCT-3V, "3D-HEVC test model," https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-16.0/ (September 2017).
10. H. Zhang et al., "Efficient depth intra mode decision by reference pixels classification in 3D-HEVC," in *IEEE Int. Conf. on Image Processing (ICIP)* (2015).
11. R. Conceição et al., "Complexity reduction for 3D-HEVC depth map coding based on early skip and DIS scheme," in *IEEE Int. Conf. on Image Processing (ICIP)* (2016).
12. G. Sanchez et al., "DFPS: a fast pattern selector for depth modeling mode 1 in three-dimensional high-efficiency video coding," *J. Electron. Imaging* **25**(6), 063011 (2016).
13. F. Bossen, "Common test conditions and software reference configurations," JCT-VC document L1100 (2013).
14. G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," Technical Report VCEG-M33, ITU-T SG16/Q6, Austin, Texas (2001).