Depth Modeling Modes Complexity Control System for the 3D-HEVC Video Encoder

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

***Abstract*—This paper presents a complexity control system for depth maps intra-frame prediction of the 3D-High Efficiency Video Coding (3D-HEVC) standard. The proposed system uses a Proportional-Integral-Derivative controller over the Simplified Edge Detector heuristic to skip the Depth Modeling Modes (DMMs) evaluation dynamically according to a defined target rate. When analyzing the proposed system under Common Test Conditions, the proposed controller stabilizes the system to the target rate (i.e., the percentage of DMMs evaluation) after encoding a few frames, with negligible encoding efficiency impacts. The BD-rate degradation varies from 0.50% to 0.20%, on average, when the target rates vary from 5% to 15%. These target rates imply in an aggressive reduction in the DMMs evaluations, skipping the DMMs from 85% to 95% of the cases.**

Keywords—3D-HEVC; Intra-Frame Prediction; Depth Maps; Complexity Control; Complexity Reduction

I. Introduction

The Multiview Video plus Depth (MVD) [1] is one of the most advanced methods for 3D video representation. In the MVD format, only a limited number of views and their corresponding depth maps are captured from different angles (typically from cameras equipped with infrared sensors), encoded, multiplexed into a 3D video bitstream, and transmitted. With the addition of geometric information from depth maps, it is possible to synthesize an arbitrary number of intermediary views by applying view synthesis techniques at the decoder side, such as Depth-Image-Based Rendering (DIBR) [2].

The MVD data format is adopted by the state-of-the-art 3D-High Efficiency Video Coding (3D-HEVC) standard [3]. The 3D-HEVC is an extension of High Efficiency Video Coding (HEVC) standard, and it can achieve a high compression rate providing good video quality. In MVD, depth samples provide the distance between a camera and objects as a grayscale image, while texture samples provide the colors of the image.

The characteristics of depth maps are quite different from texture frame since depth maps have large homogeneous regions delimited by sharp edges, whereas texture frames contain a complex content with sudden variations in sample values. Although end-users do not view depth maps, the geometrical information is crucial for generating the synthesized views. Besides, distortion on edges information during the encoding process may cause inaccurate representation between pixels of background and foreground after view synthesis process. Based on this fact, preserving the edges information to provide high quality in synthesized views can become the most critical task for depth map coding in 3D-HEVC. To deal with this challenge, 3D-HEVC introduces new coding tools for a better exploration of depth maps proprieties, such as Depth Modeling Modes (DMM) [4], Segment-wise DC Coding (SDC) [5], Depth Intra Skip (DIS) [6], and View Synthesis Optimization (VSO) [7].

Since HEVC intra-frame prediction presents limited efficiency when encoding sharp edges and can produce undesired artifacts in the synthesized views, DMMs are introduced as additional intra-frame prediction modes for coding 3D-HEVC depth maps. DMMs are composed of two encoding modes: DMM-1 and DMM-4 that evaluate wedgelets and contours, respectively. However, as a drawback of enhancing synthesized views quality, DMM modes cause a significant increase of 3D-HEVC encoder complexity. Evaluations using the Common Test Conditions (CTC) under the all intra-frame configuration [8] showed that depth maps complexity is 5.8 times higher than texture coding and the DMMs processing use 35.34% of the computation time of the 3D-HEVC depth maps encoder [9]. Therefore, new techniques have been proposed to reduce the complexity of the depth maps intra-frame prediction.

Several works have already presented solutions to reduce the prediction complexity of depth maps intra-frame such as [10]-[14]. In [10], DMMs evaluations are skipped if the variance of the encoding block is higher than a threshold or if the first mode in intra-frame prediction list is the planar mode. Saldanha et al. [11] propose to compute the encoding block border gradient and specifically reduce the complexity of DMM-1 algorithm by only evaluating wedgelets starting in a high gradient value. In our previous work [12], a Simplified Edge Detector (SED) algorithm is proposed to skip DMM evaluations when the encoding block is classified as homogeneous. This classification is performed by analyzing the highest difference of the four corners samples of the block.

Fu et al. [13] proposed a technique to reduce the DMM-1 complexity based on the estimation of sharp edges by pre-defined regions; the technique only evaluates search patterns in the most probable region. Zhang et al. [14] proposed an intra-mode decision algorithm to reduce the number of HEVC intra modes evaluated and to skip DMM evaluations. Their algorithm is based on thresholds taking advantage of Rate-Distortion cost (R-D cost) information in the rough mode decision.

Although the mentioned works can meaningfully reduce the 3D-HEVC coding complexity, there are applications where complexity control is necessary. For the 3D video systems popularization, it is desired that many different devices could be capable of encoding and decoding 3D videos. Since different

ISBN 978-0-9928626-7-1

devices have a different computational capacity and different power restrictions, a complexity control system is important to allow that all devices could be able to provide a good encoding efficiency for any device. Moreover, considering that many battery-based devices will have an embedded 3D-HEVC encoder, then a complexity control system is interesting to allow a dynamic adaptation according to the available battery.

Some complexity control solutions can be found for the 2D version of HEVC. Correa et al. [15] propose a complexity control for 2D-HEVC, which is based on the coding tree structures adjustment according to a target complexity. In [16] and [17] complexity control techniques were proposed to adjust 2D-HEVC encoding complexity that limits the Coding Tree Units (CTU) maximum depth. In [16] a region-of-interest model is established to define different weights for regions according to their importance, whereas in [17] the certain saliency threshold is used to decide whether the Coding Unit (CU) should be split into smaller CUs.

To the best of our knowledge, there is no published work presenting methods to deal with the complexity control for the 3D-HEVC standard. Therefore, this paper presents the first complexity control system for the 3D-HEVC standard that targets only the DMMs processing. In this proposed solution, the SED algorithm [12] is applied to decide for an encoding depth block if DMM evaluations should be skipped according to a target computational complexity.

II. SED Algorithm & Motivational Analysis

In our previous work [12], we demonstrated that the highest difference of the four corner samples of a block (called D_{max}) contains significant information to classify this block as homogeneous or edge. Fig. 1 depicts a slice from *Undo_Dancer* depth map with four edges blocks and four homogeneous blocks detached. The D_{max} values of those detached blocks are provided below. One can notice that edge blocks contain D_{max} with high values, while homogeneous regions contain low D_{max} values.

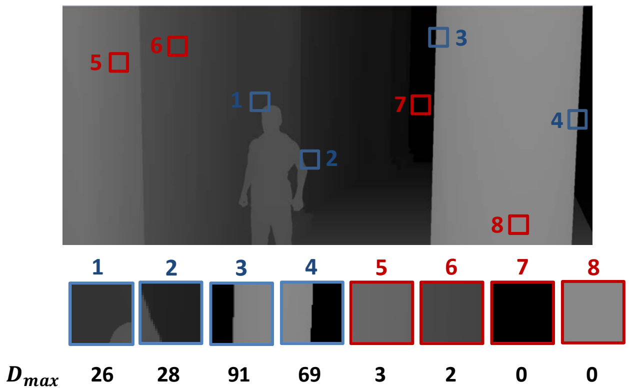

Fig. 1. Depth map example with detached edges (1-4) and homogeneous regions (5-8).

Motivated by this analysis, the Simplified Edge Detector (SED), whose flowchart is presented in Fig. 2, compares the D_{max} against a fixed threshold to perform this classification. When the D_{max} is higher than the threshold, then the encoding block is classified as an edge block, and DMMs evaluation is necessary. Otherwise, the encoding block is classified as a homogeneous region and the DMMs evaluations are skipped.

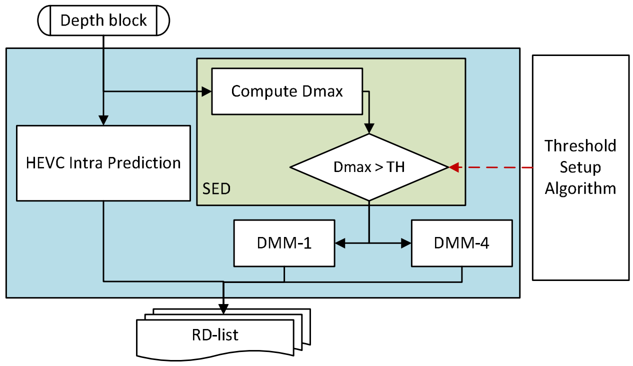

Fig. 2. Flowchart of the Simplified Edge Detector (SED).

The SED classification proposed in [12] is based on a fixed threshold value according to the video resolution and block size. The thresholds were defined in an offline statistical analysis regarding the resolutions 1024×768 and 1920×1080. Then, its decision cannot be directly expanded to lower video resolutions such as CIF (352×288) and SD (720×480) or higher video resolutions such as 4K (3840×2160) and UHD (7680×4320).

Moreover, the performance of SED decision depends on the encoding video characteristic, which introduces non-determinism regarding the complexity of the DMMs. Fig. 3 illustrates this idea presenting the D_{max} heat map for two frames of distinct 1080p video sequences ((a) *Shark* and (b) *PoznanStreet* [8]) considering 8×8 blocks, where red colored blocks indicate the highest values of D_{max}, green regions means intermediary values, and blue regions show the lowest values.

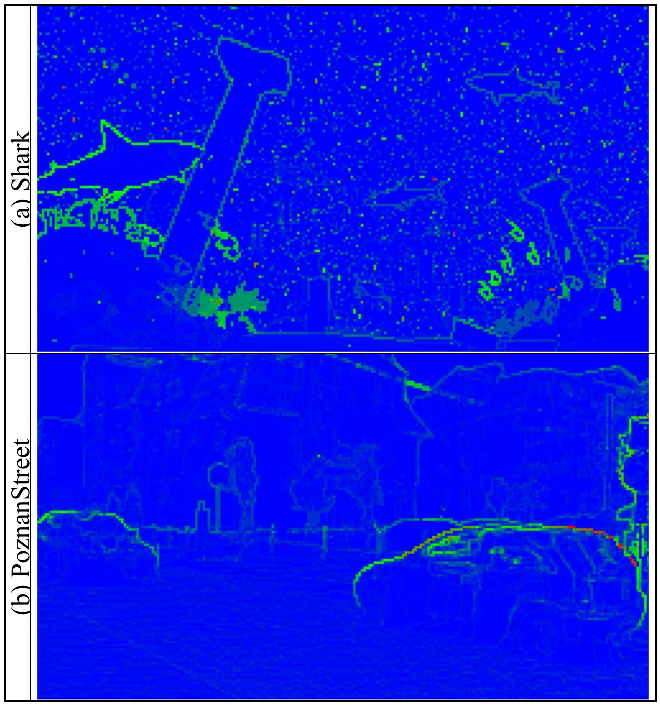

Fig. 3. SED heat map for the first frame central view of (a) Shark and (b) Poznan_street video sequences.

It is notable that if SED uses a fixed threshold as proposed in [12], then a variable number of DMMs should be skipped for different types of videos since each video has a different quantity of borders and details as exemplified in Fig. 3.

ISBN 978-0-9928626-7-1 © EURASIP 2017

Fig. 4 shows an example of encoding the 100 first frames of the CTC [8] sequences using the fixed thresholds SED [10]. Notice that each video has a different profiling as previously highlighted in the example shown in Fig. 3. Moreover, in a short execution of a video, there are frames where DMMs are evaluated from 3% to 20% of cases. Besides, different block sizes have a different DMMs evaluation complexity when comparing Fig. 4(a) (4×4 blocks) and Fig. 4(b) (32×32 blocks), when SED threshold is fixed. For example, *Shark* video sequence achieves about 20% of DMMs evaluated considering 32×32 blocks, whereas, for 4×4 blocks, the highest DMMs evaluation achieved is about 6%. Therefore, with the different profiling of each video and the variety of available devices providing different computational capacity and power restrictions for encoding 3D videos, is important to create complexity-controlled solutions to allow performing DMMs operation for different encoding video characteristics and without surpassing the system resources limit. In the case of the system employs fewer evaluations than its capacity, it is possible to compensate it by applying further assessments in the next frames, increasing the encoding quality.

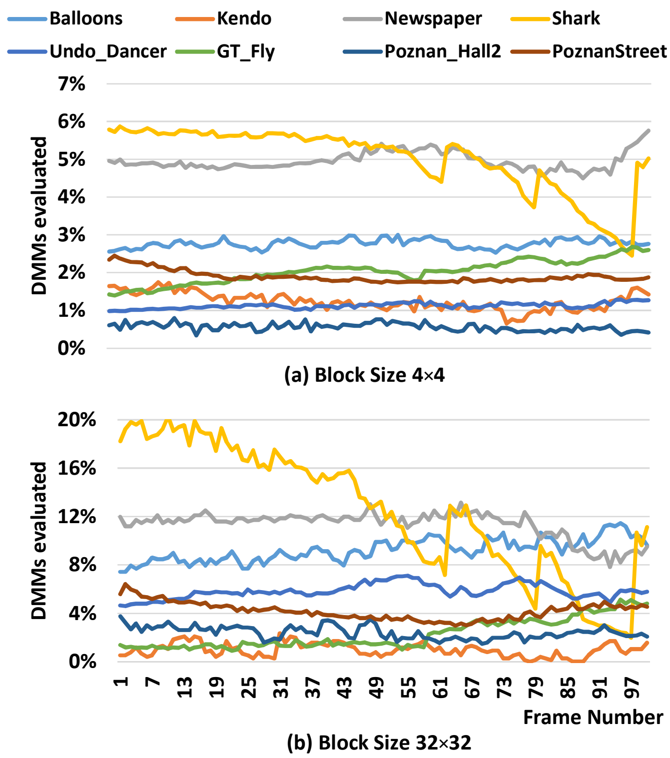

Fig. 4. SED DMMs evaluation for (a) 4×4 blocks and (b) 32×32 blocks.

III. Proposed Control System

Based on the analysis of Section II, we proposed a Proportional-Integral-Derivative (PID) controller to adapt the SED threshold for controlling the DMMs evaluation according to the desired target complexity dynamically. The diagram of the proposed control system is displayed in Fig. 5 and it is replicated for each block size. The delay in Fig. 5 means the value from the previously encoded frame. Complementing the control block diagram, Fig. 6 presents the pseudo-code of our control system.

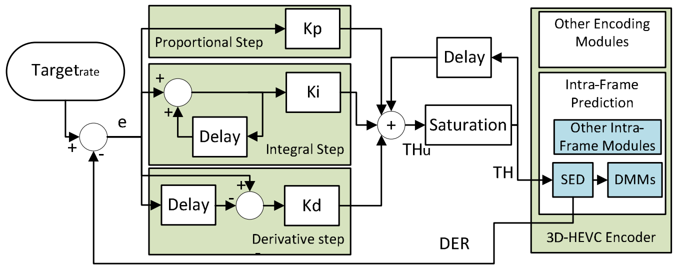

Fig. 5. Diagram of the proposed PID control system.

```
TH              ← initial threshold value
Targetrate      ← desired rate
N               ← number of frames
e_sum           ← 0
last_e          ← 0
for fn = 1 to N
    Encode frame using TH
    DER ← percentage of DMMs evaluated
    e ← Targetrate – DER
    e_sum ← e_sum + e
    THu ← Kp×e + Ki×e_sum + Kd×(e - last_e) + TH
    if (THu < 1)
        TH ← 1
    else
        TH ← THu
```

Fig. 6. Algorithm of the control system.

A target DMMs evaluation rate ($Target_{rate}$) should be selected as the input of the system. For example, when a target of 5% is selected, then 95% of the DMMs evaluation should be skipped. This $Target_{rate}$ can be selected according to the available resources of the system and it can be changed from frame to frame, according to the available resources (e.g., it can be reduced if the system is running with low battery). Let *fn* be the number of the current frame being encoded, then the error of *fn* (*e[fn]*) can be computed using (1), where the $Target_{rate}$ is subtracted from DMM Evaluation Rate (*DER*) of the previous frame. The DER is always delivered by SED when a frame encoding is finished.

$$e[fn] = Target_{rate} - DER \tag{1}$$

The computed error follows three paths (i) a proportional step, (ii) an integration step, and (iii) a derivation step, whose results are added to the threshold of the previous frame (*TH[fn-1]*) in (2) aiming to generate the unbounded threshold (*THu*).

$$THu[fn] = Kp \times e[fn] + Ki \times \sum_{i=0}^{n} e[i] + Kd \times (e[fn] - e[fn-1]) + TH[fn-1] \tag{2}$$

In proportional step, the error value is multiplied only by a *Kp* constant. The proportional step results in a change in the threshold that is proportional to the error obtained by current frame evaluation.

In integration step, all previous errors values are added and multiplied by a *Ki* constant. The integral step is necessary for eliminating the residual evaluation error of the system. On a first analysis, the integration step may sound a complex operation because all previous errors need to be added. However, this operation can be reduced to a single sum multiplied by *Ki* constant because all errors until the current frame have already

ISBN 978-0-9928626-7-1

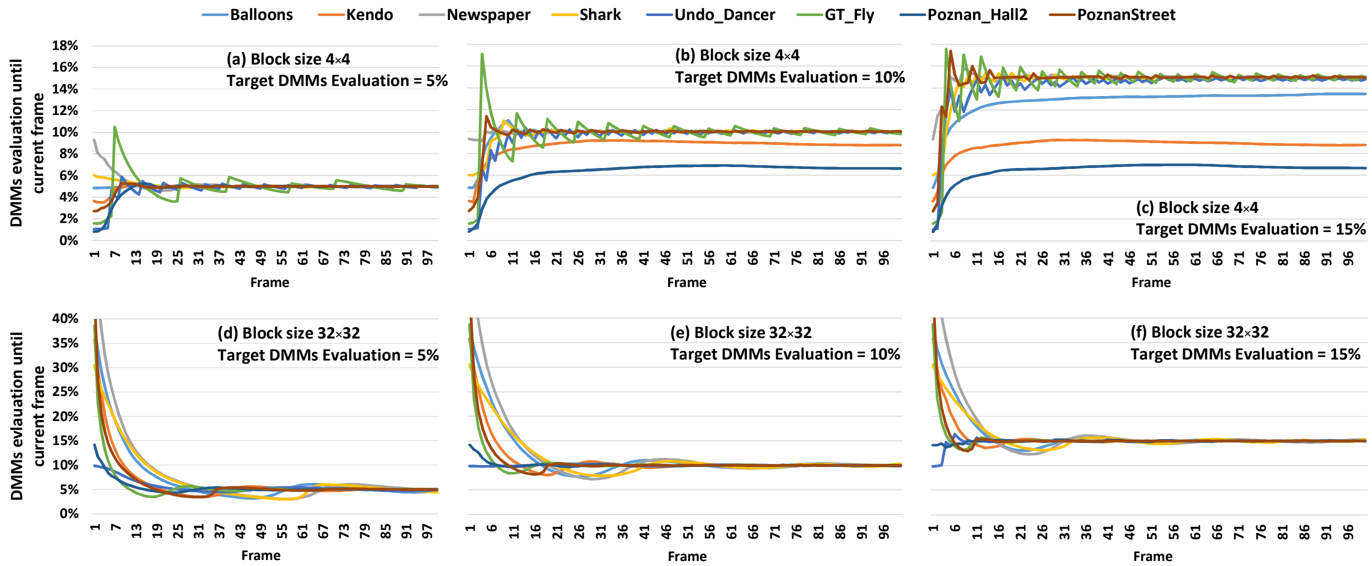

Fig. 7. Complexity control evaluations for some scenarios and target evaluations.

been added and this information is already available, as one can see in the algorithm displayed in Fig. 6.

In derivative step, the current error is subtracted from the error obtained in the previous frame and multiplied by a *Kd* constant. The derivative step helps to predict the system behavior fast, enabling a fast control actuation.

By adding these resulting values with the last frames threshold, the unbounded threshold (*THu*) is generated. A lower bound saturation is applied before delivering this threshold to SED. This saturation prevents the threshold to go below one, when 100% of DMMs evaluations would be performed. Moreover, in case $D_{max} = 0$, the region has the highest probability to be a homogeneous region, then the DMMs should be always skipped.

IV. Results & Discussion

The proposed control system was implemented in 3D-HEVC Test Model (3D-HTM) version 16.0. The CTC videos sequences were evaluated in all-intra configuration for three DMM evaluation target rates: 5%, 10%, and 15%. These target rates considered the information presented in previous experiments and aggressively reduces the number of evaluated DMMs. Fig. 7 depicts the DMMs use rate convergence for the desired target rate. This experiment considered the first 100 frames of the CTC depth video sequences for block sizes 4×4 and 32×32 (considering only the central view). Several experiments were performed in *Balloons* video sequence to select the constants *Kp*, *Ki,* and *Kd* that well fits the system without leading to an unstable behavior. Therefore, they were defined, respectively, with the values 10, 10, and 1.

Analyzing Fig. 7, one can notice that the DMM evaluations for smaller block sizes show a slower convergence than for larger blocks. However, after a short period, all videos have converged to the target evaluation rate. In some cases, such as the encoding *Kendo* video sequence with a target DMMs evaluation of 15%, the assessments are limited to a maximum of 7.5%. This limitation occurs because the proposed control system has a lower bound saturation that does not allow the encoder to perform the DMMs evaluation for blocks that contains a $D_{max} = 0$, which has the highest chance to be a homogeneous block. Therefore, the maximum DMMs evaluation in our system is limited to the video content.

The proposed complexity control system causes different impact in the encoding efficiency according to the target evaluation rate. When a higher number of DMMs evaluation is targeted, a better encoding efficiency should be obtained. Table I shows the encoding efficiency regarding Bjontegaard Delta-Rate (BD-rate) when compared with the original 3D-HTM 16.0. It demonstrates the impact caused by the three targets selected in our evaluation. In our assessment, choosing a target of 5% causes an average increase on BD-rate of 0.50%. This value decreases as the target DMM evaluation increases, reaching an average of 0.20% when using 15% of target DMMs evaluation.

The worst result is noticed in *Newspaper* video sequence with a BD-rate increase that ranges from 1.11% up to 0.62% regarding complexity targets. This impact occurs because the original video sequence has a poor depth map quality and regions where the DMMs tend to be chosen (such as sharp edges) are composed of many distortions, causing more coding efficiency losses. When analyzing video sequences with a good depth map quality, e.g., *GT_Fly*, one can notice that the BD-rate increase is negligible.

These impacts are completely acceptable since more than 85% of DMMs evaluations were skipped. Considering that the DMMs uses at least 35.34% of the 3D-HEVC depth maps intra-frame encoder time as described in [9], the proposed solution can drastically reduce the encoder complexity, and maintain this complexity over a well-defined range of values. Table II displays the depth time saving of our solution in the three DMMs complexity targets. Notice that our solution achieved an average time saving of 29.9%, 27.9% and, 19.3%, according to the selected target. Although the DMMs evaluations are stabilized as demonstrated in Fig. 7, there is still significant variation in the time saving as Table II shows. It occurs because the evaluated

ISBN 978-0-9928626-7-1 © EURASIP 2017

videos contain different content characteristics. Therefore, other tools used in depth maps coding are responsible for different complexities (that are not stabilized), since we use the system of control only for DMMs.

TABLE I. BD-RATE RESULTS WITH DIFFERENT COMPLEXITY TARGETS.

Videos	BD-rate		
	Target = 5%	Target = 10%	Target = 15%
Balloons	0.59%	0.35%	0.27%
Kendo	0.57%	0.36%	0.25%
Newspaper	1.11%	0.78%	0.62%
GT_Fly	0.36%	0.16%	0.10%
Poznan_Hall2	0.39%	0.22%	0.13%
PoznanStreet	0.16%	0.09%	0.08%
Undo_Dancer	0.23%	0.08%	0.06%
Shark	0.60%	0.22%	0.10%
Average	**0.50%**	**0.28%**	**0.20%**

In this evaluation, *Newspaper* video sequence has achieved the highest gains regarding timesaving. Again, it occurs because of the original depth map low quality. The timesaving obtained by this sequence ranges from 24.5% to 35.4%, according to the target DMM evaluation.

TABLE II. DEPTH TIME SAVING ACCORDING TO COMPLEXITY TARGETS.

Videos	Depth coding time saving		
	Target = 5%	Target = 10%	Target = 15%
Balloons	32.0%	31.8%	21.5%
Kendo	32.3%	31.4%	20.6%
Newspaper	35.4%	34.9%	24.5%
GT_Fly	29.2%	25.9%	18.7%
Poznan_Hall2	26.2%	22.5%	13.1%
PoznanStreet	28.5%	26.5%	20.4%
Undo_Dancer	28.0%	25.2%	17.6%
Shark	27.3%	25.0%	18.2%
Average	**29.9%**	**27.9%**	**19.3%**

V. CONCLUSIONS

This paper presented a complexity control system for Depth Modeling Modes (DMMs) in 3D-HEVC depth maps intra-frame prediction. The basic algorithm of the designed system is the Simplified Edge Detector (SED) that compares the four corners samples of the encoding block with a threshold and determine if DMMs should be evaluated or not. After several evaluations, one can conclude that the number of evaluations varies significantly between frames and it may be prohibitive for applications with limited performance and/or battery. Therefore, our system was designed to manage this problem and keeps the DMMs evaluation under a target DMMs evaluation rate. The control system uses a PID controller to determine the threshold of SED algorithm dynamically. Experimental results demonstrate that our control system is capable of stabilizing the system to perform the target evaluation rate in a few frames with minor impact on the encoding efficiency. With a target rate of 5% (skipping 95% of the DMMs evaluations), the depth coding execution time is reduced by 29.9% and the BD-rate is increased in less than 1%.

ACKNOWLEDGMENT

This paper was achieved in cooperation with Hewlett-Packard Brazil Ltda. using incentives of Brazilian Informatics Law (Law nº 8.248 of 1991). Authors also would like to thank CAPES (processes 88881135737/2016-01 and 88881119481/2016-01), CNPq (processes 309707/2015-3 and 486136/2013-2) and FAPERGS (process 16/2551-0000241-0) to support the development of this work.

REFERENCES

[1] 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," Image Communication, v. 22, n. 2, pp. 217-234, Feb. 2007.

[2] C. Fehn. "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," SPIE, Stereo-scopic Displays and Virtual Reality Systems, v. 5291, pp. 93-104, May 2004.

[3] 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*, v. 22, n. 9, pp. 3366-3378, Sep. 2013.

[4] P. Merkle, K. Muller, D. Marpe, T. Wiegand. "Depth Intra Coding for 3D Video based on Geometric Primitives," *IEEE Transactions on Circuits and Systems for Video Technology*, v. 26, n. 3, pp. 570-582, Feb. 2015.

[5] H. Liu, Y. Chen. "Generic segment-wise DC for 3D-HEVC depth intra coding," *IEEE International Conference on Image Process (ICIP)*, pp. 3219-3222, 2014.

[6] J. Lee, M. Park, C. Kim. "*3D-CE1: Depth Intra Skip (DIS) Mode*," *document JCT3V-K0033*, Geneva, Switzerland, Feb. 2015.

[7] K. Muller, P. Merkle, G. Tech, T. Wiegand. "3D video coding with depth modeling modes and view synthesis optimization," *Asia-Pacific Signal & Information Processing Association Annual Summit and Conference* (APSIPA ASC), pp. 1-4, 2012.

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

[9] G. Sanchez, R. Cataldo, R. Fernandes, L. Agostini, C. Marcon. "3D-HEVC depth maps intra prediction complexity analysis", IEEE International Conference on Electronics, Circuits, and Systems, 2016.

[10] Z. Gu, J. Zheng, N. Ling, P. Zhang. "Fast Depth Modeling Mode Selection for 3D HEVC Depth Intra Coding," *IEEE International Conference on Multimedia and Expo Workshops* (ICMEW), pp. 1-4, 2013.

[11] 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.

[12] G. Sanchez, M. Saldanha, G. Balota, B. Zatt, M. Porto, L. Agostini. "Complexity reduction for 3D-HEVC depth maps intra-frame prediction using simplified edge detector algorithm," *IEEE International Conference on Image Processing* (ICIP), pp. 3209-3213, 2014.

[13] C. Fu et al. "Fast Wedgelet Pattern Decision for DMM in 3D-HEVC," *IEEE International Conference on Digital Signal Processing (DSP)*, pp. 447-481, 2015.

[14] H. Zhang et al. "Adaptive fast intra mode decision of depth map coding by Low Complexity RD-Cost in 3D-HEVC," *IEEE International Conference on Digital Signal Processing (DSP)*, pp. 487-491, 2015.

[15] G. Correa, P. Assunção, L. Agostini, L. Cruz. "Complexity control of HEVC through quadtree depth estimation," *The International Conference on Computer as a Tool* (EUROCON), pp. 81-86, 2013.

[16] X. Deng, M. Xu, S. Li, Z. Wang. "Complexity control of HEVC based on region-of-interest attention model," *Visual Communications and Image Processing Conference* (VCIP), pp. 225-228, 2014.

[17] N. Shan, W. Zhou, Zhemin Duan. "Complexity control for HEVC by visual saliency detection," *IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1-5, 2016.