Joint Complexity and Rate Optimization for 3DTV Depth Map Encoding

Sebastiaan Van Leuven*, Hari Kalva†, Glenn Van Wallendael*, Jan De Cock*, and Rik Van de Walle*
*Multimedia Lab, Dep. of Electronics and Information Systems, Ghent University-IBBT, B-9050 Ledeberg Ghent, Belgium
Email: {sebastiaan.vanleuven; glenn.vanwallendael; jan.decock; rik.vandewalle}@ugent.be
†Dept. of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL, United States
Email: hari.kalva@fau.edu

***Abstract*—Current research towards 3D video compression within MPEG requires the compression of three texture and depth views. To reduce the additional complexity and bit rate of the depth map encoding, we present a fast mode decision model based on previously encoded macroblocks of the texture view. Meanwhile we present techniques to reduce the rate based on predicting syntax elements from the corresponding texture view. The proposed system is able to get a reduction in complexity of 71.08% with an average bit rate gain of 4.35%.**

I. Introduction

Currently, MPEG is standardizing 3D Video [1]. The goal is to improve compression compared to the Multiview Video Coding (MVC) extension of MPEG-4/AVC -H.264 (annex H) [2]. Next generation 3D devices are targeted. Such devices, like autostereoscopic displays, might require depth information to generate a large amount of intermediate views using view interpolation [3], [4]. Also more common stereo displays will benefit from the standard. However, these displays might not require depth information. Consequently, the depth is encoded solely for a limited set of users. Therefore the overhead, both in complexity and rate distortion (RD), should be limited. This preliminary research is based on the MVC extension and is aimed to reduce both the complexity and bit rate in H.264/AVC based 3D video standardized encoders.

Current research is focusing on 3D video encoding using the MVC extension, a base view is encoded as a regular H.264/AVC (AVC) bitstream, while additional views are encoded using the decoded output of the base view as an additional prediction frame. Additionally, depth can be encoded although this requires a relatively high complexity. In [5] the complexity of the depth map encoding is reduced, but the RD is not improved. An even more complex scheme is presented in [6], where the views are warped before prediction. However, such system require an even higher complexity, and also hardware design is more complex. Moreover, the applied warping should also be standardized. Optimizations have been proposed to reduce the syntax [7], however, the complexity is not reduced. Therefore we propose a model to reduce both depth map encoding complexity and the bit rate, while the only minor implementation changes are required.

II. Methodology

The MPEG CfP considers a three and two view case, for eight sequences and four rate points per sequence (resulting in 64 encoded bit streams). Our system is evaluated for the three view case, therefore the data of two sequences (*Poznan_Hall2* and *Kendo*) is analyzed for the highest and lowest rate point. This analysis results in our proposed model. Study of the syntactical overhead showed potential RD improvements. Both the reduced complexity and syntactical improvements have been evaluated with eight HD sequences (*Poznan_Hall2*, *Poznan_Street*, *Undo_Dancer*, *GT_Fly*, *Kendo*, *Balloons*, *Lovebird1*, and *Newspaper*) and four rate points for the three view case. The test conditions (GOP size, quantization, intra period) from the MPEG CfP are used.

III. Proposed Method

A. Complexity Reduction

An analysis of the mode distribution of the depth map results in the probability that a mode is selected in the depth view based on the mode in the co-located macroblock of the corresponding texture view. This probability is given by: $p = P(MODE_{Depth}|MODE_{Tex})$. Where $MODE_{Depth}$ is the macroblock mode used in the depth and $MODE_{Tex}$ the macroblock mode used in the co-located macroblock in the corresponding texture view. The analysis shows that when $MODE\_Skip$ or $MODE\_16\times16$ are selected in the texture, this mainly results in $MODE\_Skip$ or $MODE\_16 \times 16$ being selected in the depth. Based on this analysis, the set of probable modes for the depth is given by :

$$MODE_{Depth} = \begin{cases} S & \text{if } MODE_{Tex} \in S \\ A & \text{if } MODE_{Tex} \notin S \end{cases}$$

Where S is the subset of unpartitioned modes $S = \{MODE\_Skip, MODE\_16 \times 16\}$ and A the set of all modes $A = \{MODE\_Skip, MODE\_16\times16, MODE\_16\times 8, MODE\_8 \times 16, MODE\_8 \times 8, MODE\_Intra\}$. After evaluating the probable modes, the most RD-optimal mode is selected. The analyzed sequences achieve a high accuracy, 96.76%, using this model (Table I).

B. Rate Distortion Improvement

Due to the low bit rate of depth maps, syntactical data has a high impact. For the low rate points of the MPEG CfP sequences, the syntactical data in depth maps accounts for approximately 50%. Next to the complexity reduction, our

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

TABLE I
ACCURACY OF THE PROPOSED MODEL.

	Poznan_Hall2		*Kendo*	
	R1	R4	R1	R4
accuracy (%)	98.76	96.34	97.59	94.37

TABLE II
NUMBER OF MACROBLOCKS FOR WHICH THE PROPOSED MODEL RESULTS IN LESS SYNTACTICAL DATA.

	Poznan_Hall2		*Kendo*	
	R1	R4	R1	R4
Accuracy (%)	37.09	59.48	59.54	68.09

proposed technique is also able to reduce the *mb_type* signaling. The syntax element *mb_type* indicates the macroblock type and *mb_skip_flag* indicates whether a macroblock is a skip macroblock. Since *mb_type* of the depth map is based on the mode of the texture the *mb_type* does not have to be transmitted for every macroblock. When $MODE_{Tex} \in S$, the depth map macroblock type can be determined based on the *mb_skip_flag* in the depth. When *mb_skip_flag* $= 1$, the macroblock type of the depth is skipped, otherwise it will be $MODE\_16 \times 16$. Consequently, *mb_type* can be omitted for such cases. Table II shows the accuracy of this reduction for the analyzed sequences.

Furthermore, for each macroblock an *end_of_slice_flag* is transmitted, indicating if the current macroblock is the last macroblock of the slice. We propose to use the same slice configuration for texture and depth, so the data corresponding to the *end_of_slice_flag* should not be transmitted. Additionally, all data corresponding to chroma prediction can be neglected. This include the *intra_chroma_pred_mode* flag and chroma residual data.

IV. EXPERIMENTAL RESULTS

Results are calculated on a cluster consisting of 2.27 GHz dual quad core nodes. The software is compiled in g++ 4.1.2 for 64 bit. Each sequence is executed as a single thread. The complexity reduction is expressed as the time saving (TS) for encoding with our proposed model (T_{Fast}) compared to the reference encoder ($T_{Original}$), and is given by (1).

$$TS\ (\%) = \frac{T_{Original}\ (ms) - T_{Fast}\ (ms)}{T_{Original}\ (ms)} \qquad (1)$$

Results of the complexity measurements can be found in Table III, the average complexity of four rate points per view is given because the complexity reduction is relatively constant. The overall complexity reduction is 71.08%, so only 28.92% of the complexity is required compared to the original encoder. Since the enhancement layer complexity depends on $MODE_{Tex}$, the complexity reductions are not constant.

Metrics to evaluate the quality of multiple 3D depth maps are not yet commonly used and available. Therefore, the proposed system is evaluated by the RD. Significant bit rate gains (up to 13.52%) are noticed, while a comparable quality is maintained. Worst case, a reduction of 2 dB is measured(33.1 dB to 31.1 dB for *Newspaper*). On average, 4.35% in bit rate reduction is achieved. In general, the proposed system results in higher bit rate reduction for low rate points due to the impact of the syntactical data.

TABLE III
COMPLEXITY REDUCTION OF THE PROPOSED MODEL (AVERAGE=71.08%) FOR LEFT (L), CENTER (C), RIGHT VIEW (R) AND TOTAL SEQUENCE.

Test Sequence	L	C	R	overall
Poznan_Hall2	72.62	70.80	70.81	71.16
Poznan_Street	76.10	70.35	72.16	72.18
GT_Fly	71.61	72.43	71.66	72.03
Undo_Dancer	66.15	68.29	70.55	68.46
Kendo	69.16	68.49	67.90	68.45
Balloons	74.91	73.00	72.45	73.20
Loverbird1	78.87	70.29	69.29	71.59
Newspaper	77.38	73.64	71.22	73.41

V. CONCLUSIONS

To lower the cost for depth maps in 3DTV systems, the encoding complexity and bandwidth of depth maps should be reduced. Based on an analysis of encoded depth maps we propose a model to reduce the number of modes for depth map encoding depending on the selected mode in the corresponding texture view. Because of the nature of the proposed scheme, we are also able to reduce the bandwidth of the resulting bit stream. An implementation of our model shows that an average complexity reduction of 71.08% is achieved. Meanwhile, the total bit rate for the depth maps shows an average reduction of -4.35%, with a negligible quality loss.

REFERENCES

[1] MPEG, "Doc. MPEG-W12036: Call for Proposals on 3D Video Coding Technology," Tech. Rep., MPEG, Mar. 2011.

[2] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, "Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264 and ISO/IEC 14496-10 Advanced Video Coding, Edition 5.0 (incl. MVC extension)," Tech. Rep., MPEG / ITU-T, March 2010.

[3] K. Yamamoto, M. Kitahara, H. Kimata, T. Yendo, T. Fujii, M. Tanimoto, S. Shimizu, K. Kamikura, and Y. Yashima, "Multiview video coding using view interpolation and color corrections," *IEEE Tran. Circuits and Systemd for Video Technology*, vol. 17, no. 11, pp. 1436–1449, Nov. 2007.

[4] E. Martinian, A. Behrens, J. Xin, and A. Vetro, "View Synthesis for Multiview Video Compression," in *Picture Coding Symp. (PCS)*, Apr. 2006.

[5] G. Cernigliaro, M. Naccari, F Jaureguizar, J. Cabrera, E Pereira, and N. Garcia, "A new fast motion estimation and mode decision algorithm for H.264 depth maps encoding in free viewpoint TV," in *IEEE International Conference on Image Processing (ICIP) 2011*, sept. 2011.

[6] Sang-Tae Na, Kwan-Jung Oh, and Yo-Sung Ho, "Joint coding of multi-view video and corresponding depth map," in *IEEE International Conference on Image Processing (ICIP) 2008.*, oct. 2008, pp. 2468 –2471.

[7] Ismael Daribo, Christophe Tillier, and Beatrice Pesquet-Popescu, "Motion vector sharing and bitrate allocation for 3d video-plus-depth coding," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, pp. 1–13, 2009.