Depth Map Compression with Diffusion Modes in 3D-HEVC

Yun Li, Mårten Sjöström, Ulf Jennehag and Roger Olsson
Dept. of Information Technology and Media
Mid Sweden University
Sundsvall, Sweden
yun.li@miun.se, marten.sjostrom@miun.se, ulf.jennehag@miun.se, and roger.olsson@miun.se

***Abstract*—For three-dimensional television, multiple views can be generated by using the Multi-view Video plus Depth (MVD) format. The depth maps of this format can be compressed efficiently by the 3D extension of High Efficiency Video Coding (3D-HEVC), which has explored the correlations between its two components, texture and associated depth map. In this paper, we introduce two modes for depth map coding into HEVC, where the modes use diffusion. The framework for inter-component prediction of Depth Modeling Modes (DMM) is utilized for the proposed modes. They detect edges from textures and then diffuse an entire block from known adjacent blocks by using Laplace equation constrained by the detected edges. The experimental results show that depth maps can be compressed more efficiently with the proposed diffusion modes, where the bit rate saving can reach 1.25 percentage of the total depth bit rate with a constant quality of synthesized views.**

Keywords-Depth map coding; HEVC; diffusion modes

I. INTRODUCTION

The Multi-view Video plus Depth (MVD) video format consists of two components: texture and depth map, where a combination of these components enables a receiver to generate arbitrary virtual views. The 3D extension of High Efficiency Video Coding (3D-HEVC) [1] utilizes different prediction techniques to improve the compression efficiency for MVD data. We have previously devised an edge-based compression scheme by diffusion for depth images [2], as the depth image can be assumed to be piece-wise smooth bounded by sharp edges. The question is if better compression of depth maps can also be achieved by implementing diffusion modes block-wise in 3D-HEVC.

Three dimensional video representation using depth map reduces the number of views being transmitted, but coding of depth maps with the current techniques H.264/AVC [3] or its multi-view coding (MVC) extension [4] will introduce visible distortions in synthesized views. Therefore, they are not recommended for depth coding [5]. Various schemes have been developed to address problems of depth map coding. In paper [6], a comparative study showed that Block Truncation Coding (BTC) outperforms the Discrete Cosine Transform (DCT) and the Karhunen-lòeve Transform KLT, and the adaptive BTC was devised that adaptively selects the block size for the BTC. Weighted mode filtering with depth dynamic range reduction [7] and Edge-weighted Optimization Concept (EWOC) with adaptive filtering [8] have been proposed for depth compression. Model based intra coding approach using a depth lookup table and encoding the residuals in pixel domain in 3D-HEVC was devised in paper [9]. The edge-based depth image coding schemes [2][10] utilize diffusion to interpolate the smooth areas bounded by depth edges. There are still many other algorithms for depth map coding, but in this research work we focus on improving our previously proposed edge-based diffusion scheme [2].

The edge-based depth image compression scheme can preserve the transitions on the depth better than traditional video and image encoders [2]. However, such a scheme implies a very expensive encoding of edge contour information in terms of bit rate.

To solve this issue, edges can be extracted from the co-located texture image. Inter-component prediction for depth map coding has recently been implemented in 3D-HEVC [11]. It may employ an inter-component predicted wedgelet partitioning or a predicted contour partitioning for intra coding, the former separates a depth block into two parts by a straight line, the latter divides the block into parts of arbitrary shapes. Intensities of each part are then represented by constant values. The wedgelet partition for a depth block is found by searching for the best wedgelet pattern in the co-located texture luminance block. The contour partition is also detected from the texture luminance. This partition is selected depending on the pixel values in relation to their mean within the texture block, whereby the partition may be of arbitrary shape. The two depth values given to the different parts of the depth block are predicted from the partially reconstructed depth. Fig. 1 shows a depth block, where $P1$ and $P2$ are decoded values in the adjacent blocks. The edges in the current block are detected from the co-located texture luminance by thresholding, and the parts A and B are predicted by the mean of $P1$ and $P2$, respectively.

Another issue with our previously proposed edge-based scheme is the lack of rate distortion control for optimizing the compression ratio. Therefore, one of the solutions for this is to implement the diffusion process in a block-wise manner in HEVC.

A block-based diffusion method based on Laplace equation for H.264 was proposed in [12]. It detects an edge map from the depth map and encodes these edges by the

 ISBN: 978-1-61208-265-3

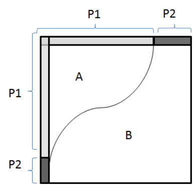

Figure 1. A depth block for inter-component prediction: the edge between part A and B are detected from the co-located texture block, and parts of the adjacent reconstructed blocks available for prediction are showed in dark and light grey.

bi-level image compression tool JBIG. The method uses these edges as constraints with the Laplace diffusion when reconstructing the depth map on blocks of a fixed size.

In this work, we propose two new modes based on block-wise Laplace diffusion. We replace two inter-component prediction modes in the 3D-HEVC by the proposed modes in order to save bits used for signaling. The novelties of this work are: (1) block-based diffusion modes are introduced into HEVC using inter-component prediction framework; (2) the block size is allowed to be further split in the same way as the original inter-component prediction modes; (3) the diffusion is conducted in two-step if isolated parts still exist in the block.

The overall aim of the work is to improve compression ratio for depth maps with a sustained 3D video quality. The work is limited to reusing the inter-component prediction framework, and the goals is to investigate the rate-distortion ratio for the new proposed diffusion modes, where quality is measured on synthesized views.

The sequel of the paper is organized as follows. We illustrate the proposed modes in Section 2, and test arrangements and evaluation criteria in Section 3. Section 4 presents the results and analysis, and Section 5 concludes the work.

II. PROPOSED METHOD

Fig. 2 illustrates all eight Depth Modeling Modes (DMM) in the 3D-HEVC software [13]. They are derived from the 3D-HEVC test model [1]. Among them, the modes (1), (2), (7) and (8) employ wedgelet partitions detected by a search on the depth block or predicted from the previously coded blocks, i.e., they are non-inter-component prediction modes. The inter-component prediction modes (3), (4), (5) and (6) derive, on the other hand, the partition information from the co-located texture block. Mode (2), (4), (6) and (8) employ so called delta constant partition value coding, i.e., they encode the difference between the mean of the original signal and the predicted constant value, which is the mean of the available adjacent prediction signal.

The original mode (5) in Fig. 2 is denoted as DMM-TEX-CONTOUR in the context of this paper. In this mode, the partition is detected from the co-located texture luminance by thresholding. As mentioned, the partition can also be detected by searching for the best Wedgelet pattern in

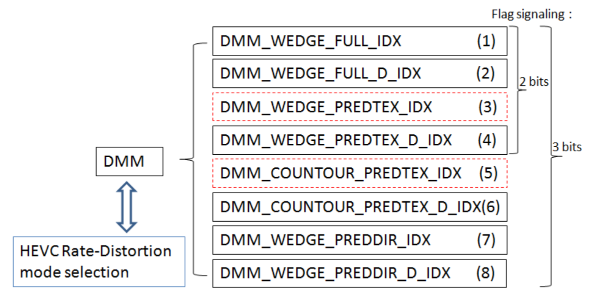

Figure 2. Depth modeling modes. (1) and (2) Explicit wedgelet signaling. (3) and (4) Inter-component predicted wedgelet partitioning. (5) and (6) Inter-component predicted contour partitioning. (7) and (8) Intra-predicted wedgelet partitioning.

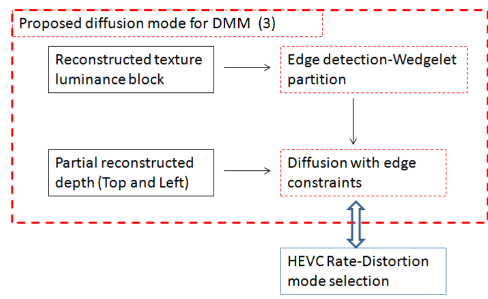

Figure 3. Proposed diffusion mode for DMM-TEX-WEDGE.

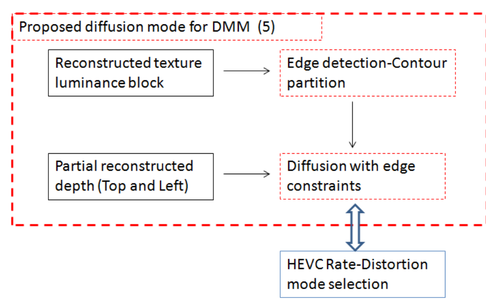

Figure 4. Proposed diffusion mode for DMM-TEX-CONTOUR.

the co-located texture block, i.e., in mode (3). This inter-component predicted wedgelet partition may avoid a possible inconsistency of the contour detection from the texture [11]. We also denote the original mode (3) in Fig. 2 as DMM-TEX-WEDGE.

The DMM-TEX-WEDGE and DMM-TEX-CONTOUR were replaced by the proposed diffusion modes for intra prediction. We replaced the existing modes instead of adding new modes in the inter-component prediction framework because no additional bits had been required for signaling the proposed modes. The original partitioning methods were kept and the obtained edges are used as constraints in the Laplace diffusion.

The proposed modes, illustrated in Fig. 3 and 4, thus include two processes: Edge detection and Diffusion with edge constraints, which are defined as follows:

Edge detection: The modes kept the original partitioning

 ISBN: 978-1-61208-265-3

methods using texture luminance for the edge detection. They require no extra bits for encoding the depth edges, whereas explicit coding of depth edges require substantial amount of coding bits. The methods for obtaining the edges are different for the contour and the wedgelet partitioning.

Edge detection-Wedgelet partition: The wedgetlet partitioning is carried out by an efficient wedgelet search on the co-located texture block for the least distortion. Edge is the straight line that separates two parts.

Edge detection-Contour partition: The contour partition for the depth block is made by a thresholding process, in which parts from the partitioning are obtained based on if the value in the co-located texture block is above or below the mean value of this texture block. Edges are located at the transitions between the parts.

Diffusion with edge constrains: The new diffusion modes for intra prediction are also showed in Fig. 1. The parts A and B are diffused from $P1$ and $P2$ respectively.

Laplace equation

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0, \quad (1)$$

is employed for the diffusion. The unknowns of the equations are solved by a method described in [12]. The method refines the diffusion iteratively as:

$$f^{(n+1)}(x,y) = \frac{1}{N(x,y))} \sum_{(i,j)\varepsilon u_4(x,y)} C(i,j))f^n(i,j), \quad n = 1,2,3...M) \quad (2)$$

$$C(x,y) = \begin{cases} 1, & if\ f(x,y)\ \ is\ \ available) \\ 0, & else, \end{cases} \quad (3)$$

$$N(x,y) = \sum_{(i,j)\varepsilon u_4(x,y)} C(i,j)). \quad (4)$$

The equations describe that the diffusion for a depth map block $f^{(n)}$ is refined iteratively with the number of iteration (n). $u_4(x,y)$ represents the four neighbors (up, right, down and left) around the current refined pixel with position (x,y) in the block. $C(i,j)$ denotes the availability of these neighbors (e.g., the pixels taken into calculation are available and belong to the same part), and $N(x,y)$ sums up the number of the available neighbors. The iteration stops with a convergence condition in Equation 5a. In addition to this condition, we also impose a time constrain for the diffusion, which is to limit the number of iterations M. Therefore, the diffusion process stops when either of the conditions a or b is satisfied:

$$\begin{cases} a. & |f^{(n+1)} - f^{(n)}| < 0.05 \\ b. & n >= M. \end{cases} \quad (5)$$

A. Two-step diffusion

The contour partition may appear much more complex than the one showed in Fig. 1. The parts can be arbitrary shapes and even be isolated within a block. An example is depicted in Fig. 5. Our approach to fill these isolated parts is by using a two-step diffusion. In the first step, parts that are connected with the available prediction pixels are diffused, which is illustrated in Fig. 5(d). In the second step, the diffusion is carried out without the edge constraint for only those isolated parts. Fig. 5(e) shows the final diffused block. As to the maximum iterations for the diffusion in Equation 5, we set $M = 20$ for the diffusion step 1 and $M = 10$ for the step 2.

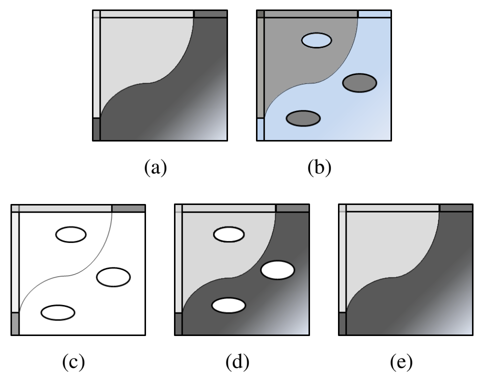

Figure 5. Two-step diffusion: (a) Original depth block, (b) co-located texture block, (c) detected edges on the texture, (d) diffusion after the first step, and (e) diffusion after the second step.

Such a diffusion process might produce erroneous depth values for the isolated parts, but these edges detected from the co-located textures might also not exist in the depth block. The HEVC rate-distortion process decides if the proposed modes are chosen.

III. Test arrangements and evaluation criteria

The test arrangements and evaluation criteria are described as follows:

A. Implementation and Test setup

The proposed modes have been implemented in 3DV HEVC Test Model (3DV-HTM) software version 4.1 [13]. The evaluation partially followed the Call for Proposals on 3D Video Coding Technology [14]. However, we evaluated only the intra-frame coding to better understand the effectiveness of the proposed intra diffusion modes. Therefore, the bit rate anchors were not followed. We chose two-view configurations and four test sequences with fixed Quantization Parameter (QP) pairs for texture and depth. The MPEG test sequences [14]: Poznan Street [15], Poznan Hall [15], Undo Dancer, and Newspaper were selected. The first 50 frames from these sequences were evaluated.

We used Poznan Street view 3, Poznan Hall view 6, Undo Dancer view 2 and Newspaper view 4 for the evaluation of depth. Virtual views were rendered at camera position 3.5 for Poznan Street, camera position 6.5 for Poznan Hall, position 3 for Undo dancer and position 5 for Newspaper for the assessment of synthesized views. The virtual views were synthesized from the decoded texture and the decoded depth, and compared to the virtual views synthesized from the original texture and the original depth. VSRS [16] version 3.5 was employed for the view synthesis.

The View Synthesized Optimization (VSO) [17] was turned off, i.e., in the HEVC rate-distortion optimization, the distortion is measured on the depth map instead of on the synthesized view when encoding of depth map. The QPs in (texture, depth) format were (20, 30), (25, 34), (30, 38) and (35, 42). These QPs were selected because the bit rate for the depth should be significantly lower than for the texture for an optimized bit rate allocation between texture and depth [17]. The results using the alternated 3D-HEVC with the proposed modes were compared to results using the original 3D-HEVC in the same testing conditions.

B. Evaluation criteria

The results were calculated using the BD-PSNR model [18]. In this model, a curve is fitted through the PSNR values of four bit-rate points. The difference between the integrals divided by their respective integration intervals is the average difference for two curves. In the evaluation, the bit rate change for depth was computed over the bit rates for the depth map versus PSNR of the decoded depth map, whereas the bit rate change for the synthesized views was calculated over the bit rates for the depth map versus PSNR of the synthesized view.

The complexity of the modes is presented as a ratio of total coding time between the proposed scheme and the 3D-HEVC.

IV. Results and Analysis

The results are illustrated in Table I. The bit rate saving is around 0.64 percent for Poznan Hall, 0.47 for Poznan Street and 0.28 for Newspaper when only the depth quality is considered. When the evaluation of PSNR is on the synthesized views, around 0.49 percent bit rate savings were achieved for Poznan Hall and 0.31 percent for newspaper. Better bit rate savings were obtained for the synthetic sequence Undo Dancer, where 1.54 percent for the depth and 1.25 percent for the synthesized views were achieved. The results further show that there is no improvement for the Poznan Street sequence when considering the synthesized views.

Table II summarizes the complexity of the proposed modes. The complexity increases in average by 6.2 percent for encoding and 3.4 percent for decoding. An exception is for Undo Dancer sequence, where the decoding time is 4.2 percent less than for the 3D-HEVC. This implies that, in some cases, the proposed diffusion modes are more efficient

TABLE I. BD-PSNR For The Tested Sequences (The Bit Rate Change In Percentage Of The Total Depth Bit Rate)

Sequence	BD-rate(depth) (%)	BD-rate(virtual view) (%)
Undo Dancer	-1.536	-1.247
Newspaper	-0.282	-0.312
Poznan Street	-0.465	0.026
Poznan Hall	-0.642	-0.488
Average	-0.731	-0.505

TABLE II. Coding Complexity (Time Ratio Between Proposed And Reference Schemes)

Sequence	Encoding	Decoding
Undo Dancer	1.041	0.958
Newspaper	1.055	1.096
Poznan Street	1.049	1.055
Poznan Hall	1.102	1.026
Average	1.062	1.034

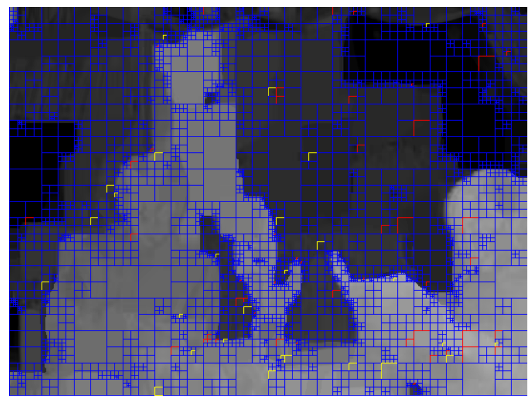

Figure 6. A depth image from the first frame of Newspaper: the blocks marked with red and yellow use the proposed modes that replaced the DMM-TEX-WEDGE and DMM-TEX-CONTOUR respectively.

in decoding than some of the other intra modes in the 3D-HEVC.

An example of block fragmentations and modes assignments are plotted in Fig. 6. Our proposed modes are marked with red and yellow color, which represent the two proposed modes that replaced DMM-TEX-WEDGE and DMM-TEX-CONTOUR, respectively. The total area covered by the proposed modes is 2.33 percent of the entire image among all intra prediction modes in Fig. 6.

The test results illustrate that better compression of depth maps can be achieved with the proposed modes in 3D-HEVC, and that the decoding complexity increases by less than 4 percent. The proposed modes target only inter-component prediction framework, and they cover a very small percentage of the entire depth map. Thus the effectiveness seems less significant. By replacing further intra-modes by diffusion modes, it is likely that further depth compression may be achieved.

The experimental results also demonstrate that the improvement for the quality of decoded depth is consistent. This implies that the Laplace diffusion process can better approximate the original depth signals than the constant

 ISBN: 978-1-61208-265-3

partition value coding in 3D-HEVC under the given testing conditions. The fast advancement in hardware processing power will likely make high computational complexity less of a problem in the future.

This work aimed at improving depth compression (reducing bandwidth consumption) for a better quality of synthesized views in 3D-HEVC, which is the state of the art in coding of 3D video contents. With the proposed diffusion modes, the proposed scheme outperforms the original 3D-HEVC. As coding of 3D video contents has been attracting many research attentions, we also aim at comparing our scheme with other novel methods and improving the proposed scheme further in the future research.

V. CONCLUSION

We have implemented two modes using diffusion in 3D-HEVC for coding of depth map and replaced two inter-component prediction modes by the proposed modes. They utilize edges from the associated texture and diffuse depth values in the block by using Laplace equation with texture edge constraints.

The experimental results illustrate that the proposed modes can improve the compression efficiency for depth map coding, and that the complexity increases by 3.4 percent in average for the decoding. When considering the quality of synthesized views, the bit rate saving can reach around 1.25 percentage of the total depth bit rate for the tested MVD sequences. The bit rate saving is efficient, considering that the proposed modes have been implemented in the inter-component prediction framework only and cover a very small percentage of the depth image among all intra prediction modes.

Future works consist of better edge detection schemes to reduce the partitioning errors for diffusion, investigating the possibility of introducing diffusion into further intra modes, optimizing the proposed modes with View Synthesized Optimization (VSO) enable and subjective quality oriented encoding by using the diffusion modes for a better view synthesis.

ACKNOWLEDGMENT

This work has been supported by grant 2009/0264 of the Knowledge Foundation, Sweden, by grant 00156702 of the EU European Regional Development Fund, Mellersta Norrland, Sweden, and by grant 00155148 of Länsstyrelsen Västernorrland, Sweden.

REFERENCES

[1] "3D-HEVC Test Model Description Draft 1," ITU-T SG 16 WP 3 JCT3V-A1005_d0, July 2012.

[2] Y. Li, M. Sjöström, U. Jennehag, and R. Olsson, "A scalable coding approach for high quality depth image compression," in 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2012, Oct. 2012, pp. 1 –4.

[3] "Advanced video coding for generic audiovisual services," ITU-T Recommendation H.264 and ISO/IEC 14496-10, Jan. 2012.

[4] Y. Chen, Y.-K. Wang, K. Ugur, M. M. Hannuksela, J. Lainema, and M. Gabbouj, "The emerging mvc standard for 3d video services," EURASIP J. Appl. Signal Process., vol. 2009, Jan. 2008, pp. 8:1–8:13.

[5] K. Muller, P. Merkle, and T. Wiegand, "3-d video representation using depth maps," Proceedings of the IEEE, vol. 99, no. 4, April 2011, pp. 643 –656.

[6] H. Nayyar and A. Wei, "A Comparative Study of Depth-Map Coding Schemes for 3D Video," Image and Video Compression, Stanford University, Mar. 2011.

[7] V.-A. Nguyen, D. Min, and M. N. Do, "Efficient techniques for depth video compression using weighted mode filtering," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 23, no. 2, Feb. 2013, pp. 189 –202.

[8] S. Schwarz, M. Sjöström, and R. Olsson, "Depth map upscaling through edge-weighted optimization," in Proc. SPIE 8290, Three-Dimensional Image Processing (3DIP) and Applications II, Feb. 2012, pp. 829 008–8.

[9] F. Jager, M. Wien, and P. Kosse, "Model-based intra coding for depth maps in 3d video using a depth lookup table," in 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2012, Oct. 2012, pp. 1 –4.

[10] J. Gautier and O. Meur, "Efficient depth map compression based on lossless edge coding and diffusion," Picture Coding Symposium, 2012, pp. 81–84.

[11] P. Merkle, C. Bartnik, and K. Muller, "3D video: Depth coding based on inter-component prediction of block partitions," in Proc. Picture Coding Symposium, 2012, pp. 149–152.

[12] J. Chen, F. Ye, J. Di, C. Liu, and A. Men, "Depth map compression via edge-based inpainting," Picture Coding Symposium, 2012, pp. 57–60.

[13] 3DV HEVC Test Model (3DV-HTM) version 4.1. Retrieved: 09, 2010. [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-4.1/

[14] "Call for Proposals on 3D Video Coding Technology," ISO/IEC JTC1/SC29/WG11, Mar. 2011.

[15] M. Domaski, T. Grajek, K. Klimaszewski, and M. Kurc, "Pozna Multiview Video Test Sequences and Camera Parameters," ISO/IEC JTC1/SC29/WG11, 2009.

[16] "Report on experimental framework for 3D video coding," ISO/IEC JTC1/SC29/WG11 MPEG2010/N11631, Oct. 2010, Guangzhou, China.

[17] G. Tech, H. Schwarz, K. Muller, and W. Thomas, "3D video coding using the synthesized view distortion change," Picture Coding Symposium, 2012, pp. 25–28.

[18] G. Bjontegard, "Calculation of average PSNR differences between RD-curves," ITU-T VCEG-M33, Mar. 2001.