DEPTH MAP CODING BASED ON VIRTUAL VIEW QUALITY

Chao Yang, *Ping An*, *Deyang Liu*, *Liquan Shen*

Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai, China
School of Communication and Information Engineering, Shanghai University, Shanghai, China

ABSTRACT

Multi-view video plus depth (MVD) is a 3D video representation. In MVD, the depth map provides the scene distance information and is used to render the virtual view through Depth Image Based Rendering (DIBR) technique. The depth map coding error will induce distortion in the rendered virtual views. This paper proposes a mathematic model that can estimate the synthesized virtual view distortion induced by depth map compression, and the model is employed to the rate distortion optimization (RDO) in the depth map coding. Based on the rendered virtual view quality, a Lagrangian optimization adjustment scheme at Coding Unit (CU) level is proposed to improve the depth map encoding efficiency. Experimental results demonstrate that the proposed method can improve the BD-PSNR of virtual view for 0.62 dB, and the encoding complexity reduces compared with the view synthesis optimization (VSO) technique in the 3D-HEVC Test Model (HTM).

Index Terms— depth map distortion, depth map encoding, exponential model, Lagrangian optimization, virtual view distortion

1. INTRODUCTION

Multi-view video plus depth (MVD) 3D video representation enables functionalities like 3D television and free viewpoint video [1]. MVD includes colorful texture image and grey depth map, the depth map is not displayed to the viewers but it is used to render the arbitrary virtual view with Depth Image Based Rendering (DIBR) [2] technique, which will provide the viewers with more realistic 3D visual effect. From the coding efficiency point of view, the bitrate of the depth map accounts for about 40~60% of the texture bitrate, which makes MVD possible to reduce the total bitrate of 3D video [3].

During the encoding process, quantization will induce distortions in both texture image and depth map, which will in turn induce distortions in the rendered virtual view images. The depth map provides the scene geometry information, and distortion in depth map will cause pixel location deviation in the rendered virtual view images and induce virtual view distortion.

Many researches exploit the relationship between the depth map distortion and the virtual view distortion. Shao *et al.* proposed a liner model among texture image distortion, depth map distortion and virtual view distortion [4], but the model was not used to improve depth map coding efficiency. [5] and [6] proposed a linear model and employed it in depth map coding, but the model cannot estimate virtual view distortion accurately. De Silva *et al.* proposed an algorithm to minimize the virtual view distortion in intra mode [7-8], but the algorithm introduces high computation complexity. In 3D-HEVC Test Model (HTM) reference software, view synthesis optimization (VSO) technique is used to evaluate the virtual view distortion in order to improve the virtual view quality [9], in order to reduce the complexity of virtual view rendering, the rendering process is simplified which makes the estimation of virtual view distortion inaccurate.

In order to estimate the virtual view distortion accurately to improve depth map coding efficiency, an exponential model with low complexity is proposed in this paper. We employ the exponential model and the corresponding Lagrangian Multiplier (LM) into the rate distortion optimization (RDO) in HTM reference software version 7.0 to improve the depth map coding efficiency. The proposed model does not perform the actual view synthesis for each depth map frame so its complexity is lower compared with VSO.

Jiang *et al.* shows that dynamically adjusted LM at Macro Block (MB) level can achieve better performance in video coding in H.264/AVC [10]. Depth map is predominantly flat and may have noise and temporal inconsistency raised by depth estimation [11]. Different depth map areas have different effect on the synthesized virtual view, based on the characteristics of depth map, adaptive Quantization Parameter (QP) adjustment is employed in this paper to achieve better depth map coding performance.

The remainder of this paper is organized as follows. Section II derives the exponential model to estimate the virtual view distortion induced by depth map compression. Section III introduces the utility of the proposed exponential model in the RDO in HTM reference software. Section IV introduces the Coding Unit (CU) level Lagrangian optimization adjustment which can improve depth map

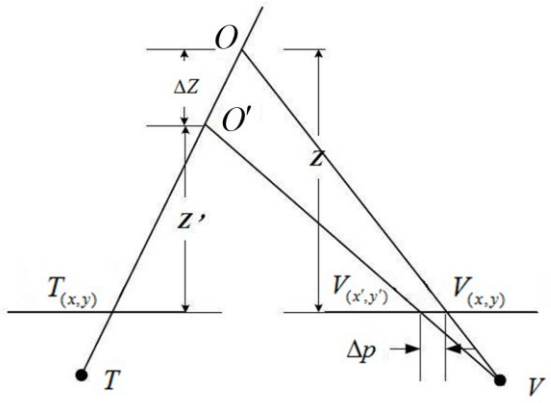

Fig.1. DIBR algorithm principle (parallel camera setup)

TABLE I
TEST SEQUENCES & RENDERED VIRTUAL VIEWS

Sequences	Reference Views	Virtual View
Balloons	1 & 3	2
Kendo	1 & 3	2
Lovebird1	4 & 6	5
Newspaper	2 & 4	3
PoznanHall2	6 & 7	6.5
PoznanStreet	4 & 5	4.5

coding efficiency. The experimental results and conclusions are given in section V and section VI respectively.

2. VIRTUAL VIEW DISTORTION METRIC

In MVD, the depth map is used to render the virtual view through a DIBR technique, which is illustrated in Fig. 1. The depth map coding error ΔZ induces pixel location deviation Δp in the rendered virtual view, which induces virtual view distortion.

The virtual view distortion can be approximately divided into texture image induced distortion and depth map induced distortion respectively [12]. As the ground truth of intermediate virtual view does not exist, in order to evaluate the virtual view distortion induced by depth map compression, the virtual view synthesized using uncompressed depth map is used as the ground truth virtual view image. Mean Squared Error (MSE) is used to evaluate the distortion, as written in Eq. (1),

$$MSE = E[(\hat{I}_i - I_i)^2] \quad (1)$$

where I_i and $\hat{I}_i$ denote the pixel in the ground truth and distorted virtual view image, respectively and $E[\ ]$ denotes the expectation taken over all pixels in one virtual view image.

The test sequences and their synthesized virtual views are demonstrated in TABLE I, the depth map is coded with different QPs ranging from 10 to 50. The relationship between depth map distortion and virtual view distortion of

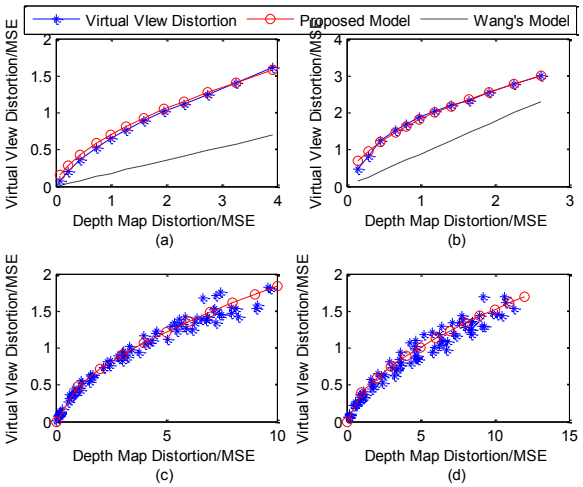

Fig.2.Relationship between depth map distortion and virtual view distortion: (a) sequence '*Kendo*'; (b) sequence '*Lovebird1*'; (c) the first GOP in '*Kendo*'; (d) the third GOP in '*Kendo*'

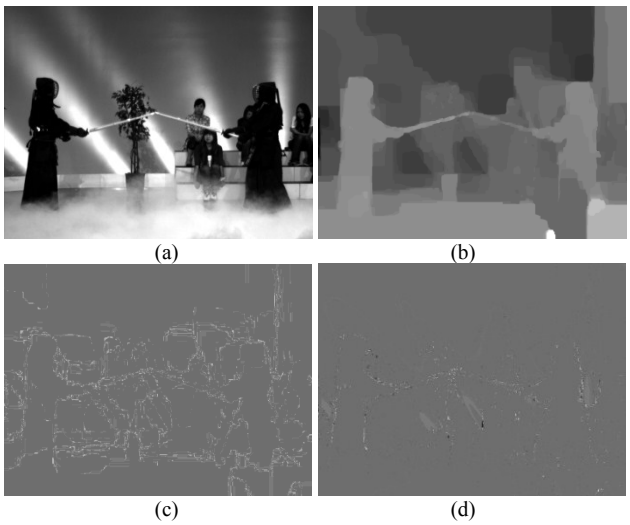

Fig.3. Test Sequence Kendo: (a) Texture video; (b) Depth map; (c) Difference between original and coded depth map; (d) Difference in synthesized virtual view after depth map coding

test sequence '*Kendo*' and '*Lovebird1*' are demonstrated in Fig. 2 (a) and (b). Fig. 2 (c) and (d) illustrate the relationship of different frames in '*Kendo*'. In the figure, the horizontal axis denotes MSE between the original and the compressed depth map, the vertical denotes MSE of the synthesized virtual view.

Based on Fig. 2, an exponential model in Eq. (2) is proposed to estimate the virtual view distortion induced by depth map compression,

$$D_v = \alpha \cdot D_d^{\ \beta} \quad (2)$$

where D_d and D_v represent the depth map distortion and the virtual view distortion, respectively, α and β are model parameters based on test sequences. Compared with Wang's linear model in [6], the exponential model in Eq. (2) can estimate the real virtual view distortion more accurate.

As we can see in Fig. 2, for different test sequences and different frames in one sequence, the curves are different, which means the parameters of the exponential model are different. In order to calculate the model parameters α and β for depth map coding, we need to pre-encode one depth map frame twice with different QPs, synthesize the virtual views with the coded depth maps and then calculate the model parameters.

In order to reduce the complexities induced by depth map pre-encoding and virtual view synthesis, while maintain the accuracy of the model parameters, we propose to pre-encode the first depth map frame in one Group of Picture (GOP) with different QPs to calculate the model parameters, and then use the fixed parameters to encode the remaining depth map frames in the current GOP. In 3D-HEVC hierarchical structure, one GOP consists of 8 frames. Since the depth map frames in one GOP are correlated, encoding one GOP with fixed parameters can guarantee the depth map coding efficiency. On the other hand, we update the model parameters at GOP level, when depth map content changes, the model can be estimated correctly for efficient depth map coding.

Coding distortion usually emerges in complex areas [13], for depth map, sharp edges like boundaries between objects in different depths are complex areas [14], as shown in Fig. 3 (b). Depth map distortion leads to geometric error in the synthesized view and affects the synthesized virtual view quality. For this reason, depth map mainly distorts in the complex areas after compression, as shown in Fig. 3 (c). The virtual view images synthesized through DIBR also distort in these areas, which is demonstrated in Fig. 3 (d).

In texture image, pixel intensities in the same depth are similar, while differ from those in other depths, as we can see in Fig. 3 (a). Depth map provides disparity information, so depth map distortion induces pixel location deviation in virtual view images. When the pixels deviate from one depth to another, the virtual view distortion is significant. On the contrary, if pixels stay in the same depth after deviation, the distortion in virtual view is negligible.

Because of this, with the augment of depth map distortion, virtual view image distortion does not increase linearly. The exponential relationship in Eq. (2) can estimate virtual view distortion more accurately, as shown in Fig. 2.

3. DEPTH MAP RATE DISTORTION OPTIMIZATION

In HTM reference software, for mode selection and motion estimation, Lagrangian Optimization (LO) is used to determine each candidate mode and parameter [15]. By adding efficient coding options in the rate-distortion sense to the codec, the overall coding performance will increase. The optimization is to choose the most efficient coded representation in the rate-distortion sense for each block. The optimization task is complicated because various coding options contain varying distortion at different bitrates. The Lagrangian optimization used in RDO could be written as:

$$\arg\min_{S} J = \arg\min_{S}(D_d + \lambda_d \cdot R_d) \tag{3}$$

where S is candidate of coding modes, D_d and R_d represent the depth map coding distortion and coding bits respectively, and λ_d is the LM, which is not chosen arbitrarily. In LO, λ_d is determined by Eq. (4) [16].

$$\lambda_d = -\frac{\partial D_d}{\partial R_d} \tag{4}$$

Given that the depth map is not used to display but render virtual views, D_d in Eq. (3) should be modified in order to improve the rendered virtual view quality while encode depth map. For this reason we use the depth map virtual view distortion model in Eq. (2) to replace the distortion metric D_d. As the distortion metric is modified, the LM is also supposed to be modified according to Eq. (4).

As the LM has a relationship with distortion and bits in Eq. (4), from Eq. (2) we can get the modified λ_v for the depth map coding written in Eq. (5),

$$\lambda_v = -\frac{\partial D_v}{\partial R_d} = -\frac{\partial D_v}{\partial D_d}\frac{\partial D_d}{\partial R_d} = \alpha \cdot \beta \cdot {D_d}^{\beta-1} \cdot \lambda_d \tag{5}$$

where λ_d is the original LM for the depth map RDO in HTM.

With virtual view distortion metric D_v and modified LM λ_v, the new Lagrangian Optimization formulation for RDO is complete, written as Eq. (6).

$$\arg\min_{S} J = \arg\min_{S}(D_v + \lambda_v \cdot R_d) \tag{6}$$

With Eq. (6), we are able to consider the virtual view distortion while encode depth map, which can improve the rendered virtual view quality.

4. LAGRANGIAN OPTIMIZATION ADJUSTMENT

In HEVC depth map coding, Lagrangian Multiplier λ_d is calculated by Eq. (7),

$$\lambda_d = QPfactor \cdot 2^{(QP-12)/3} \tag{7}$$

where $QPfactor$ is a constant parameter.

Wang adjusted LM dynamically in H.264/AVC to achieve a better coding performance [17]. The LM in 3D-HEVC depth map coding is similar to that in H.264/AVC, adjusting LM dynamically at CU level can also achieve better performance in 3D-HEVC depth map coding. Furthermore, in DIBR, distortions in complex areas like boundaries between different depths induce significant distortions in synthesized virtual view images, while homogenous areas like backgrounds have negligible distortions. Since depth map has this special characteristic, in order to improve the coding efficiency, we adjust LM and *QP* dynamically based on the complexity of depth map. We employ the gradient

TABLE II
BDPSNR PERFORMANCE OF DIFFERENT METHODS

Sequences	Proposed	Wang's Method
Balloons	0.36	-0.84
Kendo	0.12	-1.07
Lovebird1	0.29	-0.94
Newspaper	2.44	1.47
Poznan_Hall2	0.13	-1.32
Poznan_Street	0.36	-1.43
Average	0.62	-0.69

TABLE III
ENCODE TIME SAVING OF DIFFERENT METHODS COMPARED WITH VSO

Test Sequences	Encode Time Reduction (%)	
	Proposed Method	Wang's Method
Balloons	-17.8	-21.1
Kendo	-17.2	-22.2
Lovebird1	-18.8	-21.8
Newspaper	-20.4	-24.6
Poznan_Hall2	-23.5	-25.9
Poznan_Street	-19.0	-21.8
Average	-19.5	-22.9

magnitude G to represent the complexity of CU in the depth map, and G is calculated as follow:

$$G = \frac{1}{M \cdot N}\{\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}(I_{i,j} - I_{i+1,j})^2 + (I_{i,j} - I_{i,j+1})^2\} \quad (8)$$

where $I_{i,j}$ indicates the depth map pixel intensity located at (i,j), M and N denote the horizontal and vertical dimensions of the CU respectively. Complex areas derive large G values.

Since the complexities of CUs in one depth map frame are various, the fixed LM and QP can hardly satisfy the optimal coding efficiency. Ideally, LM and QP should adjust in different depth map CUs. Hence, we employ the adjustment factor k_i to further revise LM:

$$\lambda_i = k_i \cdot \lambda \quad (9)$$

where λ_i is the LM for the ith CU in the current depth map frame, and k_i is the adjustment factor:

$$k_i = \begin{cases} 1.25 & G_{CU} > G_{Pic} \\ 1.0 & Others \end{cases} \quad (10)$$

where G_{CU} is the gradient magnitude of the current CU and G_{Pic} is the gradient magnitude of the current frame.

As LM adjusts along with the complexity of CU, QP should also adjust along with the complexity. Let QP_i represent the QP of the current CU and QP represent the QP of the current frame, we revise QP_i as follow:

$$QP_i = \begin{cases} QP-2 & G_{CU} > G_{Pic} \\ QP & Others \end{cases} \quad (11)$$

Both LM and QP adjustment values are obtained through extensive experiments.

5. EXPERIMENTAL RESULTS

We employ the proposed distortion estimation model and RDO adjustment scheme into HTM reference software version 7.0, the detailed test settings are provided in TABLE I and the test configuration is based on the Common Test Conditions (CTC) in 3D-HEVC [18].

We compare the proposed method and Wang's method [6] with the VSO in HTM [9], i.e. the VSO is set as the benchmark and we calculate the BDPSNR [19] of the two methods with the benchmark. The BDPSNR of the two methods are presented in TABLE II. As we can see from the table, the BDPSNR with the proposed method is 0.62 dB higher compared with VSO and is much better than that of Wang's method, which has proven the effectiveness of the proposed method.

In order to evaluate the complexity of the proposed method, Eq. (12) is employed to calculate the coding time reduction:

$$\Delta T = \frac{T_{method} - T_{VSO}}{T_{VSO}} \times 100\% \quad (12)$$

where T_{method} is the encoding time of the proposed method or Wang's method, and T_{VSO} is the encoding time of VSO.

TABLE III demonstrates that the coding time of the proposed method reduces approximately 20% compared with VSO. As the proposed metric needs not to render every block in depth map coding, its computational complexity is reduced compared with VSO. For the proposed method needs pre-encoding to calculate the model parameters, this has induced extra complexity which makes its complexity a bit higher than Wang's method, but compared to VSO, the complexity of pre-encoding is negligible.

6. CONCLUSIONS

In this paper, an exponential model is proposed, which can estimate the virtual view distortion induced by depth map distortion accurately. This exponential model is employed to the RDO mode selection scheme along with the new LM derived from the exponential model. A Lagrangian optimization adjustment scheme at CU level based on depth map complexity is also proposed to increase the depth map coding efficiency. Experimental results with different test sequences demonstrate that the proposed method can bring considerable BDPSNR improvement compared with VSO. And the proposed method has lower coding complexity which brings about 20% time saving compared with VSO.

7. ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China, under Grants 61172096, U1301257, 61571285 and 61422111.

8. REFERENCES

[1] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Multi-view video plus depth representation and coding," *in Proc. IEEE ICIP*, vol. 1, pp. 201-204, Sept. 2007.

[2] X. Xu, L.-M. Po, K.-W. Cheung, K.-Ho Ng, K.-M. Wong, and C.-W. Ting, "A foreground biased depth map refinement method for DIBR view synthesis," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 805-808, Mar. 2012.

[3] E. Bosc, V. Jantet, M. Pressigout, L. Morin, and C. Guillemot, "Bit-rate allocation for multi-view video plus depth," *in Proc. 3DTV Conf: True Vision-Capture, Transmission Display 3D Video (3DTVCON)*, pp. 1-4, 2011.

[4] F. Shao, G. Jiang, M. Yu, K. Chen, and Y.-S. Ho, "Asymmetric coding of multi-view video plus depth based 3-D video for view rendering," *IEEE Trans. on Multimedia*, vol. 14, no. 1, pp. 157-167, Feb. 2012.

[5] B. T. Oh, J. Lee, and D.-S. Park, "Depth Map Coding Based on Synthesized View Distortion Function," *IEEE J. Sel. Topics Signal Process.*, Vol. 5, no. 7, pp. 1344-1352, Nov. 2011.

[6] L. Wang, and L. Yu, "Rate-distortion Optimization for Depth Map Coding with Distortion Estimation of Synthesized View," *IEEE Int. Symposium on Circuits and Systems (ISCAS 2013)*, Beijing, China, pp.17-20, May 2013.

[7] D. V. S. X. De Silva, and W. A. C. Fernando, "Intra mode selection for depth map coding to minimize rendering distortions in 3D video," *IEEE Trans. on Consumer Electron.*, vol. 55, no. 4, pp. 2385-2393, Nov. 2009.

[8] D. De Silva, W. Fernando, and H. Arachchi, "A new mode selection technique for coding depth maps of 3D video," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 686-689, Mar. 2010.

[9] G. Tech, K. Wegner, Y. Chen, and S. Yea, "3D-HEVC Test Model 3," *MPEG number m28377*, Geneve, Switzerland, January 2013.

[10] M. Jiang, and N. Ling, "On Lagrange multiplier and quantizer adjustment for H. 264 frame-layer video rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 5, pp. 663-669, May 2006.

[11] Y. Zhang, S. Kwong, S. Hu, and C.-C. J. Kuo, "Efficient Multiview Depth Coding Optimization Based on Allowable Depth Distortion in View Synthesis," *IEEE Tans. Image Process.*, vol. 23, no. 11, pp. 4879-4892, 2014.

[12] H. Yuan, Y. Chang, M. Li, and F. Yang, "Model based bit allocation between texture images and depth maps," *in Proc. Int. Conf. CCTAE*, vol. 3, pp. 380-383, Aug. 2010.

[13] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards-Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 22, no. 12, pp. 1668-1683, Dec. 2012.

[14] W. S. Kim, A. Ortega, P. L. Lai, D. Tian, and C. Gomila, "Depth map distortion analysis for view rendering and depth coding," *in Proc. IEEE Int. Conf. Image Process.*, pp. 721-724, 2009.

[15] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards-Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 22, no. 12, pp. 1668-1683, Dec. 2012.

[16] G. J. Sullivan, and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Mag.*, vol. 15, no. 6, pp. 74-90, 1998.

[17] M. Wang, and B. Yan, "Lagrangian multiplier based joint three-layer rate control for H. 264/AVC," *IEEE Signal Process. Lett.*, vol.16, no. 8, pp. 679-682, 2009.

[18] D. Rusanovskyy, K. Mueller, and A. Vetro "Common test conditions of 3DV core experiments," Joint Collaborative Team on 3D Video Coding Extensions (JCT-3V) Document JCT3V-E1100, 5th Meeting: Vienna, Austria, 2013

[19] G. Bjontegaard, "Calculation of average PSNR difference between RD-curves," 13th VCEG-M33 Meeting, Austin, TX, Apr. 2001