Quality Improvement for View Synthesis in 3D-HEVC by Post-Processing of Depth Map

Jie-Ru Lin⁺ and Mei-Juan Chen

Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan

Abstract. The depth map in multiview plus depth (MVD) format of 3D-HEVC is able to synthesize series virtual views in different camera positions. However, the block-based encoding structure of the video coding standard results in blocking effect especially for depth map, which not only degrades the prediction accuracy of the encoder, but also affects the video quality of the synthesized view. This paper proposes a post-processing algorithm for depth map to preserve significant edges and smooth jagged contours resulted from blocking effect. We separate the encoded depth map into three regions by Otsu's automatic segmentation. Then, the vision-based edge detection, which combining the JNDD (Just Noticeable Depth Difference) model, is utilized to obtain the visible depth edges in 3D display. Moreover, the geometric edges of the depth map are also acquired by the edge classification in MPEG-7. Finally, the Bilateral filter, JNDD filter and Gaussian filter are applied to the encoded depth map under various cases. The experimental results show that the proposed post-processing method obviously maintains the vital edge information of the depth map when removing the jagged contours and provides better video quality in synthesized view.

Keywords: 3D-HEVC, depth map, post-processing, synthesized view.

1. Introduction

The 3D extension of HEVC (3D-HEVC) [1] has been developed as the newest coding standard for 3D video with the input format of multiview plus depth (MVD). The depth map can help to synthesize multiple virtual views in series camera locations by Depth-Image-Based Rendering (DIBR) [2] technique. However, the block-based coding structure of 3D-HEVC greatly affects the encoded depth map by blocking effect, which degrades both the prediction accuracy of depth map in encoder and video quality in depth map. Fig. 1 compares the uncompressed depth map and the compressed depth map encoded by 3D-HEVC with different quantization parameters (QP). The depth map degrades evidently at the object edges by the jagged contours when the QP grows. This paper proposes a post-processing method applied on depth map to flatten jagged contours and preserve the necessary edge information. First, the depth map is segmented into three regions by Otsu's automatic segmentation [3]. Then, the edge classification in MPEG-7 [4] and the vision-edge detection acquire the edge information in respect of picture geometry and human perception, respectively. Finally, various filters are implemented on depth map under different region categories and edge types. The rest of the paper is organized as follows. Section 2 describes the proposed post-processing algorithm on depth map. Section 3 demonstrates the experimental results and Section 4 concludes this research.

2. The Proposed Depth Map Post-processing

2.1. Depth Map Segmentation

We separate the depth map into three regions by Otsu's automatic segmentation [3]. First, the histogram distribution is calculated. Then, there are two thresholds (TH1 and TH2) for classifying the depth map into three groups. By Otsu's method, the algorithm will recursively try every combination of TH1 and TH2 to

⁺ Corresponding author. Tel.: +886-03-8634072

E-mail address: 810523001@gms.ndhu.edu.tw

find out the most appropriate thresholds. The segmentation results by the best combination of the two thresholds will make the pixel variance between each group the largest. In other words, the pixel value within the group will be similar, and the pixel value between each group will be distinguishable. As a result, the depth map can be classified into background, middle-ground and foreground by (1), which *PIX* indicates the pixel value. Fig. 2 shows the segmentation results by Otsu's method.



(a) Uncompressed (b) QP34 (e) QP45 Fig. 1: Comparison of the depth map of the Balloons sequence under different QPs in view 1.



Fig. 2: Depth map segmentation of Kendo sequence by Otsu's method.



Fig. 3: Depth edges of Kendo sequence by the edge classification in MPEG-7.

2.2. Edge Classification in MPEG-7

The edge classification in MPEG-7 [4] firstly divides the depth map into various non-overlapped 4×4 blocks. Then, the edge classification in MPEG-7 defines the edge direction into five categories, including horizontal, vertical, 45° , 135° and non-edge. The representative edge is found out by the analysis of the pixel distribution within the 4×4 block. The edges detected by the edge classification in MPEG-7 are shown in Fig. 3, in which the color-marked pixels are with edges while the others are non-edge.

2.3. JNDD Model, JNDD Filter and Vision-Based Edge Detection

Just Noticeable Depth Difference (JNDD) model [5] explores the human visual sensitivity about perceiving the relative change of scene depth in 3D display. The JNDD model can be simply summarized in (2). We can obtain the threshold D_{IND} for each pixel at location (m,n) according to its pixel value $(Od_{m,n})$. It means that people will notice the relative change of scene depth in 3D display once the change of the pixel value in depth map exceeds D_{IND} . The similar experiment is conducted in [6] by considering other environmental stimulations. The value of D_{IND} is added an offset by -10 in [6] and we adopt the modification as shown in (3). In addition, we modify the JNDD mean filter of [5] to JNDD median filter. The modified JNDD median filter smooths the unnoticeable pixel changes in depth map while preserving the original pixel with the noticeable changes.

$$D_{JND} = \begin{cases} 21, \ if \ 0 \le Od_{m,n} < 64:Bin1\\ 19, \ if \ 64 \le Od_{m,n} < 128:Bin2\\ 18, \ if \ 128 \le Od_{m,n} < 192:Bin3\\ 20, \ if \ 192 \le Od_{m,n} \le 255:Bin4 \end{cases}$$
(2)

Furthermore, we adopt the vision-based edge detection in our previous work [7], which is modified from the criterion of Sample Adaptive Offset (SAO) in HEVC [8]. If one of the pixel samplings in Fig. 4 satisfies the condition in Table 1, it means that there is a vision-based edge and we mark all the pixels in the 3×3 blocks as edge pixel.



Fig. 4: Pixel samplings in four kinds of direction.

Table 1: The criteria for detecting the vision-based edge

NO.	Condition	Category
1	$(a - c \ge D_{JND_offset}) \&\&$	
1	$(b - c \ge D_{JND_offset})$	
2	$\left(\left(a-c \ge D_{JND_offset}\right) \&\&(c == b)\right) \parallel$	
2	$((c == a)\&\&(b - c \ge D_{JND_offset}))$	Edaa
2	$\left((c == a)\&\&(c - b \ge D_{JND_offset})\right) \parallel$	Euge
3	$((c-a \ge D_{JND_offset}) \& \& (c == b))$	
	$(c - a \ge D_{JND_offset}) \&\&$	
4	$(c - b \ge D_{JND_offset})$	
0	Non of the above	Non-edge

2.4. Gaussian Filter and Bilateral Filter

In the proposed post-processing method, we also make use of Gaussian filter and Bilateral filter [9][10]. The formulas of Gaussian filter are described in (4)(5). The Gaussian coefficients at location (x,y) of the window coordinate is denoted as G(x,y) in (4), where I(x,y) is the pixel value at position (x,y) and standard deviation (σ) symbolizes the weighting parameter to adjust the Gaussian distribution. The processed pixel value by Gaussian filter at position (m,n) is denoted as $I_G(m,n)$ in (5). Bilateral filter is an edge preserved filter shown in (6)(7), which is controlled by standard deviations of both spatial (σ_d) and pixel range (σ_r) parameters. w(m,n,x,y) indicates the weighting function. (m,n) is the location of the processing pixel. (x,y) denotes the location of the window coordinate. $I_B(m,n)$ is the processed pixel value by Bilateral filter at position (m,n) and I(x,y) represents the pixel value at position (x,y).

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}$$
(4)

$$I_{G}(m,n) = \frac{\sum_{x,y} I(x,y) * G(x,y)}{\sum_{x,y} G(x,y)}$$
(5)

$$w(m,n,x,y) = e^{-\frac{(m-x)^2 + (n-y)^2}{2\sigma_d^2} - \frac{(l(m,n) - l(x,y))^2}{2\sigma_r^2}}$$
(6)

$$I_B(m,n) = \frac{\sum_{x,y} I(x,y) * w(m,n,x,y)}{\sum w(m,n,x,y)}$$
(7)

2.5. Overall Framework of the Post-Processing

For the pixels with vision-based edge, we don't perform any filter on them because any change in these pixels will lead to the perceivable difference in 3D display and we cannot guarantee whether it upgrades the experience of the viewer or not. As a result, the post-processing is only executed in the pixels with non vision-based edges. Besides, among the three regions, we think people focus more on foreground, which is

followed by middle-ground and background. Therefore, more cautious parameter settings with smaller window size (W) is designed for foreground and rougher processing is for background.

In regard to the pixels with edges by the edge classification in MPEG-7, we prefer to apply the Bilateral filter to flatten the unsmooth regions while preserve the important edge information at the same time. For the pixels without any edge information, they are prior. Consequently, the JNDD median filter is applied to deal with the depth map. Gaussian filter is only for the non-edge background region with less necessary information. The detail processing is tabulated in Table 2. In addition, the allowable depth distortion (ADD) model [11] is also combined into the proposed algorithm. The median filter is performed on the whole depth map in the final step.

Table 2: The criteria of the proposed post-processing algorithm under various situations								
		Foreground	Middle-ground	Background				
Non vision-based edge	Edge in MPEG-7	-	Bilateral filter (W =11, σ_d =0.8, σ_r =0.8)	Bilateral filter (W =13, σ_d =1, σ_r =1)				
	Non edge in MPEG-7	JNDD median filter (W=11)	JNDD median filter (W=13)	Gaussian filter ($W=15$, $\sigma=1$)				



(a) Uncompressed







(c) HTM-16.2

(d) Proposed

Fig. 5: Subjective comparison of the synthesized view of Kendo sequence, where Fig. 5(b) shows the enlargement of the ground truth in Fig. 5(a). Fig. 5(c)(d) are the enlargements synthesized by the unprocessed/processed depth map.

synthesized	Δ Average			DD Ditroto	DD DCND
	Bitrate (%)	PSNR (dB)	Time (%)	DD-Ditrate	DD-PSNK
Kendo	0.017	0.015	6.540	-0.379%	0.016
Balloons	-0.072	-0.022	6.707	0.074%	0.000
Poznan_Street	0.023	-0.001	1.917	-0.003%	0.001
Poznan_Hall2	0.051	0.011	-0.181	-0.396%	0.009
Shark	-0.171	0.034	18.978	-1.080%	0.042
Undo_Dancer	-0.090	0.005	1.480	-0.400%	0.012
Average	-0.040	0.007	5.907	-0.364%	0.013

Table 3: Performance of the synthesized view of proposed algorithm compared to the original encoder

3. Experimental Results

Rather than additionally adding the noise model on the depth map to simulate the degradation after encoding and quantization as other researches, we directly implement the post-processing method on the reference software version 16.2 (HTM-16.2) [12][13] to verify how the proposed post-processing algorithm improves the encoding results of the depth map. Various benchmark sequences with two-view case of MVD format are encoded under QP pairs of color texture and depth map containing QP(25,34), QP(30,39), QP(35,42) and QP(40,45). In addition, the quality and coding efficiency of the virtual views are also evaluated to confirm the performance the proposed method. Table 3 tabulates the performance of the synthesized view compared to the original encoder (HTM-16.2). The average PSNR improvement of the synthesized view is 0.007dB. For the coding efficiency, the BD-Bitrate is reduced by 0.364% and the BD-PSNR is increased by 0.013dB, which implies that the progress in the quality of the depth map also apparently promotes the coding efficiency. Fig. 5 demonstrates the subjective comparison of the synthesized view of the Kendo sequence. From the view synthesized by the encoded depth map by HTM-16.2 in Fig. 5(c), there are obvious jagged edges and blocking effect compared to the view synthesized by the

uncompressed depth map in Fig. 5(b). On the contrary, the degradation is significantly removed in the view synthesized by the depth map with the proposed post-processing approach applied as shown in Fig. 5(d), which is much more close to Fig. 5(b). The quality of the synthesized view is well-maintained by our method compared to original encoder.

4. Conclusion

This paper proposes a post-processing technique for depth map to eliminate the degradation of the depth map and improve the quality of the synthesized view. We separate the depth map into foreground, middleground and background by Otsu's method. Then, the edge classification in MPEG-7 and vision-based edge detection are implemented to detect the geometric and vision-based edges. Finally, various filters are applied under various conditions to smooth the depth and preserve the essential edge information at the same time. The experimental results demonstrate that the quality of the depth map and the coding efficiency of the synthesized view processed by the proposed method outperform those of the original encoded videos.

5. Acknowledgements

The authors would like to thank the Ministry of Science and Technology, Taiwan, for supporting this research under grant MOST 106-2221-E-259-009-MY3.

6. References

- [1] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro, and Y. K. Wang. Overview of the Multiview and 3D Extensions of High Efficiency Video Coding. *IEEE Trans. Circuits Syst. Video Technol.* 2016, 26 (1): 35-49.
- [2] C. Fehn. Depth-image-based Rendering (DIBR), Compression, and Transmission for A New Ppproach on 3D-TV. in *Proc. SPIE Conference Stereoscopic Displays and Virtual Reality Systems XI*. 5291(2004): 93-104.
- [3] N. Otsu. A Threshold Selection Method from Gray Level Histograms. *IEEE Trans. Syst., Man, Cybern., Syst.* 1979, 9(1): 62-66.
- [4] ISO/IEC/JTC1/SC29/WG11. MPEG-7 XM Document: MPEG-7 Visual Part Experimentation Model Version 10.0. MPEG Document N4063. 2011.
- [5] D. V. S. X. D. Silva, E. Ekmekcioglu, W. A. C. Fernando, and S. T. Worrall. Display Dependent Preprocessing of Depth Maps Based on Just Noticeable Depth Difference Modeling. *IEEE J. Sel. Topics Signal Process.* 2011, 5(2): 335-351.
- [6] S. W. Jung and S. J. Ko. Depth Sensation Enhancement Using the Just Noticeable Depth Difference. *IEEE Trans. Image Process.* 2012, 21(8): 3624-3637.
- [7] J. R. Lin and M. J. Chen. Fast Depth Coding of 3D-HEVC Based on Visual Perception. in Proc. 2016 Taiwan Academic Network Conference (TANET 2016).
- [8] C. M. Fu, E. Alshina, A. Alshin, Y. W. Huang, C. Y. Chen, C. Y. Tsai, C. W. Hsu, S. M. Lei, J. H. Park, and W. J. Han. Sample Adaptive Offset in the HEVC Standard. *IEEE Trans. Circuits Syst. Video Technol.* 2012, 22(12): 1755–1764.
- [9] C. Tomasi and R. Manduchi. Bilateral Filtering for Gray and Color Images. *in Proc. 1998 Sixth International Conference on Computer Vision*. 1998, 839-846.
- [10] P. Wennersten, J. Strom, Y. Wang, K. Andersson, R. Sjoberg, and J. Enhorn. Bilateral Filtering for Video Coding. in Proc. 2017 IEEE Visual Communications and Image Processing (VCIP). 2017.
- [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 Trans. Image Process.* 2014, 23(11): 4879-4892.
- [12] 3D-HEVC reference software version 15.1 (HTM-15.1), available online at https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSoftware/tags/HTM-15.1/
- [13] Y. Chen, G. Tech, K. Wegner, and S. Yea. Test Model 11 of 3D-HEVC and MV-HEVC. Document JCT3V-K1003. 2015.