EFFECT OF A SYNTHESIZED DEPTH VIEW ON MULTI-VIEW RENDERING QUALITY

Jin Young Lee, Jaejoon Lee, and Du-Sik Park

Samsung Advanced Institute of Technology, Samsung Electronics Co., Ltd.

ABSTRACT

In general, a depth image in a multi-view video plus depth (MVD) format is employed as additional data to synthesize a virtual texture view. In this paper, the proposed method generates a virtual depth view through depth view synthesis and then adaptively uses the synthesized virtual depth view during encoding, according to an evaluation on effect of the synthesized depth view on the multi-view rendering quality. Experimental results demonstrate that the proposed method significantly reduces a depth image coding bit rate without degrading the rendering quality.

Index Terms — Depth map coding, depth view synthesis, multi-view video plus depth (MVD), rendering quality.

1. INTRODUCTION

With the increasing demand of advanced multi-media video systems, such as 3D television (3DTV) and free-viewpoint television (FTV), the 3D display market has been recently getting bigger. 3DTV offers more realistic and natural 3D scenes and FTV provides interactive selection for arbitrary viewpoints [1], [2]. However, such systems require a large amount of the multi-view video information simultaneously captured from lots of cameras. A simulcast coding method was used to store and transmit these multi-view video data. It encodes each view's data separately employing a state-of-the-art codec such as H.264/AVC [3] without consideration of inter-view redundancies between different views. In order to decrease the inter-view redundancies efficiently, a multi-view video coding (MVC) structure in [4] made use of not only temporal but inter-view predictions with a hierarchical B picture. As a result, the prediction structure could achieve the very high coding performance.

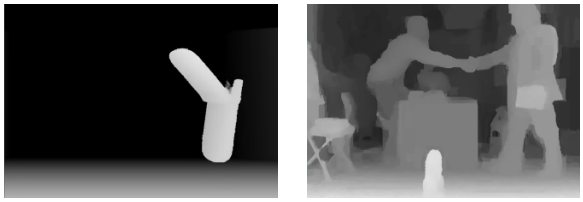
In addition, a multi-view video plus depth (MVD) format [5] is popular for 3DTV and FTV. It enables an encoder to send a small amount of the video data for the several views instead of all the views, because a decoder can synthesize virtual views for the remaining view positions by employing a depth image based rendering (DIBR) method [6]. A depth map means a distance between a 3D point in a visual scene and an optical center of a camera. Unlike a texture image, a depth image predominantly contains flat regions with sharp edges at boundary areas between objects and backgrounds. Also, a depth map is used as extra data to synthesize virtual texture views, so it is not displayed to viewers. Hence, many efficient coding methods specialized on the depth map have been introduced. An edge-aware intra prediction scheme in [7] reduced the prediction error energies in blocks with an arbitrary edge shape. The proposed method in [8] employed a virtual depth view synthesized from the neighboring views as a reference picture in a motion estimation process. In [9], [10], some blocks in depth images were adaptively skipped without encoding, based on correlations of the previously encoded texture images. As a result, a bit rate allocated for the multi-view depth map coding was drastically reduced. In [11], a down-sampled depth image was encoded to decrease a coding bit rate, and then it was reconstructed through up-sampling. In order to maintain clear object boundaries in the reconstructed depth image, the corresponding texture image was employed for the up-sampling. A wavelet-based depth map coding method was presented to improve the rendering quality in [12].

In this paper, we evaluate effect of a synthesized depth view on the multi-view rendering quality. The synthesized depth view is obtained from the neighboring views through virtual depth view synthesis. Also, based on the evaluation, the proposed method adaptively uses the synthesized depth view during encoding at the B view. The remaining of this paper is organized as follows. In section 2, the virtual depth view synthesis is discussed. Section 3 investigates how the synthesized depth view gives an influence on the multi-view rendering quality. In section 4, a simple and efficient depth image coding method is introduced. Finally, experimental results are showed and we conclude the paper in sections 5 and 6, respectively

2. VIRTUAL DEPTH VIEW SYNTHESIS

A general goal of view synthesis is to make virtual texture views with camera geometry information and depth maps. However, we apply a concept of the texture view synthesis to find a virtual depth view of the B view position [8]-[10]. The depth view synthesis is performed by 3D warping the reconstructed depth images at the I and P views into the B view position. First, a real depth value is calculated from a depth pixel with the following equation.

$$Z_{ref}(x,y) = \frac{1}{\frac{D_{ref}(x,y)}{255}\left(\frac{1}{Z_{near}} - \frac{1}{Z_{far}}\right) + \frac{1}{Z_{far}}} \qquad (1)$$



(a) (b)

Fig. 1. Virtual depth view synthesized from the reconstructed neighboring depth views in (a) *Mobile* and (b) *Book Arrival* test sequences.

where $Z_{ref}(x, y)$ and $D_{ref}(x, y)$ indicate a real depth value and an 8-bit intensity of the depth map at a pixel position (x, y), respectively. Z_{near} and Z_{far} denote the nearest and farthest depth values, respectively. Next, by applying a well-known pinhole camera model, a depth pixel (x, y) in the reference view (ref) can be projected into a world coordinate (u, v, w) using the following transformation.

$$[u, v, w]^T = R_{ref} \cdot A_{ref}^{-1} \cdot [x, y, 1]^T \cdot Z_{ref}(x, y) + T_{ref} \qquad (2)$$

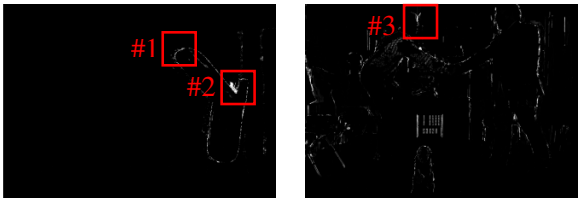
where R, A, and T represent a rotation matrix, an intrinsic matrix, and a translation vector in the camera parameters, respectively. The world coordinate (u, v, w) is mapped into a local coordinate (x', y', z') of the target virtual view (tar) using the following transformation.

$$[x', y', z']^T = A_{tar} \cdot R_{tar}^{-1} \cdot \{[u, v, w]^T - T_{tar}\} \qquad (3)$$

The corresponding pixel location of the virtual image at the target view becomes $(x'/z', y'/z')$. Finally, we can achieve the synthesized virtual depth image by assigning the pixels of the reference view which is closer to the target virtual view into the corresponding pixels of the target one. During the synthesis process, some pixels in the target virtual view are not mapped from the reference view owing to occlusion areas. These positions are generally defined as holes. In our experiment, the holes in the virtual depth view synthesized from the I view were filled with the non-holes in the virtual depth view synthesized from the P view and vice versa. Fig. 1 describes the virtual depth images of the B view position synthesized from the reconstructed depth maps at the I and P views in *Mobile* and *Book Arrival* test sequences.

3. RENDERING QUALITY EVALUATION OF A SYNTHESIZED DEPTH VIEW

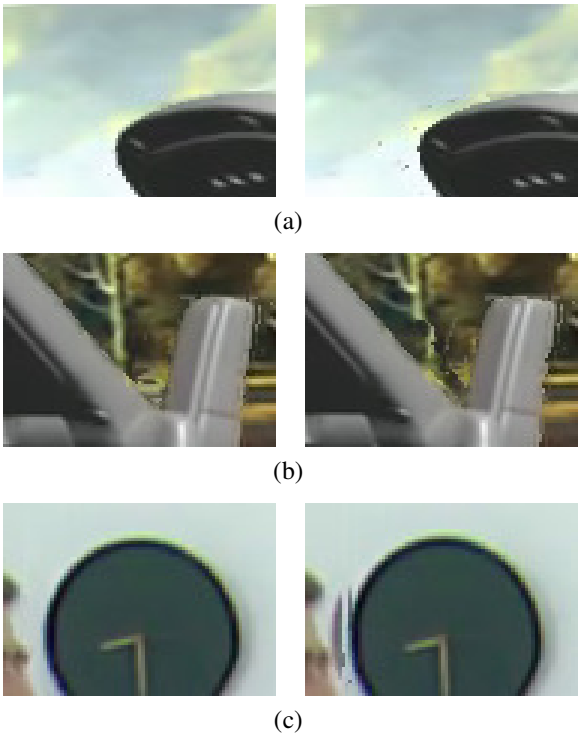
We make use of the virtual depth map synthesized from the reconstructed depth images at the I and P views during the depth image coding at the B view. However, the synthesized virtual depth view generally includes blurring and ghosting artifacts, so it might have a negative influence on the multi-view rendering quality. In order to investigate whether using the synthesized depth view for rendering the virtual views is acceptable or not, we evaluated the quality of the virtual



(a) (b)

Fig. 2. Difference between virtual texture images rendered with the encoded and the synthesized depth images in (a) *Mobile* and (b) *Book Arrival* test sequences.

texture images rendered using the encoded depth maps and the synthesized depth maps, respectively. Fig. 2 illustrates sum of the squared difference between the rendered views. Most of large distortions exist in boundary regions between objects and backgrounds such as marked regions in Fig. 2, while the remaining areas have relatively small distortions. Fig. 3 describes a negative impact of the synthesized depth images on the rendering quality, only for the marked regions in Fig. 2. Some annoying visual artifacts are shown around the boundary regions of the virtual texture images that are rendered using the synthesized depth images, whereas there are scarcely visual artifacts in the virtual images which are rendered using the encoded depth maps. It indicates that the boundary areas in the synthesized depth view have the very bad influences on the rendering quality.



(a)

(b)

(c)

Fig. 3. Example of visual artifacts: Virtual texture images rendered from the encoded (left) and the synthesized (right) depth maps in the marked (a) #1, (b) #2, and (c) #3 areas.

4. EDGE-BASED DEPTH MAP CODING

Based on the rendering quality evaluation of the synthesized view, the proposed method only encodes the areas including object boundaries in the depth map of the B view and skips the remaining ones without encoding. The skipped areas are replaced by the co-located ones in the virtual depth image synthesized from the reconstructed depth maps at the I and P views.

The boundaries are classified by edge information of the synthesized depth image. Edge detection is performed with a gradient of edge vectors ($Gx_{i,j}$, $Gy_{i,j}$), which are calculated as follows.

$$\begin{aligned} Gx_{i,j} &= -1\times p_{i-1,j-1} + p_{i+1,j-1} - 2\times p_{i-1,j} \\ &\quad + 2\times p_{i+1,j} - 1\times p_{i-1,j+1} + p_{i+1,j+1} \\ Gy_{i,j} &= p_{i-1,j-1} + 2\times p_{i,j-1} + p_{i+1,j-1} \\ &\quad - 1\times p_{i-1,j+1} - 2\times p_{i,j+1} - p_{i+1,j+1} \end{aligned} \tag{4}$$

where p denotes a pixel value of the synthesized depth map at a pixel position (i, j), and it is multiplied by a Sobel mask. A magnitude of the gradient is computed as follows.

$$G = \sqrt{Gx_{i,j}^{\ 2} + Gy_{i,j}^{\ 2}} \tag{5}$$

Since all coding modes are processed by a 16×16 block, sum of the magnitudes in the 16×16 block is employed for the edge detection. If the sum is greater than or equal to a threshold value, we determine that the block contains edges and the proposed method compresses the edge block and its adjoining blocks to avoid the visual artifacts around the boundary areas. The threshold value was set to 5000 for our experiment. On the other hand, if the sum is less than the threshold value, the proposed method skips the block.

Fig. 4 shows edge blocks of the synthesized depth maps in *Mobile* and *Book Arrival* test sequences. White denotes a block that includes edges, which represents that it should be encoded due to the great influence on the rendering quality. On the contrary, black corresponds to a block that does not contain edges, which signifies that the proposed method can skip the block instead of encoding. As displayed in Fig. 4, a number of the regions except for the boundaries are skipped, so the proposed method can significantly reduce the depth coding bit rate.

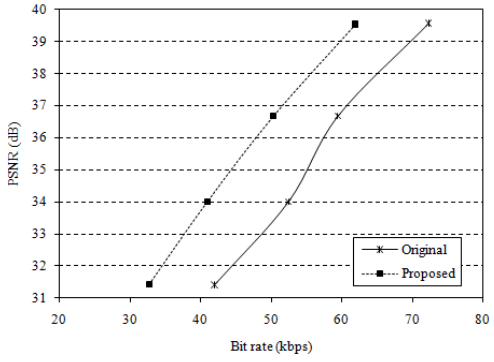
(a) (b)

Fig. 4. Edge blocks of the synthesized depth images in (a) *Mobile* and (b) *Book Arrival* test sequences.

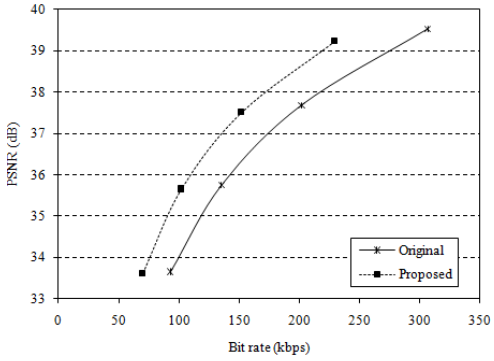
5. EXPERIMENTAL RESULTS

Experiments were performed on reference software JMVC 7.0. We employed two MPEG test sequences, which were *Mobile* (3-5-7 views) [13] and *Book Arrival* (10-8-6 views) [14] sequences, in a three view configuration (I-B-P views) [15]. 100 frames in each test sequence were encoded with GOP of 8 at four different quantization parameters (QP) of 28, 32, 36, and 40. An inter-view prediction was enabled for both anchor and non-anchor pictures. The proposed method was only applied to the depth sequences of the B view. The Bjontegaard Delta bit rate (BDBR) and PSNR (BDPR) [16] were used to measure overall improvement of the proposed method.

Fig. 5 depicts rate-distortion (RD) curves of the original method and the proposed method. They represent a bit rate for the depth map coding at the B view and average PSNR over two middle virtual views. These middle views were synthesized from the decoded texture and depth images at the I and B views and from the decoded texture and depth images at the P and B views, respectively, employing a view synthesis program (VSRS 3.5) from Nagoya University [17]. The uncompressed versions of the middle views were used as a reference [15]. The RD curves demonstrate that the proposed method significantly outperforms the original method. In addition, Table 1 shows the coding performance



(a)



(b)

Fig. 5. RD curves of the original method and the proposed method in (a) *Mobile* and (b) *Book Arrival* test sequences.

Table 1. Coding performance of the proposed method.

Sequences	BDBR	BDPR
Mobile	-18.35%	3.091dB
Book Arrival	-22.70%	1.226dB

of the proposed method. Since it only encodes edge blocks and skips non-edge blocks, the drastically high coding gain can be achieved.

Fig. 6 describes the subjective quality test results for the two synthesized middle views in *Mobile* and *Book Arrival* test sequences. The two middle views were displayed in the Ture3Di SDM-400 42-inch stereoscopic LCD monitor as a stereoscopic view pair. Ten professional subjects took part in the test. The subjects were showed with two stereoscopic view pairs, which were rendered using the depth sequences decoded from the original method and the proposed method, respectively, and then asked to rate their perceived qualities, based on a double-stimulus continuous quality-scale method in ITU-R BT.500-11 [18]. The Y-axis indicates a value that subtracts a score of the proposed method from that of the original method. As shown in Fig. 6, since all the values are near zero, it can be said that the difference in the perceived quality between the original and the proposed methods is unnoticeable. Therefore, we can conclude that the proposed method reduces the coding bit rate for the multi-view depth image without degrading the rendering quality.

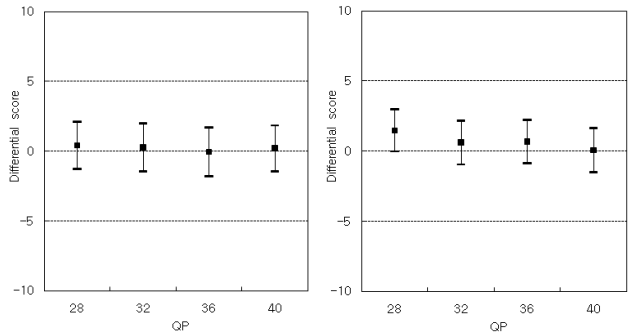


Fig. 6. Subjective quality test results in (a) *Mobile* and (b) *Book Arrival* test sequences. Error bars describe the 95% confidence interval of the mean.

6. CONCLUSIONS

In this paper, we evaluated the effect of the virtual depth view synthesized from the neighboring views on the multi-view rendering quality. The evaluation results showed that boundary areas in the synthesized depth image only cause visual artifacts in the rendered image. Hence, we proposed the edge-based depth map coding method that only encodes the boundary regions containing edge blocks and skips the remaining ones without encoding. The skipped regions are simply replaced by the co-located areas in the synthesized depth map. The experimental results demonstrated that the proposed method significantly reduces the depth coding bit rate while maintaining the rendering quality.

7. REFERENCES

[1] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards," in *Proc. of IEEE Int. Conf. on Multimedia and Exposition*, Jul. 2006.

[2] M. Tanimoto, "Overview of free viewpoint television," *Signal Process.: Image Commun.*, vol. 21, no. 6, pp. 454-461, Jul. 2006.

[3] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264-ISO/ IEC 14496-10 AVC), Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVT-G050, May 2003.

[4] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1461-1473, Nov. 2007.

[5] ISO/IEC MPEG & ITU-T VCEG, "Multi-view Video plus Depth (MVD) Format for Advanced 3D Video System," Doc. JVT-W100, Apr. 2007.

[6] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV," in *Proc. of the SPIE Stereoscopic Displays and Virtual Reality Systems XI*, Jan. 2004.

[7] G. Shen, W. S. Kim, A. Ortega, J. Lee, and H. Wey, "Edge-aware intra prediction for depth-map coding," in *Proc. of IEEE Int. Conf. on Image Processing*, Sep. 2010.

[8] S.-T. Na, K.-J. Oh, and Y.-S. Ho, "Joint coding of multi-view video and corresponding depth map," in *Proc. of IEEE Int. Conf. on Image Processing*, Oct. 2008.

[9] J. Y. Lee, H. Wey, and D.-S. Park, "A novel approach for efficient multi-view depth map coding," in *Proc. of Picture Coding Symposium*, Dec. 2010.

[10] J. Y. Lee, H. Wey, and D.-S. Park, "A Fast and Efficient Multi-View Depth Image Coding Method Based on Temporal and Inter-View Correlations of Texture Images," *IEEE Trans. Circuits and Systems for Video Technology*, to be published.

[11] M. O. Wildeboer, T. Yendo, M. P. Tehrani, T. Fujii, and M. Tanimoto, "Color based depth up-sampling for depth compression," in *Proc. of Picture Coding Symposium*, Dec. 2010.

[12] I. Daribo and H. Saito, "Influence of wavelet-based depth coding in multiview video system," in *Proc. of Picture Coding Symposium*, Dec. 2010.

[13] ISO/IEC JTC1/SC29/WG11, "Philips response to new Call for 3DV Test Material: Arrive book & Mobile," Doc. M16420, Apr. 2009.

[14] ISO/IEC JTC1/SC29/WG11, "HHI Test Material for 3D Video," Doc. M15413, Apr. 2008.

[15] ISO/IEC JTC1/SC29/WG11, "Description of Exploration Experiments in 3D Video Coding," Doc. N11630, Oct. 2010.

[16] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," Doc. VCEG-M33, Apr. 2001.

[17] ISO/IEC JTC1/SC29/WG11, "View Synthesis Algorithm in View Synthesis Reference Software 2.0 (VSRS2.0)," Doc. M16090, Feb. 2009.

[18] "Methodology for the subjective assessment of the quality of television pictures," Rec. ITR-R BT.500-11, 2002.