A Framework of 3D Video Coding using View Synthesis Prediction

Cheon Lee and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST),
123 Cheomdan-gwagiro, Buk-gu, Gwangju, 500-712, Republic of Korea
{leecheon, hoyo}@gist.ac.kr

Abstract— **The advanced 3D video system employs the multi-view video plus depth (MVD) format to support free-viewpoint navigation and comfortable 3D video. Therefore, the prediction structure of the multi-view video coding (MVC) can be used for 3D video coding. The view synthesis prediction method is designed to exploit inter-view correlation using the virtual view generation; hence it is suitable for 3D video coding. In this paper, we propose an efficient framework for 3D video coding using view synthesis prediction to compress multi-view color and depth data simultaneously. We designed the coding procedure of MVD data with four types of view synthesis methods according to the view position. The experimental results showed that the proposed framework improved the coding performance at most 0.9 dB for the multi-view color videos.**

Keywords-3D video coding, view syntehsis prediction, framework for 3D video coding, multi-view video coding.

I. INTRODUCTION (HEADING 1)

Recently, the demand for 3D contents providing more realistic and natural depth impression is growing rapidly. Employing multi-view videos and their corresponding depth data, i.e., multi-view video plus depth (MVD), is one of the promising approaches to support such demand [1]. Using the multi-view video data with depth information generated by content providers, 3D displays can select proper views to render a comfortable 3D scene. However, the encoder needs an efficient video coding method to compress a huge amount of multi-view data. Recently, the moving picture experts group (MPEG) is developing an efficient 3D video coding method to compress the MVD data [2].

The view synthesis prediction (VSP) method is designed for the multi-view video coding (MVC), which employs the view synthesis method using depth data. If the input data consist of only multi-view videos, the depth map should be estimated through encoding and decoding process. Due to the high complexity of depth estimation, it is not desirable for real-time applications. However, since the 3D video system is under development of employing the MVD data, we can utilize the VSP method efficiently. Moreover, the VSP method uses depth data to generate an additional reference frame; the MVD data is suitable for the VSP scheme.

The VSP scheme for multi-view video coding has been introduced by *Martinian* et al. who used depth information to synthesize a picture [3]. Based on this, *Yea* et al. proposed a VSP method that employs an optimal mode decision including view synthesis prediction [4]. Using the basic concept of VSP, we had proposed the view interpolation method and coding method for multi-view video coding including the depth estimation using two neighboring reconstructed views [5]. As a second phase, we had developed a VSP method for multi-view depth video coding as well as the coding method [6]. In this paper, we propose a framework of 3D video coding using the previously developed VSP methods to encode color and depth data simultaneously.

II. PREVIOUS WORKS ON VIEW SYNTHESIS PREDICTION

A. View Synthesis Prediction Method

The view synthesis prediction method has been developed based on the prediction structure of MVC, which utilizes the inter-view prediction and hierarchical B-picture coding [7]. Figure 1 describes the prediction structure using view synthesis prediction [6]. The dotted lines describe the reference view(s) for view synthesis and the solid lines refer to prediction direction. This structure includes the generated frame, i.e., *VS frame*, as an additional reference frame during motion estimation.

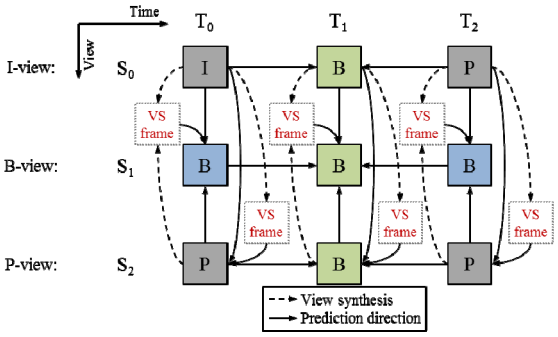


Fig. 1. Prediction structure using view synthesis

B. View Synthesis using Depth Data

Using the depth map, the virtual view image can be obtained via DIBR (depth-image based rendering) [8]. Since the depth map describes the distance of a pixel between the camera and object in a scene, we can define the geometric relations between views. Firstly, we warp the depth image at

the reference viewpoint to the virtual viewpoint by defining the pixel correspondence. Then, we modify the warped depth map using a median filter. Using both the warped depth image and the reference color image, we obtain a synthesized color image of the virtual viewpoint. Next, we fill the hole generated by viewpoint shifting. Finally, we obtain the synthesized image at the virtual viewpoint by blending the synthesized views properly. The whole processes are performed at both encoder and decoder to generate the additional reference frame.

C. *Coding Method: 'VSP modes'*

In the previous work in [6], we designed five additional prediction modes named *VSP modes*; VSP_SKIP, VSP_16x16, VSP_16x8, VSP_8x16, VSP_P8x8. All modes refer to the added VS frame exclusively. Especially the VSP_SKIP mode refers to the co-located block in the VS frame, i.e., its motion is a zero vector. Figure 2 illustrates the encoding structure including VSP modes. In the motion prediction part, the VSP modes compete with the conventional estimation modes. The best mode is selected based on RD-optimization.

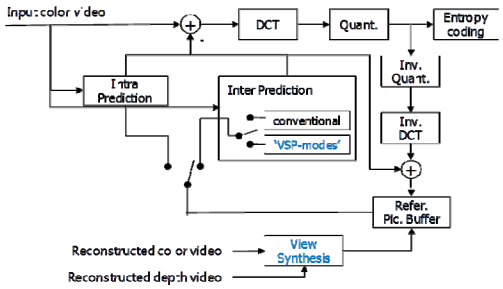


Fig. 2. Encoding structure using 'VSP modes'

The decoder structure is similar with that of the encoder. Using the reconstructed adjacent views, the view synthesizer generates the current view frame using depth map and add it to the reference lists, i.e., LIST_0 and LIST_1.

III. 3D VIDEO CODING FRAMEWORK

A. *Data Format of 3D Video System*

The MVD representation for input data is a popular data format for the 3D video system. Instead of sending multi-view videos to support multi-view 3D displays, the limited number of color videos are sent to the decoder as well as the corresponding depth data. For example, as shown in Fig. 3, the MVD data can be composed of three color videos and their depth data. In fact, the number of views to be transmitted is not fixed; it varies according to the depth composition. If the 3D scene is composed with wide depth range, the encoder may choose sparse views or vice versa. In this work, we use only three views.

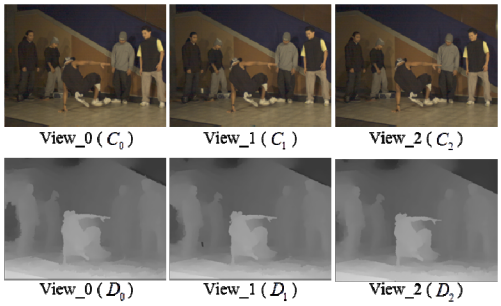


Fig. 3. Multi-view plus depth (MVD) data: three-view case

B. *Framework for 3D Video Coding*

The view synthesis prediction method uses view generation using depth map; hence we propose the coding framework for 3D video coding as illustrated in Fig. 4. Firstly, we encode the leftmost depth video, i.e., D_0, without inter-view prediction; it uses only the hierarchical B-picture coding of the MVC coding structure. We named this coding *I-view coding* since it refers to the view itself. Then we encode the leftmost color view, i.e., C_0 using I-view coding.

After encoding the leftmost view, we encode the rightmost view. Similarly with the leftmost view, we first encode the depth video of the rightmost view, i.e, D_2, but the view synthesis prediction method is employed. By referring to the reconstructed depth data of the leftmost view, i.e., D_0, we synthesize the additional reference frame for the rightmost view and add it to the reference lists. Since this view refers to the I-view frames, we named this *P*-view coding. After encoding depth map of the rightmost view, we encode the color video of the same view, i.e., C_2, by synthesizing the additional reference frame using the reconstructed frame of the leftmost view.

The coding of the center view's video is conducted after encoding both sides' view. Since this view can refer to both I-view and P-view simultaneously, we named this *B-view*

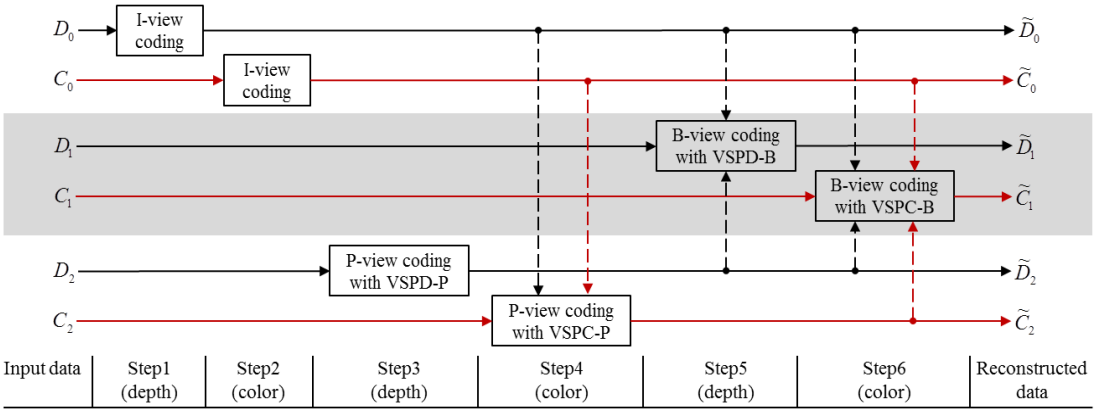
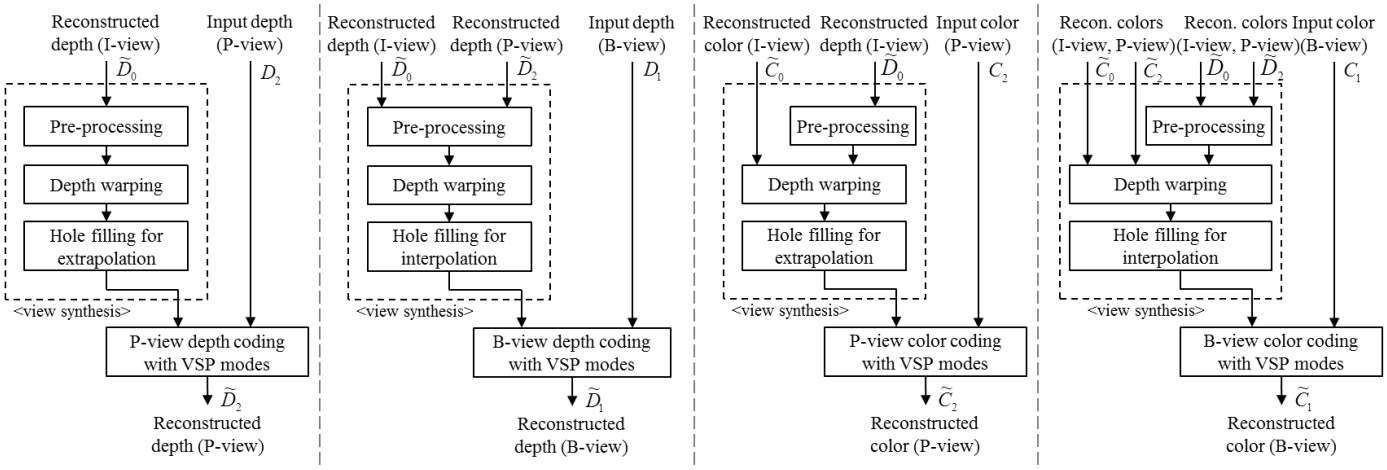


Fig. 4. Proposed framework for 3D video coding using view synthesis prediction



(a) VSPD-P (b) VSPD-B (c) VSPC-P (d) VSPC-B

Fig. 5. Generation methods of additional reference frame for view synthesis prediction

coding. Using both sides' depth images, we synthesize the additional depth image for the center view and add it to the reference lists. For the color video, we use the similar method. Using two depth frames and their color frames, we synthesize the centers' reference frame and add it to the reference lists.

C. View Synthesis Prediction for Depth: P-view Coding

For the depth map of the rightmost view, i.e., D_2, the encoder can refer to the reconstructed depth map of the leftmost view, i.e., $\widetilde{D}_0$, by the prediction structure. Since this view is encoded by P-view coding, we can use the view synthesis prediction for depth coding; hence, we named this coding *VSPD-P coding*. Figure 5.(a) describes the depth map generation for additional reference frame of VSPD-P coding. Since this view can only refer to one reconstructed depth map, we use the hole filling method for extrapolation.

D. View Synthesis Prediction for Depth: B-view Coding

For the depth map of the center view, i.e., D_1, the encoder can refer to two reconstructed depth map from the leftmost- and rightmost views, i.e., $\widetilde{D}_0$ and $\widetilde{D}_2$. Since this view is encoded by B-view coding, we can use the view synthesis prediction for depth coding; hence, we named this coding *VSPD-B coding*. Figure 5.(b) describes the depth map generation for additional reference frame of VSPD-B coding. Since this view can refer to two reconstructed depth maps, we use the hole filling method for interpolation.

E. View Synthesis Prediction for Color: P-view Coding

For the color frame of the rightmost view, i.e., C_2, the encoder can refer to the reconstructed color frame and depth map and of the leftmost view, i.e., $\widetilde{D}_0$ and $\widetilde{C}_0$. Since this view is encoded by P-view coding, we can use the view synthesis prediction for color coding; hence, we named this coding *VSPC-P coding*. Figure 5.(c) describes the color image generation for additional reference frame of VSPC-P coding. Since this view can only refer to one reconstructed view, we use the hole filling method for extrapolation.

F. View Synthesis Prediction for Color: B-view Coding

For the color image of the center view, i.e., C_1, the encoder can refer to two reconstructed color frames and their depth maps from the leftmost- and rightmost views, i.e., $\widetilde{D}_0$, $\widetilde{D}_2$, $\widetilde{C}_0$ and $\widetilde{C}_2$. Since this view is encoded by B-view coding, we can use the view synthesis prediction for depth coding; hence, we named this coding *VSPC-B coding*. Figure 5.(d) describes the depth map generation for additional reference frame of VSPC-B coding. Since this view can refer to two reconstructed color images, we use the hole filling method for interpolation.

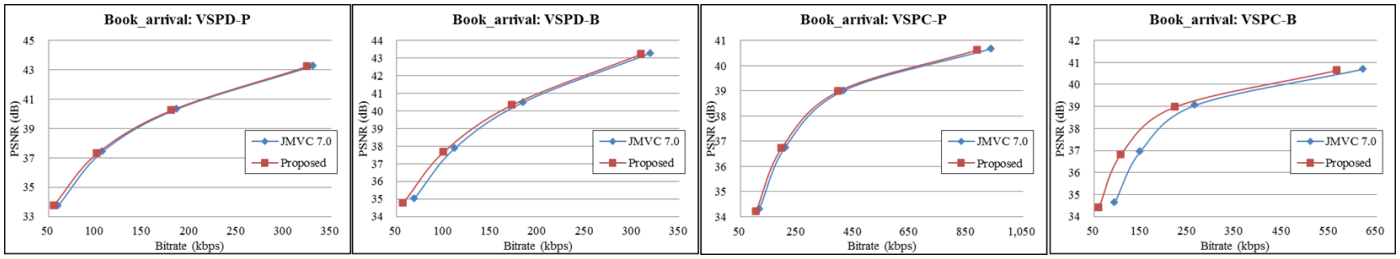
IV. Experimental Results

A. Coding Conditions

In order to evaluate the proposed framework for 3D video coding, we conducted coding experiments under the coding conditions of 3D video coding guided by MPEG 3DV group. We have implemented the view synthesis prediction algorithm into the reference coding software of MVC, i.e., JMVC 7.0. The test sequences for experiments are 'Book_Arrival' and 'Balloons' which are provided by MPEG 3DV group. All the sequences are rectified and color corrected. The depth data are generated by DERS (depth estimation reference software) [9]. When we encode the input data, we used I-B-P prediction structure in MVC coding; it is three- view configuration. The search range was 96. The used QPs for the color videos were 22, 27, 32, and 37; and those of depth videos were higher than 5 (27, 32, 37, 42).

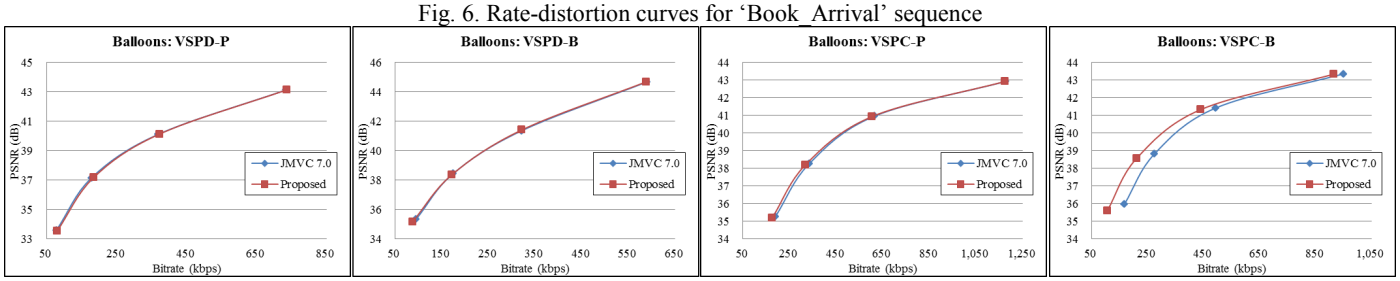
B. Results of Coding Performance

Figure 6 and 7 show the rate-distortion curves for each coding methods. As we mentioned, the highest coding gains are obtained from VSPC-B for both sequences. The second highest gains are obtained from VSPC-P, due to the high correlation between views of the multi-view color videos. On the other hand, the coding results of the depth data showed relatively low gains. The VSPD-B case showed better coding performance than that of the VSPD-P. Table 1 presents Bjontegaard measures for evaluating the coding performance



(a) VSPD-P (b) VSPD-B (c) VSPC-P (d) VSPC-B

Fig. 6. Rate-distortion curves for 'Book_Arrival' sequence



(a) VSPD-P (b) VSPD-B (c) VSPC-P (d) VSPC-B

Fig. 7. Rate-distortion curves for 'Balloons' sequence

Table 1. Bjontegaard PSNR and Bitrate Evaluation

Test Sequence	VSPD-P		VSPD-B		VSPC-P		VSPC-B	
	BDBR (%)	BDPSNR (dB)	BDBR (%)	BDPSNR (dB)	BDBR (%)	BDPSNR (dB)	BDBR (%)	BDPSNR (dB)
Book_arrival	-2.96	0.18	-5.99	0.35	-4.73	0.15	-21.33	0.86
Balloons	+2.01	-0.08	-1.13	0.06	-2.77	0.12	-16.99	0.90

[10]. The highest coding gain were the VSPC-B case of 'Book_arrival' and 'Balloons' sequence; about 21.33% and 16.99% of bit-saving were obtained. However, the VSPD-P case of the 'Balloons' case, small amount of bits was increased due to the view synthesis prediction. This is due to the increased bits for indicating the added 'VSP-modes' in the view synthesis prediction.

The coding performance of the view synthesis prediction method depends on the correlations of views in the multi-view video sequence. Since the color videos are rectified and color corrected, the correlation between views is very high. However, the correlation of depth map is relatively low because those are generated independently. Moreover, coding performance of the view synthesis prediction depends on the quality of the synthesized image. Generally, the quality of the synthesized image of the B-view is higher than that of the P-view case.

V. CONCLUSION

In this paper, we proposed the framework for 3D video coding method using view synthesis prediction. The 3D video system involves multi-view video plus depth data format for supporting wide angle of view for 3D viewing. The view synthesis prediction method is beneficial to this format because the generation of additional reference frame needs depth map. The proposed framework of 3D video coding uses the view synthesis prediction method for both P-view and B-view coding. By using the proposed view synthesis methods for each view, we generated the additional reference frame. Through coding experiments, we obtained coding gains for both views. The B-view coding of a color showed the highest coding performance at most -21.3% bit-saving.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0030822).

REFERENCES

[1] A. Smolić and P. Kauff, "Interactive 3-D Video Representation and Coding Technologies," *Proceedings of the IEEE, Special Issue on Advances in Video Coding and Delivery,*, vol. 93, pp. 98-110, Jan. 2005.

[2] ISO/IEC JTC1/SC29/WG11, "Call for Proposals on 3D Video Coding Technology," *in MPEG output document* N12036, March 2011.

[3] E. Martinian, A. Behrens, X. Jun, and A. Vetro, "View Synthesis for Multiview Video Compression," in *Picture Coding Symposium 2006*, Beijing, China, pp. 1-5, April 2006.

[4] S. Yea and A. Vetro, "View Synthesis Prediction for Multiview Video Coding," *Signal Processing: Image Communication,* vol. 24, pp. 89-100, Jan. 2009.

[5] C. Lee, K.J. Oh, and Y. S. Ho, "View Interpolation Prediction for Multi-view Video Coding," in *Picture Coding Symposium (PCS)*, pp. 1-4, Nov. 2007.

[6] C. Lee, B. H. Choi, and Y. S. Ho, "Efficient multiview depth video coding using depth synthesis prediction," *Optical Engineering,* vol. 20, July 2011.

[7] ISO/IEC JTC1/SC29/WG11, "Study Text of ISO/IEC 14496-10:200x/FPDAM 1 Constrained Baseline Profile and supplementary enhancement information," *in MPEG output document* N10540, April 2009.

[8] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV," in *Stereoscopic Displays and Virtual Reality Systems XI*, pp. 93-104, Jan. 2004.

[9] DERS. *Depth Estimation Reference Software*. Available: http://wg11.sc29.org/svn/repos/MPEG-4/test/trunk/3D/depth_estimation/DERS/DERS

[10] G. Bjontegaard, "Calculation of Average PSNR Differences between RD Curves," *ITU-T SG16/Q6, 13th VCEG Meeting* VCEG-M33, April 2001.