International Journal of Multimedia and Ubiquitous Engineering
Vol.8, No.5 (2013), pp.375-382
http://dx.doi.org/10.14257/ijmue.2013.8.5.37

Efficient Prediction Structure for View Scalability in 3DVC

Jinmi Kang and Kidong Chung

Department of Computer Engineering Pusan National University, Jangjun-dong, Geumjeong-gu, Busan, Korea

wolff98@pusan.ac.kr, kdchung3@melon.cs.pusan.ac.kr

Abstract

With the development of 3D devices such as TV, monitors, and mobile phones, 3D display can support multiple viewpoints in the user's terminal. For these applications, higher coding efficiency for stored or transmitted data is significant, since the number of views has been increased with the technical development of 3D display. To adapt the coder to 2D and diverse views devices, it is important to support view scalability in 3DVC. In this paper, we propose a prediction structure of 3D video to support view scalability considering reliable performance in coding efficiency. In order to use correlation between color video and depth map, an inter-layer prediction structure in SVC is applied to 3DVC. It improves the coding efficiency of the additional depth map with supporting the view scalability. The experimental results show that the proposed prediction structure reduces the bitrates by 0.5~8.3% compared with the reference prediction structure, JMVC 8.5.

Keywords: *3DVC, view scalability, depth map coding, 3D prediction structure*

1. Introduction

3D video technology has been drawing attention from everywhere since 3D movies won success in the market for last several years. With the rise of 3D video industry, studies are conducted on 3D technology, one from stereo video to 3D video coding. MPEG has started supporting 3D video technology by working on the efficient way to compensate disparity between stereo videos using MPEG-2 standard in 1995. Further activities has been organized on standardizing 3D video coding technology, and the results are MPEG-4 Part 2, Part 10, MPEG-A Part 11, MPEG-C Part 3. Joint Video Team (JVT), which has been cofounded by MPEG and VCEG, has completed setting Multi-view Video Coding (MVC) standard as the extension of H.264/AVC in 2009 [1]. Recently, standardization activity on 3D Video Coding (3DVC) is in progress by Joint Collaborative Team on Video Coding (JCT-VC), as there is a need for standardizing depth map data. Depth map is a set of value, which indicates the distance between camera and the subject. It is represented as a grayscale image. A Call for Proposals on 3DVC technology has been issued by MPEG [2], and a new standard for 3DVC is under development in the form of 3D HEVC Test Model Software (3DVC-HTM) based on High Efficiency Video Coding (HEVC) and 3D AVC Test Model Software (3DVC-ATM) based on H.264/AVC, respectively.

3D applications such as 3D monitor and 3D television are mainly binocular display creating 3D effect using stereo videos. For these applications, higher coding efficiency for stored or transmitted data is significant, since the number of views has been increased with the technical development of 3D display. The backward compatibility with 2D videos for a single view device is also required to provide services to users using 2D display device. To

ISSN: 1975-0080 IJMUE

adapt the coder to 2D and diverse views devices, it is important to support view scalability in 3DVC.

Scalability in terms of video coding is a hierarchical coding technology, which allows once-coded video to be adaptively transferred via the various networks according to the end users' environment. Scalable Video Coding (SVC) project has been driven by the JVT, it has been finalized as an amendment of H.264/AVC in 2007 [3]. The standard supports spatial, temporal and quality scalability through hierarchical coding, such as various resolution, control of the frame rate, and diverse quality [4]. In addition, view scalability should be considered for MVC and 3DVC. It is defined view scalability as the functionality that enables the multi-view video coding bitstream to be displayed on a multitude of different terminals and transmitted over varying networks. 3DVC should support view scalability to serve once-coded video data to multi-user terminals with various devices.

Since MVC standard has completed, several coding algorithms to handle view scalability for MVC have been introduced [5, 6]. In [5], an experimental analysis of MVC for prediction structures is presented. The results show that prediction combined with temporal prediction and inter-view prediction is highly efficient. Further, Realistic Multi-view Scalable Video Coding (RMSVC) is proposed for supporting realistic 3D services in [6]. The RMSVC scheme is implemented by integrating the structures of MVC and SVC schemes. However, these algorithms are only designed for color videos. A new 3DVC structure should support extended view scalability to depth map added in 3DVC [7]. In brief, view scalability is needed to send 3D video over the various networks with proper number of view which end-user terminal can technically support.

In this paper, we propose a prediction structure of 3D video to support view scalability considering reliable performance in coding efficiency. We present scalable prediction structure exploiting the correlation between color video and depth map. The motion information of the depth map is similar to the corresponding color video, as the contours of objects are the same. Many studies are conducted on the idea exploiting the correlation between color video and depth map to improve coding performance. We use a layered structure of SVC when coding depth map in order to use the correlation. SVC structure is composed of a base layer and enhancement layers using inter-layer prediction. Inter-layer prediction is used to remove redundancy of motion information among the layers, including macroblock partition, reference picture indices and motion vector. By exploiting the inter-layer prediction, we can improve the coding efficiency of depth map. The proposed structure can support view scalability with existing 2D devices, due to the compatibility with H.264/AVC of the color video as base layer.

The rest of this paper is organized as follows. The MVC structure and view scalability are introduced in Section 2. We briefly explain the prediction structures of MVC and the view scalability. The proposed prediction structure of 3DVC supporting view scalability is described in Section 3. Experimental results are given in Section 4 to show the efficiency of the proposed method. Finally, we conclude this paper in Section 5.

2. Typical Prediction Structures

2.1. Inter-layer Prediction

The basic design of the SVC can be classified as layered video coding. A SVC bitstream consists of a base layer and several enhancement layers. In each layer, the schemes for hybrid video coding such as temporal prediction and intra prediction are employed as in standard H.264/AVC. The bitstream of the base layer is compatible with H.264/AVC for existing 2D devices. To provide spatial, temporal and quality scalabilities, SVC offers several additional

techniques. The inter-layer prediction is one of them to improve the coding efficiency as illustrated in Figure 1. If the enhancement layer has corresponding base layer, the information of the base layer is used as a reference by inter-layer prediction. According to predicted information, there are three approaches such as intra, motion and residual. In particular, inter-layer motion prediction is applied to our prediction structure for 3DVC. Fig. 1 shows the hierarchical B prediction structure to temporal direction with a group of pictures (GOP) length of 8. In the base layer, the first picture of a GOP are coded by using intra prediction as an instantaneous decoder refresh (IDR) picture and called key pictures. The remaining pictures of a GOP are hierarchically predicted by using the two nearest pictures of the next higher temporal level as references.

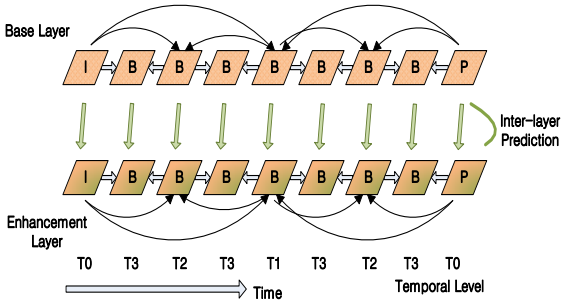

Figure 1. Inter-layer Prediction

2.2. Inter-view Prediction

Figure 2 shows a straightforward structure to compress multi-view videos, and it is called simulcast coding. Each view video is encoded like single view independently. Simulcast is usually used as a reference to compare with prediction structures that additionally use inter-view prediction. Inter-view prediction is a key feature of the MVC design. Since the cameras of a multi-view capture the same scene from adjacent views, there is a redundancy among the views. To improve the coding efficiency of MVC, this is applied as inter-view prediction by removing inter-view redundancy.

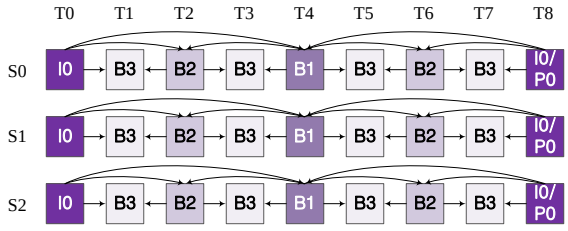

Figure 2. Simulcast prediction structure

The typical MVC structure with inter-view prediction is illustrated in Figure 3. The prediction structure uses the hierarchical B structure for temporal prediction and the IBPBP structure for inter-view prediction. The prediction structure of the first view is called base view, as it is identical to the simulcast prediction structure with hierarchical B pictures for temporal prediction only. But for the other views, all intra-coded key pictures are replaced by inter-coded pictures using inter-view prediction. For the remaining pictures of each GOP, the prediction structure does not change and remains to be temporal prediction only. Furthermore,

synchronization and random access features are provided by still coding the key pictures of the base view in intra mode.

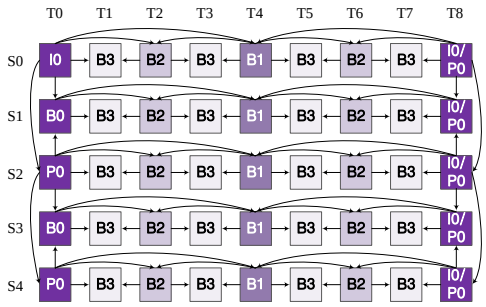

Figure 3. MVC Prediction structure (IBP)

2.3. Prediction Structures for View Scalability

View scalability is a technique that allows single bitstream to support multiple views. Since various 3D applications such as television, internet streaming video and mobile each have different views to display, 3D video server need to support view scalability for enhancing the interoperability. In MVC prediction structure, the most methods to improve coding efficiency are restricting the inter-view structure. These schemes are only based on color video of multiple views. In 3D coding structure, it is difficult to apply these structures directly. Because 3D video is consist of color video and additional depth map, MVC prediction structure makes it difficult to remove redundancies between color video and depth map. The 3D prediction structure is needed to remove the redundancy between color video and depth map. In addition, it is required to provide the view scalability.

3. Prediction Structure for View Scalability

As seen above, the current 3D video is composed of conventional color video and additional depth map. The 3D video coding standard focuses on supporting backward compatibility with conventional video coding standard, and increasing depth map coding efficiency. There are two different standards for 3D video, namely 3D-ATM which is compatible with MVC, and 3D-HTM which is compatible with HEVC. In this study, we propose a 3D video coding structure considering 3D-ATM standard.

Depth map is an image of 8 bit grayscale pixels which represents the distance between camera and object. Depth map is not projected on the screen, but it is used to create multi-view video more efficiently than by transmitting multi-view images synthesized in the receiver. Motion information of the object in depth map is similar to that of the corresponding object in color video of the same view. Thus, studies are conducted to increase coding efficiency using this correlation between color video and depth map. Nevertheless, existing studies focus on increasing coding efficiency of depth map without considering compatibility with the other standard.

The most basic multi-view video coding structure is identical to MVC prediction structure. Figure 4 shows the basic structure that color video(a) and depth map(b) are independently encoded with MVC prediction structure. This structure cannot exploit the advantage of depth map because inter-view prediction is only used. The coding efficiency of depth map can be improved by removing redundancy of depth map and color video.

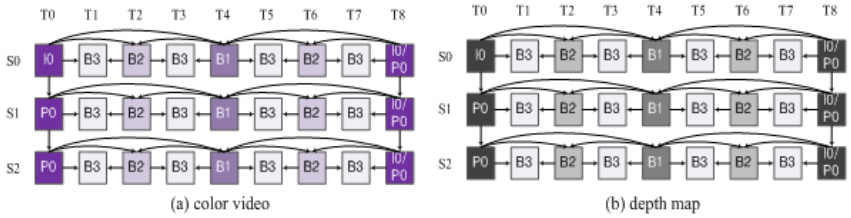

Figure 4. MVC prediction structure for 3DVC

In this paper, we propose 3D video coding structure for efficient depth map coding, which is also compatible with the existing methods for coding videos. Above all, we adapt SVC structure in order to exploit the similarity of the motion vector between depth map and color video. The color video is encoded as a base layer and depth map as an enhancement layer, as shown in Figure 5. In the enhancement layer of SVC, there is additional inter-layer prediction process using texture, motion, and residual information of the base layer which results in higher coding efficiency. The inter-layer prediction in our proposed depth map coding structure uses only motion information of color video. The reason for using motion information only is that there is less similarity between depth map and texture of color video. Hereby, the same concept is used for depth map coding, using motion information of the base layer, i.e., the color video, in order to improve the coding efficiency. Moreover, the color video as a base layer is still compatible with the existing H.264/AVC device, as it is encoded in the same manner of H.264/AVC.

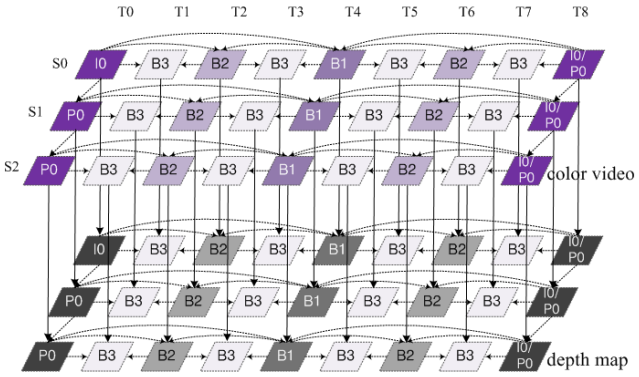

Figure 5. Proposed prediction structure

Entire coding structure of our proposed method is shown in Figure 6. Color video as a base layer, is encoded in a way equal to conventional standard MVC. Encoded motion information of color video is used in the motion estimation process of depth map coding as an enhancement layer. Inter-layer motion prediction is used to remove the redundancy of motion information among layers, including macroblock partition, reference picture indices and motion vectors. This motion information is redundant as the movement of the object in each depth map and color video is the same. Therefore, redundant motion information can be removed by encoding depth map as an enhancement layer. SVC structure is used to fully exploit the correlation between depth map and color video. Therefore, coding efficiency can be improved while achieving higher adaptability to the various network environments.

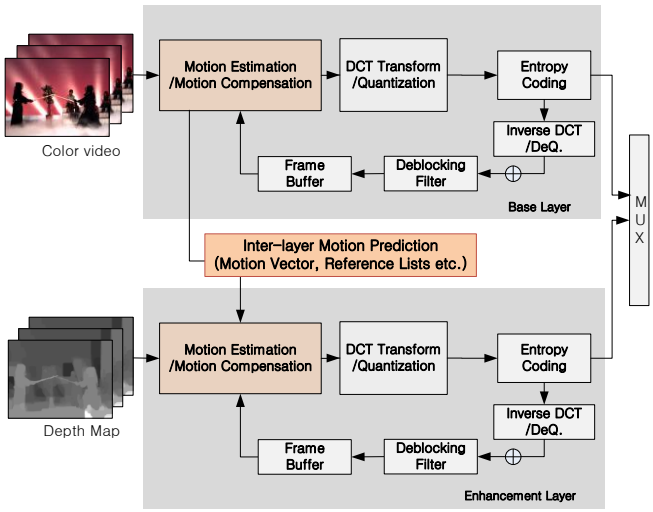

Figure 6. Entire coding structure

4. Experimental Results

In this section, we evaluate the performance of our proposed prediction structure for view scalability. We focus on the coding efficiency of the depth map prediction structure in 3DVC with several views. The simulations have been implemented in the reference software JMVC 8.5 [8]. To verify the proposed structure for the view scalability, three views are simulated. We have used four video sequences such as Kendo, Balloons, Café, and Newspaper in three view configuration (I-B-P views). Table 1 shows the configuration of the coded test sequences. The GOP size is 8 and, it means that there are three hierarchical prediction levels. For the comparison of the coding efficiency, five quantization parameters (QP) between 26 and 46 were used to obtain five rate points.

Table 1. Coding configuration

Sequence	Picture Size	GOP	Number of coded pictures
Kendo	1024 x 768	8	100
Balloons	1024 x 768	8	100
Café	1920 x 1080	8	50
Newspaper	1024 x 768	8	100

We compare the bitrates of coding between the MVC structure and the proposed structure. In this case, we only consider the depth map coder that newly added in 3DVC. In proposed structure, the color video can be coded by exploiting intra prediction and temporal prediction, and this makes it compatible with H.264/AVC.

The experimental results are given in Table 2. This is the results of only depth map coding in the enhancement layer, to evaluate the coding efficiency. The performance measures include bitrates and the peek signal-to-noise ratio of luminance component (Y-PSNR) for testing the coding efficiency and the quality of reconstructed sequences, respectively. The bitrates represented the coding efficiency is calculated by averaging the total coded bitrate with five initial QP for each view. As can be seen from Table 2, our method has achieved a significant savings from 0.5% to 8.3% which translates to Y-PSNR gains from 0.2dB to

3.1dB. From the results, we could know that the coding efficiency of proposed structure is always better than that of JMVC 8.5. As expected, the bitrate savings of the depth map is improved due to apply the color motion information.

Table 2. Experimental results

sequence	view	PSNR	Bit rate		*Δ bitrate(%)*
			JMVC	Proposed	
Kendo	0	40.19	272.37	261.26	**4.25**
	2	38.86	349.38	342.85	**1.91**
	1	40.48	238.17	237.00	**0.49**
Balloons	0	40.20	239.58	225.95	**6.03**
	2	38.49	300.14	293.64	**2.21**
	1	39.67	247.03	245.92	**0.45**
Café	0	43.08	403.99	378.56	**6.72**
	2	41.02	387.75	374.78	**3.46**
	1	40.91	352.91	348.00	**1.41**
Newspaper	0	40.07	223.45	206.31	**8.31**
	2	39.70	209.32	202.12	**3.56**
	1	38.64	242.44	240.99	**0.60**

We also depict the rate-distortion (RD) curves of the Y-signal at Figure 7. We could see that the RD curves at the last bit rates interval are larger than those of the JSVM 8.5. The RD curves show that the proposed method can have more improvement without using additional buffer or parameters. The proposed prediction structure also can support view scalability.

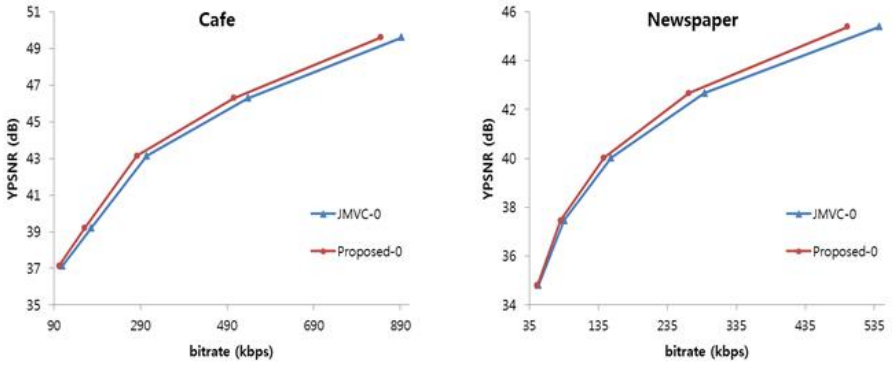

Figure 7. RD curves *café* and *newspaper*

5. Conclusion

In this paper, a new prediction structure of 3DVC has been studied by using the correlation between color video and depth map for view scalability. In the proposed structure, the inter-layer prediction of SVC is applied to exploit the motion information of color video. Color video as a base layer is encoded in a way equal to conventional standard, and it makes

compatible with H.264/AVC and MVC. The proposed method can support view scalability with 2D for a single view and diverse views devices. It also improves the coding efficiency of depth map with supporting the view scalability. The experimental results prove that the proposed prediction structure performs better than that of JMVC 8.5 in bitrates by 0.5~8.3%.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2011-0014547).

References

[1] A. Vetro, T. Wiegand and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard", Proceedings of the IEEE, **(2011)**, pp. 626-642.
[2] ISO/IEC JTC1/SC29/WG11 N12036, "Call for Proposals on 3D Video Coding Technology", Geneva, **(2011)**.
[3] Amendment G of Information technology (H.264/SVC)- Coding of audio-visual objects - Part 10: Advanced video coding, ISO/IEC 14496-10, **(2010)**.
[4] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC standard", IEEE Trans. Circuits and Systems for Video Technology, vol. 17, no. 9, **(2007)**, pp. 1103-1120.
[5] P. Merkle, A. Smolic, K. Muller and T. Wiegand, "Efficient Prediction Structures for Multiview Video Coding", IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 11, **(2007)**, pp. 1461-1473.
[6] P. Min-woo and P. Gwang-hoon, "Realistic multi-view scalable video coding scheme", IEEE Trans. Consuer Electronics. vol. 58, no. 2, **(2012)**, pp. 535-543.
[7] S. Shimizu, M. Kitahara, H. Kimata, K. Kamikura and Y. Yashima, "View Scalable Multiview Video Coding using 3-D Warping With Depth Map", IEEE Trans. Circuits and Systems for Video Technology, vol. 17, no. 11, **(2007)**, pp. 1485-1495.
[8] JMVC 8.5 reference software from CVS server, garcon.ient.rwth-aachen.de/cvs/jvt.

Authors

Jinmi Kang received the B.S. and M.S. degrees in computer engineering from Pusan National University, Busan, Korea, in 2003 and 2005, respectively. She worked for Mobile communication R&D office in LG electronics from 2005 to 2007. After 2 years working, she finished Ph.D. course in same university in 2009. His research interests include SVC, depth map coding, 3D video coding and multimedia systems.

Kidong Chung received the B.S. degree from Seoul National University, Seoul, Korea, in 1973, and M.S. and Ph.D. degree in computer science from the same university in 1975 and 1986. Since 1978, he has been a Professor of Computer Science at Pusan National University, Pusan, Korea. His research interests include 3D video coding, multimedia systems, and mobile multimedia communication.

