INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2009/M16062
February 2009, Lausanne, CH

Source: GIST (Gwangju Institute of Science and Technology)
Status: Report
Title: EE4: Coding Results on 'Pantomime' Sequence
Author: Cheon Lee and Yo-Sung Ho

1. Introduction

This document reports the experimental results of coding experiment on 'Pantomime' sequence as a response to EE4 [1]. We conducted several coding experiments and analyzed the results. We followed the general coding conditions of MVC (multiview video coding) [2]. The reference software is 'JMVC 3.0' which is the last version of MVC codec [3]. There are two types of input data; two color videos and their corresponding depth videos. We have coded only 'Pantomime' sequence as we were assigned in Busan meeting. We analyzed coding results not only the quality of synthesized images generated by view synthesis but also the quality of reconstructed images of the original input data.

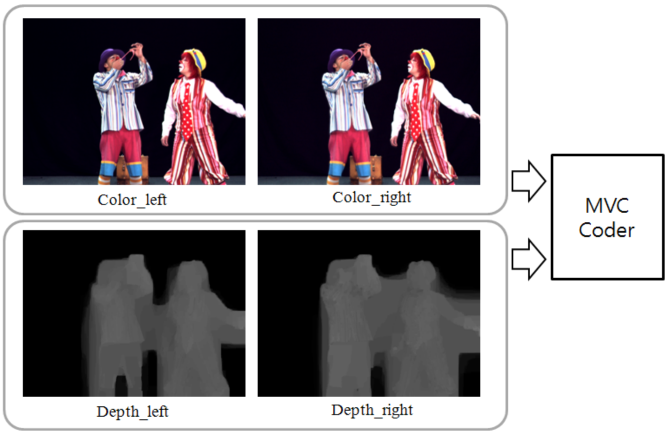

Fig. 1. Input Data of 3D Video Coder

2. Conditions of Coding Experiment

3D video coding includes the depth video coding as well as multiview video. As described in EE description document, we conducted coding experiments with two kinds of input data. Figure 1 shows an example of input data. As we are assigned to check 'Pantomime' sequence case, we got the depth data obtained by EE1 and tested in coding.

2.1. Coding Scenarios

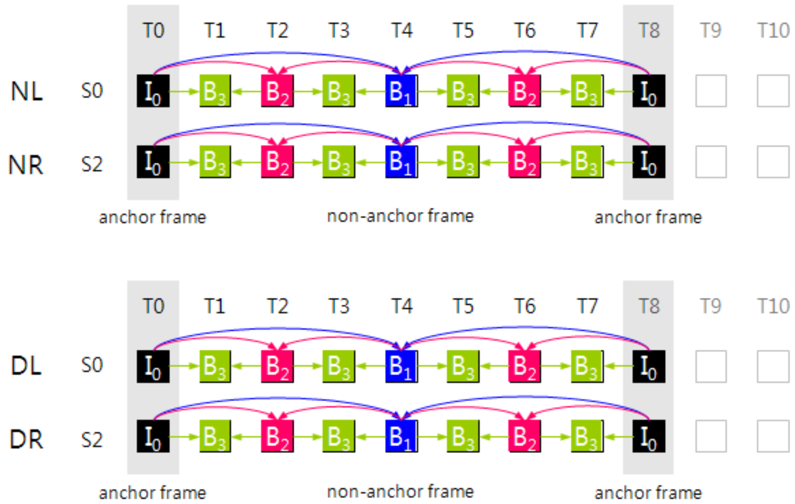

Fig. 2. Coding Scenario 1: ‘Simulcast’

Since MVC developed efficient coding structure for multiview video situation, we can use the basic structure of coding such as prediction structure and test conditions. In order to glance at the property of 3D video coding, we set three scenarios as described in Fig. 2 to 4. We conducted following three scenarios and analyzed.

Figure 2 shows the first scenario of coding so called ‘Simulcast’. All videos including depth video are coded independently using ‘JMVC 3.0’ codec maintaining hierarchical-B coding structure. There is no prediction between views. This scenario is possible to any codec as well as MVC codec.

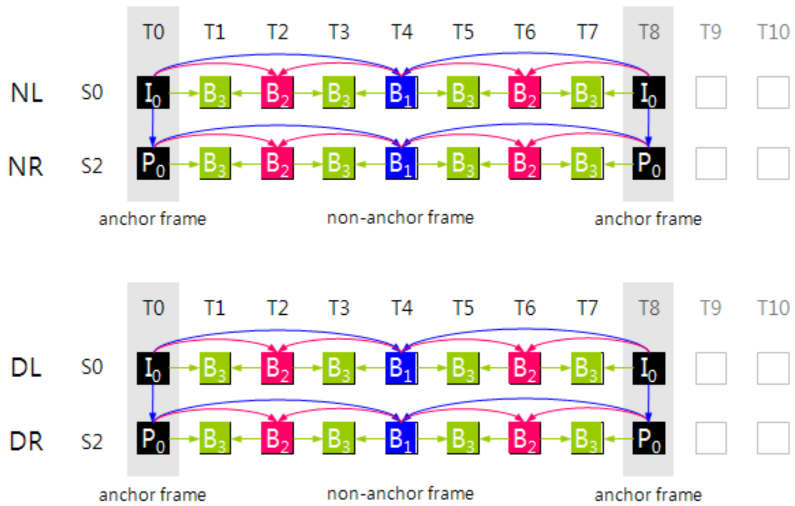

Fig. 3. Coding Scenario 2: ‘Inter-view Prediction Structure only for Anchor Frames’

The scenario 2 shown in Fig. 3 is ‘Inter-view Prediction Structure only for Anchor Frames’. Because MVC software is including inter-view prediction structure, we can test it for 3D Video data. Practically, we allowed anchor frames to refer to the adjacent anchor frames.

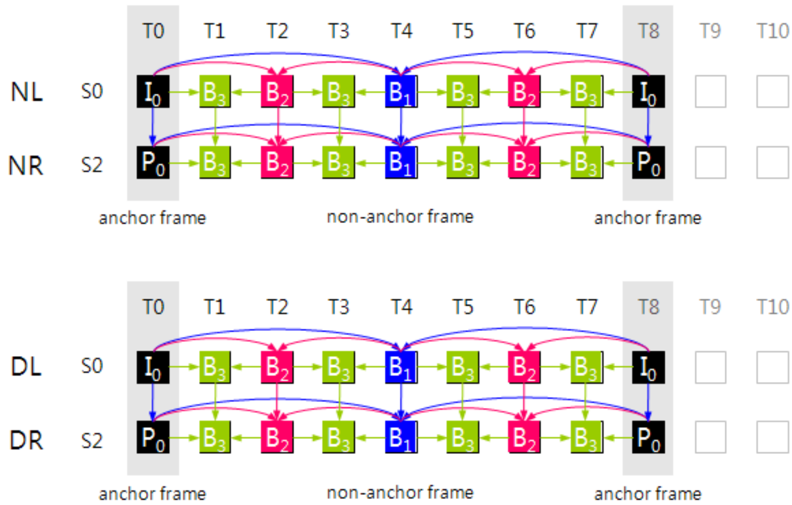

Fig. 4. Coding Scenario 3: 'Inter-view Prediction Structure for All Frames'

Similarly to scenario 2, scenario 3 is 'Inter-view Prediction Structure for All Frames'. Every frame in the right side video refers to the inter-view frame in the left view. By checking coding efficiency of those three scenarios, we could find inter-view dependency of data.

2.2. Coding Conditions

Except for QP parameters, almost of all parameters are free of selection. With above three structures, we selected a parameter set as Table 1.

Table1. Coding Condition

Number of Frames	196
Entropy Coder	CABAC
FRExt	On
Basis QPs	{22, 25, 28, 31}
GOP	15
Search Mode	Fast
Search Range	32
Rate Control	Off
Loop Filter	On

3. Experimental Results

In this chapter, we report the coding result on 3D Video data. Result of each scenario is presented in from Table 2 to 4. We mentioned above, we have coded two input data; color video and depth video. Since we are using two reference views for view synthesis, we coded two views such as view 38 and view 41.

As shown in Table 2, the PSNR value for the depth is higher as much as 8 dB on average than that of color video. In addition, bit-rate for depth video spends less 10 times than color video case.

Table 2. Coding Result of Scenario 1

Color Video								
	QP22		QP25		QP28		QP31	
Viewpoint	C38	C41	C38	C41	C38	C41	C38	C41
PSNR (dB) (average)	43.991	43.01	42.659	41.864	41.150	40.454	39.526	39.083
	43.500		42.261		40.847		39.305	
Bit-rate (kbps) (average)	3439.7	3617.7	2468.5	2514.3	1788.9	1796.6	1342.3	1328.1
	3528.7		2491.4		1792.7		1335.2	
Depth Video								
	QP22		QP25		QP28		QP31	
Viewpoint	C38	C41	C38	C41	C38	C41	C38	C41
PSNR (dB) (average)	43.189	52.211	41.423	50.560	49.26	48.498	47.533	46.184
	52.700		50.992		48.882		46.858	
Bit-rate (kbps) (average)	302.0	372.9	196.3	242.7	124.9	152.9	79.4	94.1
	337.4		219.6		138.9		86.7	

Comparing with scenario 1, scenario 2 showed better coding efficiency. It gives us that the dependency between views is high, and it can be exploited for coding.

Table 3. Coding Result of Scenario 2

Color Video								
	QP22		QP25		QP28		QP31	
Viewpoint	C38	C41	C38	C41	C38	C41	C38	C41
PSNR (dB) (average)	43.991	43.016	42.659	41.873	41.150	40.517	39.526	38.973
	43.504		42.266		40.834		39.249	
Bit-rate (kbps) (average)	3439.7	3354.0	2468.5	2272.4	1788.8	1572.5	1342.3	1135.0
	3396.9		2370.5		1680.7		1238.6	
Depth Video								
	QP22		QP25		QP28		QP31	
Viewpoint	C38	C41	C38	C41	C38	C41	C38	C41
PSNR (dB) (average)	53.189	52.187	51.423	50.483	49.266	48.362	47.533	44.797
	52.688		50.953		48.814		46.165	
Bit-rate (kbps) (average)	302.0	375.5	196.4	243.2	124.9	153.2	79.4	91.8
	338.8		219.8		139.0		85.6	

Table 4. Coding Result of Scenario 3

Color Video								
	QP22		QP25		QP28		QP31	
Viewpoint	C38	C41	C38	C41	C38	C41	C38	C41
PSNR (dB) (average)	43.991	43.020	42.659	41.867	41.150	40.493	39.526	38.915
	43.506		42.263		40.822		39.221	
Bit-rate (kbps) (average)	3439.7	3090.3	2468.5	2058.3	1788.9	1393.8	1342.3	990.3
	3265.0		2263.4		1591.3		1166.3	
Depth Video								
	QP22		QP25		QP28		QP31	
Viewpoint	C38	C41	C38	C41	C38	C41	C38	C41
PSNR (dB) (average)	53.189	52.023	51.423	50.314	49.266	48.198	47.533	44.620
	52.606		50.869		48.732		46.076	
Bit-rate (kbps) (average)	302.0	374.9	196.4	242.3	124.9	151.8	79.4	90.2
	338.5		219.3		138.3		84.8	

When a coder refers to all adjacent views, the coding efficiency of color video was the best among three scenarios as presented in Table 4. This means that although hierarchical B picture coding uses temporal redundancy mainly, inter-view prediction is more efficient than temporal prediction in some region. If we can exploit these redundancies between views, we can develop an efficiency coding method for 3D video. Figure 5 and 6 show the comparison of rate-distortion curves for every scenario.

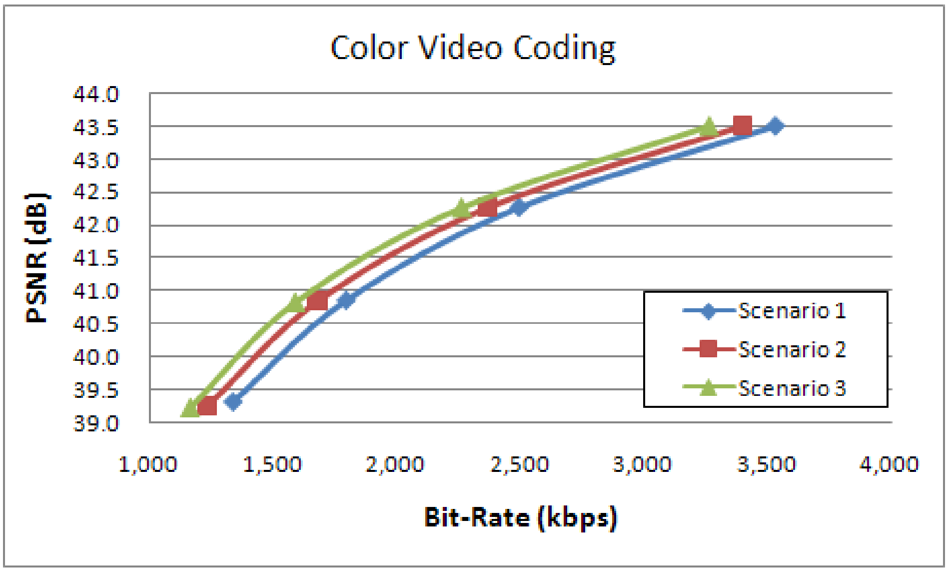

Fig. 5. Comparison of Results on Color Video

Comparing to those two types of sequence, the inter-view prediction method on color video is more efficient than depth video case. In high bit-rate, coding efficiency of depth video was similar no matter what the prediction structure is. It is that the dependency in temporal direction is higher on color video than that of depth video.

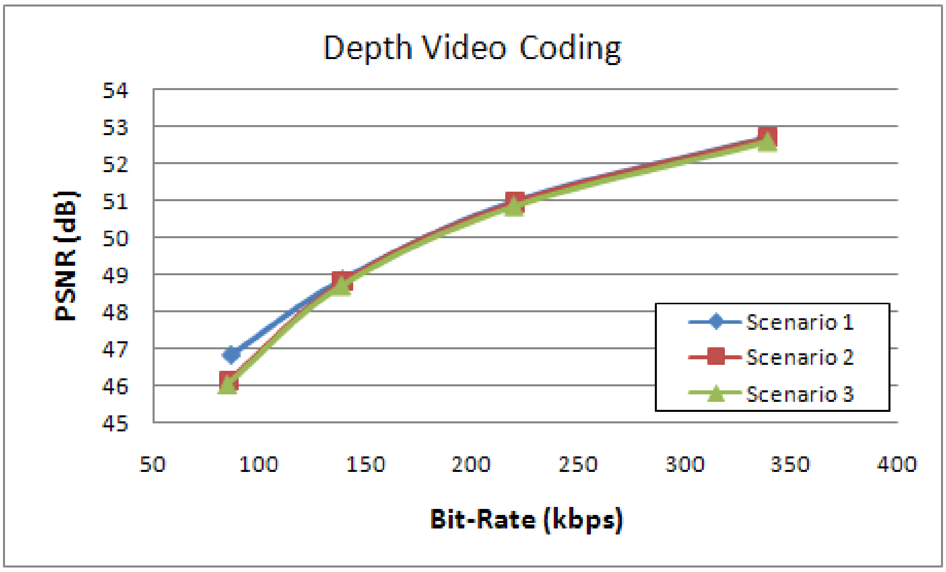

Fig. 6. Comparison of Results on Depth Video

Another quality measure could be used for coding experiment is considering the quality of synthesized image because the purpose of depth coding is to render intermediate view images using depth map. We synthesized two intermediate view images using 'VSRS' software which is provided by Nagoya University, then we calculated PSNR value of it with the original image. Table 5 represents PSNR values of the synthesized images. One considerable property is that the drop rate of PSNR on the synthesize image is lower than that of reference color video as the QP value is increase. Another property is that the average PSNR on the synthesized image is lower than that of reference color video. For instance, if the QP is 22 for color video, the PSNR of the reconstructed image for view 38 is 43.991 dB. However, when we synthesize an intermediate view image using 'VSRS', the PSNR value for view 39 is 35.6 dB. The quality of the synthesized image has been dropped significantly.

Although the PSNR value of the synthesized image is lower that of reference image, the performance of synthesis tool shows quite stable. As shown in Fig. 6, we can see the synthesized image is quite good.

Table 5. Synthesized Result using Reconstructed Data

	QP22		QP25		QP28		QP31	
Viewpoint	C39	C40	C39	C40	C39	C40	C39	C40
Scenario 1	35.6	32.9	35.4	32.8	35.1	32.6	34.7	32.3
Scenario 2	35.6	32.9	35.4	32.8	35.1	32.6	34.7	32.4
Scenario 3	35.6	32.9	35.4	32.9	35.2	32.8	34.7	32.6

Fig. 6. Synthesized Image on ‘Pantomime_39’ using Decoded Data

4. Conclusion

We have explained about the coding results of 3D video for ‘Pantomime’ sequence. Among input data, we used the obtained depth video from EE1. We tested three scenarios in order to analyze dependency between views. In depth video case, view dependency is less than that of video case. However, the bit-rate for depth coding is much less than video coding. Using the reconstructed data, we synthesized the intermediate view images using ‘VSRS’ tool and calculated the PSNR values. The synthesized image was acceptable because there was no artifact on the image except for the quantization error.

5. Acknowledgements

This work was supported in part by ITRC through RBRC at GIST (IITA-2009-C1090-0901-0017).

6. References

[1] ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding,” N10173, October 2008.

[2] ISO/IEC JTC1/SC29/WG11 “Description of Core Experiments in MVC,” N8019, April 2006.

[3] JVT of ISO/IEC MPEG & ITU-T VCEG, “WD 3 Reference software for MVC,” AC207, October 2008.