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/M16673
April 2009, London, England

Source: GIST (Gwangju Institute of Science and Technology)
Status: Proposal
Title: Bilateral Filtering of Depth Map for 3D Video Coding
Author: Cheon Lee and Yo-Sung Ho

1. Introduction

This document proposes a bilateral filtering for depth video which is generated by depth estimation. Ever since the EE started, experts evaluated the quality of depth video by testing the synthesized image. Definitely, the performance of depth estimation has been improved. Hence, the EE4 about coding experiments has been organized during the last Maui meeting because experts agreed that the estimated depth video is acceptable for reference [1]. However, we questioned the quality of the generated depth map. The generated depth map has a lot of edges through the whole image. It might consume high bit rates for coding. An efficient low pass filter maintaining significant depth changes can reduce the bits for depth coding. In this document, we propose a bilateral filter for depth map, which keeps the quality of the synthesized image and reduces the bit rate for depth coding. This filter can be included in depth estimation as a post-processing.

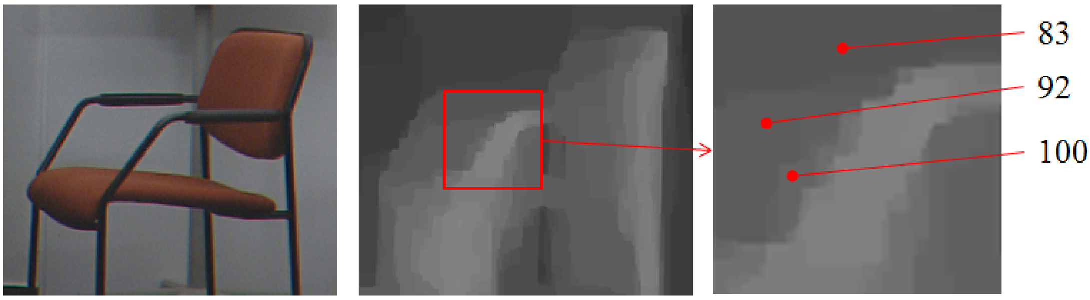


Fig. 1. ‘Book_arrival’ image and its depth map generated by DERS 4.1.

2. Generated Depth Map by DERS and Depth Coding

The depth estimation reference software on testing by MPEG 3D video group exploits a Graph Cuts algorithm to optimize the disparity value [2][3]. Since the graph cuts algorithm is one of state-of-the-art method in pattern classification, it guarantees high accuracy of the disparity estimation. However, we found there are unwanted properties in the resultant depth map as follows;

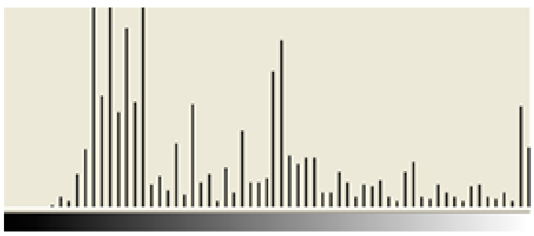
- Depth values are not matched perfectly with the shape of the object.
- Generated depth map has a lot of spatial edges.
- Homogeneous region may have several disparity values.
- Abrupt depth changes because of the disparity to depth conversion.

These properties of depth map do not affect the quality of the generated intermediate images, but those can consume a lot of bits for coding due to the high frequency components of depth edges. Figure 1 shows the properties of the depth map. The wall at the background has many different depth values, and the nearest depth difference is 8. In addition, the figure of the chair in the depth map does not fix with the color image, and it has many sharp depth edges around the chair. These properties can consume a lot of bits for depth coding.

(a) disparity map

(b) histogram of disparity map

(c) depth map



(d) histogram of disparity map

Fig. 2. Disparity and depth map ‘Book_arrival’ sequence

The depth estimation software converts disparity *d* to depth value *v* using Eq. (1).

$$v = 255 \cdot \frac{d - d_{\min}}{d_{\max} - d_{\min}} \qquad (1)$$

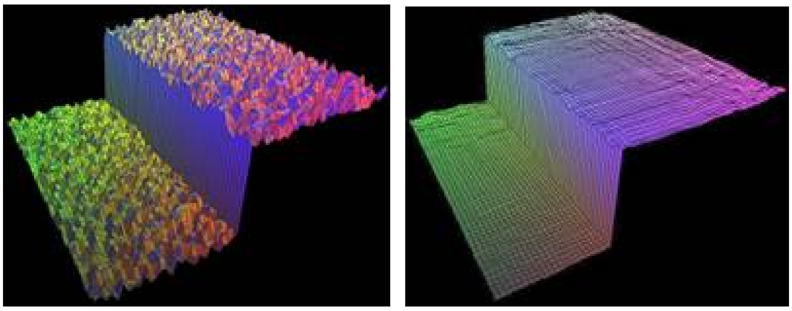
where d_{min} and d_{max} represents the minimum disparity and the maximum disparity of the disparity map. Figure 2 shows the disparity map and depth map respectively. Figure 2(a) is a disparity map of view 7 obtained by quarter-pel precision with range [16, 32] referring to the view 6, 7, and 8. As you can see the histogram of depth map, the values are enlarged by the disparity to depth conversion.

3. Bilateral Filtering on Depth Video

The depth conversion generates sharper edges in the depth map as the disparity range decreases. Likewise, the sharper edges, the more bits are consuming for coding. In order to blur the edges in the depth map, we propose a bilateral filtering as a post-processing in depth estimation. Equation 2 is the bilateral filter, which reduces noises in image in general keeping sharp edges as shown in Fig. 3

$$BF(I_p) = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_s}(\|p - q\|) G_{\sigma_r}(|I_p - I_q|) I_q \qquad (2)$$

where I_p and I_q are the target pixel and neighboring pixel in an image, respectively. W is the size of window, and S is the set of pixels in the window. σ_s is the spatial extent of the kernel, and σ_r is the minimum amplitude of an edge.



(a) input (b) output

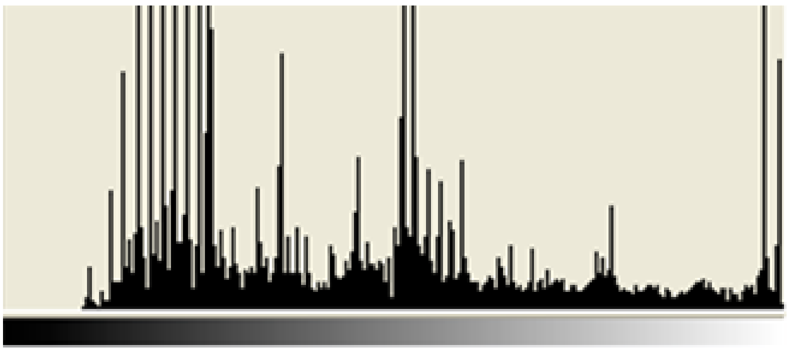
Fig. 3. The input and output images of the bilateral filter

The depth conversion generates sharper edges in the depth map as the disparity range decreases. Likewise, the sharper edges, the more bits are consuming for coding. In order to blur the edges in the depth map, we propose a bilateral filtering as a post-processing in depth estimation. Equation 2 is the bilateral filter, which reduces noises in image keeping sharp edges as shown in Fig. 3.

The important information of the depth map is significant depth changes because it gives the depth impression in general. In other words, gradual depth changes give less depth impression. Therefore, we can do blurring on the gradual depth changes to reduce the bits for coding. Figure 4 shows the resultant depth map blurred by the bilateral filter.

Many new pixel values are generated between the minimum and the maximum depth values. In addition, the significant depth changes are kept.

(a) resultant depth map



(b) histogram of the blurred depth map

Fig. 4. Blurred depth map by bilateral filter

4. Experimental Results

The objective of this proposal is to generate a good depth map for depth coding. We used the bilateral filter for depth smoothing, because it does not affect the quality of the synthesized image and reduces the consuming bits for coding significantly. To achieve this purpose, we compared the objective quality of the synthesized image and bits for depth coding. Four sequences, Book_arrival, Lovebird1, Newspaper, and Pantomime, were used for the experiments. Table 1 shows the test conditions of the depth estimation and the view synthesis.

Table. 1. Test condition of depth estimation and view synthesis

	Parameters	Settings			
		Book_arrival	Lovebird1	Newspaper	Pantomime
Depth Estimation	Reference views	8, 10	6, 8	4, 6	39, 41
	Center view	9	7	5	40
	Disparity range	16 ~ 32	1 ~ 60	18 ~ 40	1 ~ 22
	Precision	Half-pel			
	Temporal	On			
	Segmentation	Off			
	Semi-Automatic	Off			
View Synthesis	Target view	9	7	5	40
	Precision	Half-pel			
	Synthesis Mode	General Mode			
	BNR	Off			
	View Blending	0			

For depth coding, we used the multiview video coding tool JMVC 5.0. We coded the depth video of each view separately. The detailed description of test conditions is represented in the Table 2. Since 3D video considers the MVC standard as a reference codec, we used the JMVC 5.0. Figure 5 describes the data flow of the images. At first, we obtained the depth videos using DERS 4.1, and then we smoothed them with bilateral filter. Next, we coded the resultant depth video using MVC coder. Since the encoder generates the reconstructed file, we used it as input data for the VSRS tool. As a result, we had two kinds of data: synthesized images and bits of coded depth video. Using those data, we compared the performance of the proposed method. For the bilateral filtering, we used the 5x5 window and 30 for both standard deviations

Table 2. Test conditions of depth coding

Parameters	Settings
Standard software	JMVC 5.0
Coding structure	Hierarchical B
GOP size	15
Total number of frames	31
Search range	16
QP	22, 26, 30, 34, 38

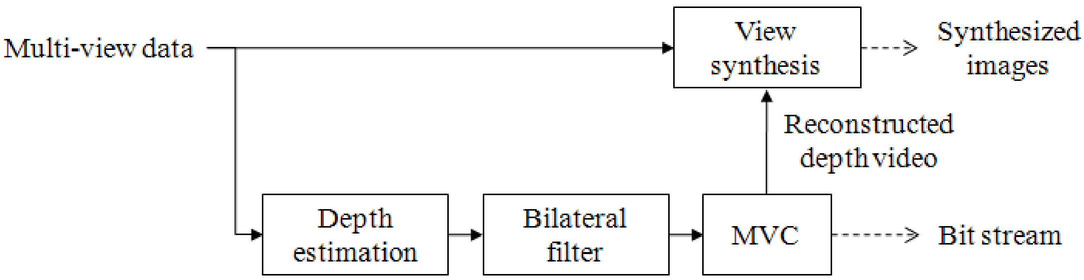


Fig. 5. Block diagram of the proposed method

Four tables from Table 3 to 6 describe the coding result and the objective quality of the synthesized image. The average rate in the table indicates the average bit rates of two input depth videos, and the PSNR value is the objective quality of the synthesized image generated by the reconstructed depth video and the original color video. The delta rate indicates the difference of the bits of the depth video. Similarly, the delta PSNR indicates the difference of two PSNR values between the previous method and the proposed method.

As you can see, the quality degradation of the synthesized image is very small, while the bit saving is considerably high. The maximum rate of bit saving is about 36 % in 'Pantomime' sequence. Each figure from Fig. 6 to 9 shows the comparison of bit rates.

Table 3. Results of 'Book_arrival' sequence

	QP	Previous		Proposed		Analysis	
		Average Rate (kbps)	PSNR of Syn. (dB)	Average Rate (kbps)	PSNR of Syn. (dB)	Delta Rate (kbps)	Delta PSNR (dB)
Book_Arrival	QP22	2,136.41	37.81	1,448.81	37.82	-687.59	0.01
	QP26	1,278.55	37.80	948.02	37.80	-330.52	0.00
	QP30	727.15	37.79	592.96	37.76	-134.19	-0.03
	QP34	402.11	37.74	353.09	37.71	-49.02	-0.02
	QP38	219.57	37.67	205.18	37.66	-14.40	-0.01

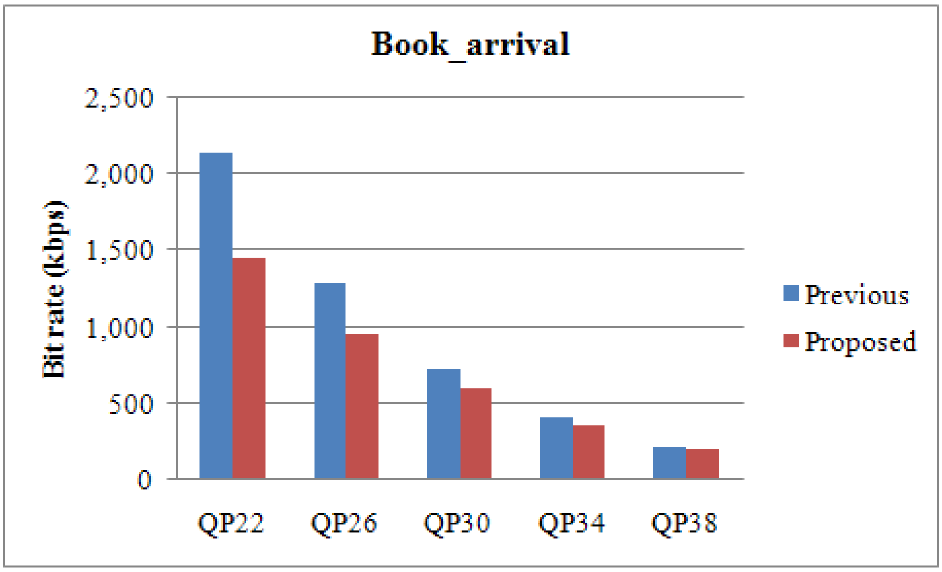


Fig. 6. Bit rates of 'Book_arrival' sequence

Table 4. Results of 'Lovebird1' sequence

	QP	Previous		Proposed		Analysis	
		Average Rate (kbps)	PSNR of Syn. (dB)	Average Rate (kbps)	PSNR of Syn. (dB)	Delta Rate (kbps)	Delta PSNR (dB)
Lovebird1	QP22	226.83	30.32	164.81	30.32	-62.02	0.00
	QP26	125.49	30.37	97.58	30.30	-27.91	-0.07
	QP30	74.95	30.41	62.25	30.31	-12.70	-0.10
	QP34	47.61	30.31	42.48	30.26	-5.13	-0.05
	QP38	33.48	30.28	31.59	30.22	-1.89	-0.07

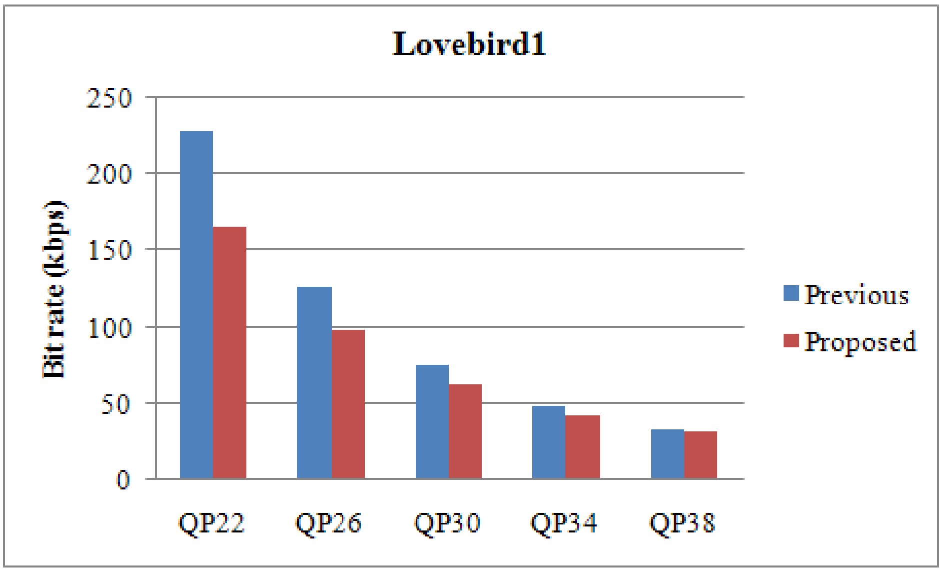


Fig. 7. Bit rates of 'Lovebird1' sequence

Table 5. Results of 'Newspaper' sequence

	QP	Previous		Proposed		Analysis	
		Average Rate (kbps)	PSNR of Syn. (dB)	Average Rate (kbps)	PSNR of Syn. (dB)	Delta Rate (kbps)	Delta PSNR (dB)
Newspaper	QP22	1,051.75	32.73	724.13	32.76	-327.62	0.03
	QP26	629.98	32.75	473.25	32.75	-156.72	0.00
	QP30	374.63	32.75	309.77	32.73	-64.85	-0.02
	QP34	223.12	32.74	198.74	32.71	-24.39	-0.03
	QP38	136.05	32.71	128.62	32.69	-7.42	-0.03

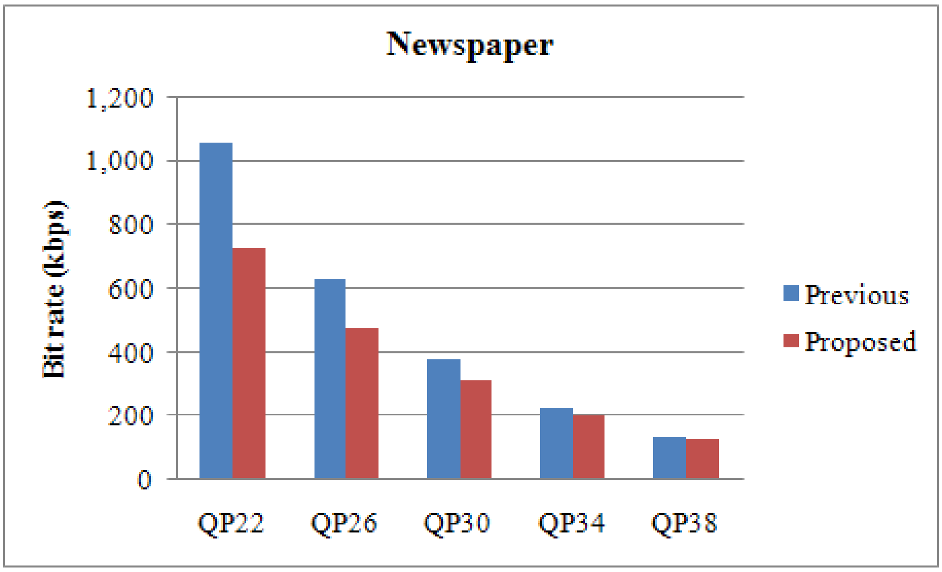


Fig. 8. Bit rates of ‘Newspaper’ sequence

Table 6. Results of ‘Newspaper’ sequence

	QP	Previous		Proposed		Analysis	
		Average Rate (kbps)	PSNR of Syn. (dB)	Average Rate (kbps)	PSNR of Syn. (dB)	Delta Rate (kbps)	Delta PSNR (dB)
Pantomime	QP22	1,058.23	36.83	676.97	36.81	-381.26	-0.02
	QP26	565.84	36.85	410.71	36.82	-155.14	-0.03
	QP30	296.07	36.86	242.66	36.83	-53.42	-0.03
	QP34	154.58	36.87	136.62	36.85	-17.96	-0.02
	QP38	82.06	36.85	76.97	36.83	-5.09	-0.02

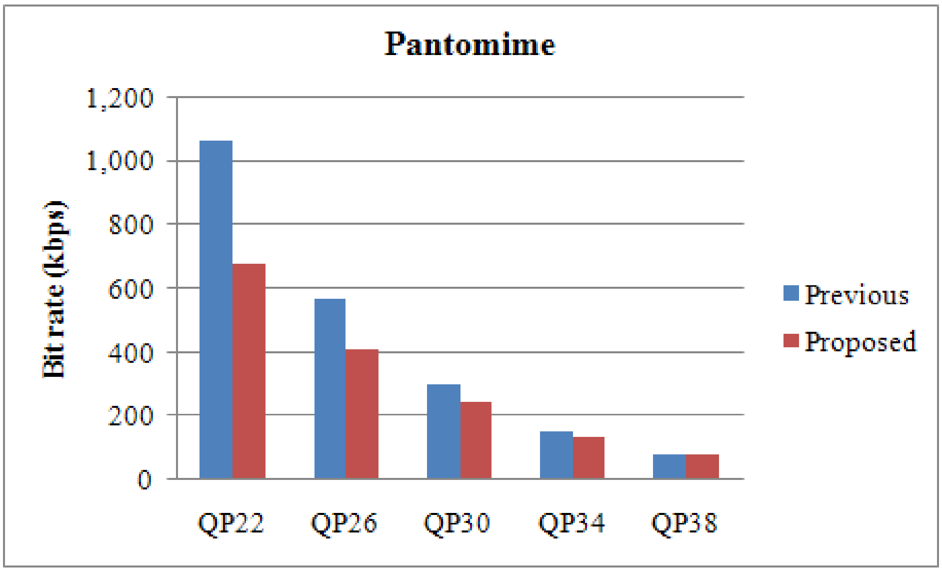


Fig. 9. Bit rates of ‘Pantomime’ sequence

5. Conclusion

Using the bilateral filtering, we obtained blurred depth video which gives significant bit saving in depth coding without quality degradation of the synthesized images. In conclusion, we suggest that this filtering process should be included in the depth estimation software as a post-processing.

6. Acknowledgements

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

7. References

[1] ISO/IEC JTC1/SC29/WG11 “Description of Exploration Experiments in 3D Video Coding”, N10360, February 2009.
[2] ISO/IEC JTC1/SC29/WG11 “Improvement of Depth Map Estimation and View Synthesis”, m15090, January 2008.
[3] Y. Boykov and V. Kologorov, “An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision,” IEEE Trans. on PAMI, September 2004.