RESIDUAL CODING OF DEPTH MAP WITH TRANSFORM SKIPPING

Cheon Lee[1], Hocheon Wey[2], Jaejoon Lee[2], and Yo-Sung Ho[1]

[1]Gwangju Institute of Science and Technology,
South Korea
{leecheon, hoyo}@gist.ac.kr

[2]Samsung Advanced Institute of Technology,
South Korea
{hc.wey, jaejoon1.lee}@samsung.com

ABSTRACT

Since advanced 3D video systems employ depth information to support free-viewpoint navigation and comfortable 3D video viewing, efficient depth map coding is necessary for future 3D video systems. Most residual data in depth map coding are generated along abrupt depth discontinuities, represented by near-zero and high-magnitude values. In this paper, we model the residual data with two representative values calculated by the K-means clustering method and send them to the decoder by skipping transformation. After best mode decision, we applied the proposed method to a block containing residual data, and then we send the quantized representative values to decoder if its coding rate is less than the conventional best mode. By conducting INTRA only coding, -20.32% bit saving was achieved.

Index Terms— 3D video coding, depth map coding, residual coding, K-means clustering

1. INTRODUCTION

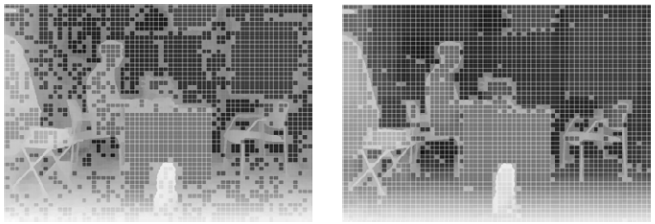
Recently, the demand for 3D contents providing realistic and natural depth impression has grown rapidly [1]. They include not only the stereoscopic video, but also the auto-stereoscopic video for 3D displaying. One popular approach of supporting the contents is employing the multi-view video-plus-depth (MVD) format [2]. Since the depth data describe the camera-object distance for each pixel, displays can generate any intermediate viewpoint images to render comfortable 3D video. However, the huge amount of data due to the increased number of views and their supplementary data, e.g., depth map, is a serious problem. Recently, to resolve this problem, the moving picture experts group (MPEG) have started developing an efficient 3D video coding method to compress the MVD data [3].

There are many approaches to compress the depth map. *Morvan* et al. [4] have used the platelet-based depth map coding with quad-tree decomposition and modeling depth edges. *Jäger* et al. [5] have proposed a depth coding method by signaling the location of depth edges for JPEG2000. *Kang* et al. [6] have used geometrical modeling in intra prediction for depth map coding. Extending to the multi-view video, *Lee* et al. [7] have proposed view synthesis prediction method, which employs the view depth synthesis method using the depth map to generate additional reference frames.

In this paper, we aimed at development of efficient depth map coding utilizing different characteristics of depth data from the color video. The variation of values of a depth map is relatively monotonic compared to color images because depth values of an object are very similar. However, depth values around object boundaries change abruptly; there are barely mixture depth values. Upon this, residual data generated by intra/inter prediction modes may contain a few high-magnitude error values around object boundaries. Consequently, many transform coefficients are generated. This property inspired us to skip the transformation and send two representative values of two groups using K-means clustering which can be useful for depth map coding. In the following chapter, we will describe the details of residual data coding regarding the depth map.

2. ANALYSIS ON DEPTH MAP CODING

Depth maps and color images have different characteristics. Generally, as we mentioned above, depth values around object boundaries change abruptly while non-boundary regions have monotonic depth values. This abrupt depth change consumes high portion of bits in depth map coding. We visualized this property with depth maps coded by QP27 and QP37 in Fig. 1. The brighter regions are coded with Intra4x4 mode and consume many bits due to the transform coefficients. From this result, we observed that the object boundary regions consume lots of coding bits and their residual data have high-magnitude values.



(a) coded by QP27 (b) coded by QP37

Fig. 1. Selected modes for depth map coding

The recent coding standard H.264/AVC uses block-based integer DCT (discrete cosine transform) to remove spatial redundancy. Such a method is very effective for color video coding since the residual data induced by prediction still have high spatial correlation. However, the case is different in depth map coding. Figure 2 is an example that one high-magnitude residual can generate many quantization coefficients in depth map coding; it may consume many bits in entropy coding. Due to the abrupt depth changes, the residual data, i.e., prediction error, can be generated as shown in the figure. Although most pixels are predicted precisely, one high-magnitude prediction error generates many high-magnitude coefficients, consuming many bits. The motivation of this work is that alternative representation method for residual data by skipping transformation can be efficient for depth map coding.

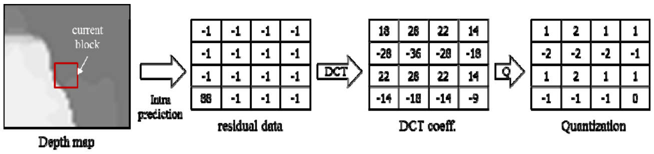


Fig. 2. Effect of transformation on object boundary

3. RESIDUAL CODING FOR DEPTH MAP

The proposed residual coding for depth map uses the K-means clustering instead of the DCT transformation to divide the residual data into two groups. By indicating the meaningful residual values to the decoder, we can reconstruct the block data. Figure 3 describes the overall procedure of the proposed method. First, we obtain the residual data using the best mode by applying the conventional prediction methods. Second, we divide them into two groups using the K-means clustering, and then we obtain two mean values and a grouping map. Third, to achieve lossy coding, we quantize mean values using the proposed quantization method. Consecutively, we calculate the rates for coding by encoding the quantized mean values and the grouping map. Finally, we select the optimal coding method between the conventional best mode and the proposed method for residual coding.

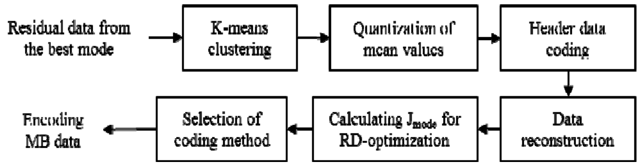
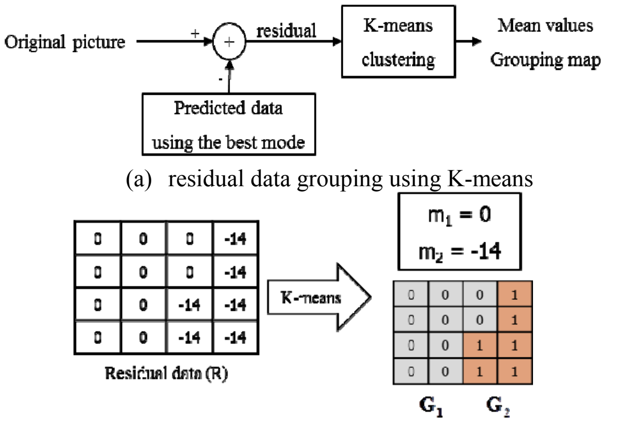

Residual data from the best mode
K-means clustering
Quantization of mean values
Header data coding
Data reconstruction
Calculating J_{mode} for RD-optimization
Selection of coding method
Encoding MB data

Fig. 3. Procedures of residual coding method

3.1. Data grouping using K-means clustering

As we mentioned above, the residual data to be processed are obtained by using the best prediction mode among the conventional modes, i.e., intra/inter prediction methods. Since the objective of the proposed method is to design an alternative coding method to DCT, we only take blocks containing coefficients into account by checking the number of coefficients of a block.

If there is any coefficient in a block, i.e., *coded_block_pattern*≠0, we obtain residual data by subtracting the predicted data of the best mode from the original picture, as shown in Fig. 4(a). Then, we conduct the K-means clustering with K=2. The outputs are two mean values, i.e., m_1 and m_2, and the grouping map. Figure 4(b) shows an example of data grouping. If residual data consists of two representative values, 0 and -14, we obtain two mean values, m_1=0 and m_2=-14. The grouping map consists of binary values. In particular, zeros indicate G_1 group having m_1 while ones indicate G_2 group having m_2. Note that most of m_1 are close to zero since the prediction method of the best mode minimizes distortion.



(a) residual data grouping using K-means

(b) example of data grouping

Fig. 4. Data grouping using K-means clustering

3.2. Quantization of mean values

After applying the proposed method, we need to select which method between the conventional best mode and the proposed method. For this, we compare the consuming rates. We first quantize the resulting mean values of K-means clustering using a midtread quantizer as described in Eq. (1).

$$Z_i = round\left(\frac{m_i}{Q_{step} \times \varepsilon}\right) \quad (1)$$

where $i \in \{0,1\}$ indicates the mean values, and the denominator of Eq. (1) is the quantization step size. We referred to the quantization step sizes of the H.264/AVC to determine Q_{step}, as presented in Table 1. In addition, we use a controller, i.e., ε, of the step size; we use a fixed value at both encoder and decoder. For de-quantization, we use Eq. (2).

$$\tilde{m}_i = Z_i \times Q_{step} \times \varepsilon \quad (2)$$

Table 1. Quantization step sizes

QP	0	1	2	3	4	5	6
Q_{step}	0.625	0.6875	0.8125	0.875	1	1.125	1.25
QP	7	8	9	10	11	12	13
Q_{step}	1.375	1.625	1.75	2	2.25	2.5	2.75
…..							
QP	45	46	47	48	49	50	51
Q_{step}	112	128	144	160	176	208	224

3.3. Entropy coding for header data

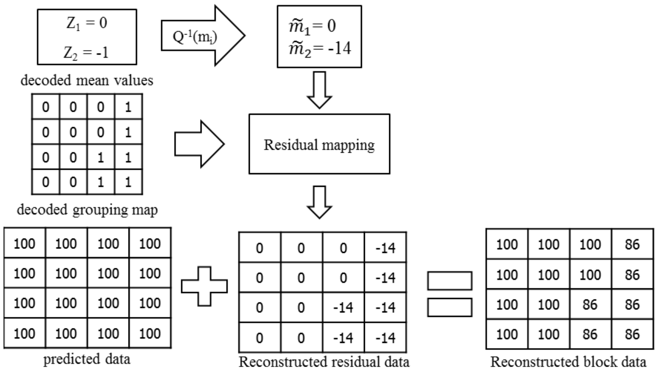
There are three header data in the proposed residual coding method; a flag bit for indicating the proposed method, two quantized mean values, and the grouping map. The flag bit is coded in MB level, hence each MB has one additional flag bit due to the proposed method. The other two data, i.e., quantized mean values and the grouping map are coded in the 4x4 block level. If a block contains transform coefficients, i.e., coded_block_flag =1, we encode these two header data instead of transform coefficients. When we encode the quantized mean values, we use the signed integer Exp-Golomb-code, which is similar to the encoding motion vector difference. After encoding the mean values, we encode the grouping map using the 8-bit fixed-length coding. By selecting 256 frequent grouping maps, we send the index value of a table to the decoder. If there is no matched table, we find the closest table.

3.4. Data reconstruction

The decoding process of the residual data is straightforward. By decoding the quantized mean values and the grouping map, we can obtain the reconstructed residual data using Eq. (3).

$$R_j = (1 - P_j) \times \widetilde{m}_1 + P_j \times \widetilde{m}_2 \qquad (3)$$

where R_j means the residual data of the j-th pixel ($j \in \{0,1,\ldots, 15\}$), and P_j means the j-th binary value indicating the group in a block. Figure 5 shows an example of data reconstruction. Since we can access the predicted data at both the encoder and decoder, we can obtain the reconstructed block data. Since the predicted data can be accessible at the decoder, we can reconstruct a block without any other information.



0	0	0	1
0	0	0	1
0	0	1	1
0	0	1	1

100	100	100	100
100	100	100	100
100	100	100	100
100	100	100	100

0	0	0	-14
0	0	0	-14
0	0	-14	-14
0	0	-14	-14

100	100	100	86
100	100	100	86
100	100	86	86
100	100	86	86

Fig. 5. Data reconstruction for a block

3.5. Mode decision

The proposed depth coding performs as an additional mode at the encoder, hence we need to select which method is efficient for coding. Here is an assumption for mode decision. Considering that the depth map is used for generation of virtual view image, distortion of depth value can be allowed if it does not affect the quality of virtual view image. In this sense, distortion due to the quantization of two mean values can be allowed because the residual data has been obtained by means of the best mode.

In order to decide a mode for a current macroblock, we calculate two cost values using Eq. (4) and (5)

$$J_{\mathrm{conv\_mode}} = R(M, C) \qquad (4)$$

$$J_{\mathrm{Kmeans\_mode}} = R(M, Z, G) \qquad (5)$$

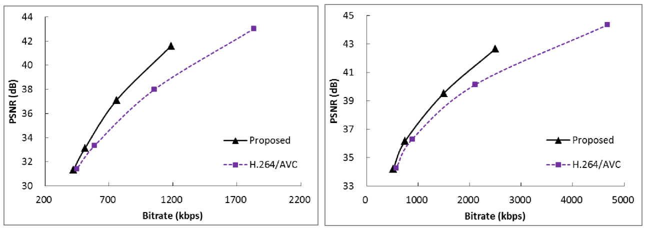
where R represents the actual coding bits for a macroblock. The cost for the conventional mode J_{conv_mode} is the coding bits for mode type M and coefficients C, and that of the proposed mode J_{Kmeans_mode} is the coding bits for mode type M, quantized mean values Z, and the index of grouping map G. By competing two costs, the encoder selects the actual coding method for a macroblock.

4. EXPERIMENTAL RESULTS

In order to evaluate the coding performance of the proposed depth map coding, we used the provided depth data by MPEG 3D video as shown in Fig. 6; we used only depth data for coding experiments [3]. The three sequences have depth data provided by MPEG 3D video coding group.

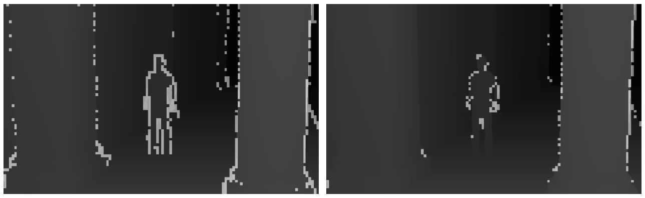
The proposed residual coding is implemented in the reference software JMVC 8.3 (joint multi-view video coding). We applied the proposed coding scheme only to 4x4 blocks. Since even in P- or B- slices use intra prediction or 4x4 block prediction, our proposed method can be applied in any slice type. We tested 100 frames for testing with four QP parameters: 18, 26, 34, and 38. The GOP (group of picture) was 8 with INTRA only coding. After reconstructing the depth video, we generated three intermediate views, where the color videos for virtual view generation were the reconstructed data coded with the same QP set. The quality of the coded depth video was evaluated with the PSNR (peak signal-to-noise ratio) measurement [8].

Figure 6 shows the experimental results of the INTRA only coding. The RD curves of the proposed method are higher than the conventional method H.264/AVC. We observed that the coding gain of high bitrates is greater than that of low bitrates. This can be explained by the fact that the coder at low bitrates loses most of coefficients due to the quantization while our method is applied to blocks containing coefficients. In this sense, we observed that the proposed method is hard to contribute the coding gain in predictive coding such as P- or B-picture coding.



(a) 'Undo_Dancer' (b) 'Newspaper'

Fig. 6. Rate-distortion curves of INTRA coding



(a) QP18 (b) QP34

Fig. 7. Coded regions with proposed method of 'Undo_Dancer'

Figure 7 demonstrates that the brighter regions are the coded macroblocks with the proposed method. Since the proposed method can be applied to depth discontinuities, most selected macroblocks are on the object boundaries. Furthermore, as the QP parameter increases, the selected regions are reduced.

To evaluate the overall coding performance, we employed the Bjontegaard Delta bitrate (BDBR) and PSNR (BDPSNR) measures [9]. As presented in Table 2, the maximum coding gain was the INTRA only case of 'Balloons' sequence,-21.91 % bit-saving or 2.10 dB quality improvement. On average, the INTRA only coding structure showed -20.32 % bit-saving.

Table 2 Coding performance of each sequence

Test data (depth)	INTRA only		Hierarchical B	
	BDBR (%)	BDPSNR (dB)	BDBR (%)	BDPSNR (dB)
Undo_Dancer	-20.46	1.41	-3.84	0.27
Balloons	-21.91	2.10	-1.66	0.07
Newspaper	-18.59	1.86	-5.37	0.22
(average)	-20.32	1.79	-3.62	0.19

5. CONCLUSION

In this paper, we proposed an efficient residual data coding for depth map using the K-means clustering instead of DCT transform. Since the depth values changes abruptly without mixture pixels around object boundaries, the magnitude of residual data is very high; hence it induces a lot of transform coefficients and consumes many bits. To resolve this, we used the K-means clustering to divide the residual data into two groups. Most of the residual data can be divided into zero-mean group and non-zero high magnitude mean group. By signaling these mean values and the grouping information to the decoder, we obtained high coding gains around the object boundaries. By experiments, the INTRA only coding structure showed -20.32 % bit saving on average.

6. ACKNOWLEDGEMENT

This work was supported by Samsung Electronics Co., Ltd.

7. REFERENCES

[1] L. Onural, "Signal Processing and 3DTV," *IEEE Signal Processing Magazine,* vol. 27, pp. 144+141-142, 2010.

[2] 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.

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

[4] Y. Morvan, P. H. N. De With, and D. Farin, "Platelet-based Coding of Depth Maps for the Transmission of Multiview Images," in *Stereoscopic Displays and Virtual Reality Systems XIII*, Jan. 2006.

[5] F. Jäger, "Contour-based Segmentation and Coding for Depth Map Compression," in *IEEE Visual Communications and Image Processing (VCIP)*, 2011.

[6] ISO/IEC JTC1/SC29/WG11, "Description of 3D Video Coding Technology Proposal by Nagoya University," *in MPEG input document* M22567, Nov. 2011.

[7] C. Lee, B. H. Choi, and Y. S. Ho, "Efficient Multiview Depth Video Coding using Depth Synthesis Prediction," *Optical Engineering,* vol. 20, July 2011.

[8] ISO/IEC JTC1/SC29/WG11, "Common Test Conditions for 3DV Experimentation," *in MPEG output document* N12745, May 2012.

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