MODIFIED INTRA PREDICTION METHOD FOR DEPTH VIDEO CODING USING MEDIAN FILTER

Cheon Lee and Yo-Sung Ho

Realistic Broadcasting Research Center (RBRC)
Gwangju Institute of Science and Technology (GIST)
1 Oryong-dong Buk-gu, Gwangju, 500-712, Korea
Tel: +82-62-715-2258, Fax: +82-62-715-2204
E-mail: {leecheon, hoyo}@gist.ac.kr

ABSTRACT

This paper presents a modified intra prediction method for depth video coding using a median filter. The intra prediction method of H.264/AVC utilizes reconstructed neighbor pixels using the 3-tap interpolation filter to predict the current block. Since depth values are monotonic except for depth discontinuity, we can detect coding errors in the reference pixels. Moreover, when the intra prediction modes predict current block using the conventional interpolation filter, unwanted residual data is induced at the depth discontinuity region. In order to improve the prediction accuracy, we determine neighboring reference pixels using the 5-tap median filter. Then, we predict the current block using modified intra prediction modes with the 3-tap median filter. Using this method, we saved the bits as much as 1.6 to 5.8 %.

Index Terms— 3D video coding, depth video coding, intra prediction, median filter

1. INTRODUCTION

Three-dimensional video is captured data using more than two cameras simultaneously to offer a realistic scene [1]. Recently, moving picture experts group (MPEG) has started the standardization activity on 3D video coding (3DVC) since 2007. The MPEG 3DVC group announced a vision on 3D video which supports various 3D displays and its new coding standard will be developed in two years [2]. Related work on the 3D video coding is the multi-view video coding (MVC) which compresses the multi-view video data using inter-view correlation.

In order to provide 3D scene through the multi-dimensional displays, input data of 3D video systems should contain the multi-viewpoint video and its supplementary data for rendering. Depth image is sufficient supplementary data for generating arbitrary viewpoint images. If the depth data and camera parameters are available, we can find the corresponding pixel positions between the reference image and the virtual image. With Mapping the corresponding pixels and filling the hole pixels, we can generate arbitrary viewpoint image. It is called as DIBR (depth-based-image rendering) [3].

The depth image is invisible to user, but it occupies a number of bits in the 3D video data. Therefore, we need to compress them efficiently. One important issue on the depth video coding is coding of the depth discontinuity effectively since those values determines rendering quality. Generally, depth discontinuity indicates the object boundaries and changes abruptly. Because of these characteristics, most of depth discontinuities are coded with the intra prediction. Therefore, if we can modify the intra prediction method for depth video coding, we can improve the coding performance.

The objective of this work is to improve the performance of intra prediction. The conventional intra mode predicts the current block using the reconstructed neighboring pixels, and chooses one best prediction mode [4]. If those reference pixels contain coding errors, prediction errors are induced simultaneously. Another problem in intra prediction is the use of the 3-tap interpolation filter which generates transition values between two distinct values. The generated intermediate value induces a lot of residual data as well.

In this paper, we proposed two steps for the intra prediction using the median filter. We reduced coding errors from the neighboring pixels and predicted the current block using the median filter.

2. INTRA PREDICTION IN H.264/AVC

2.1. Conventional Intra Prediction Method

The latest coding standard H.264/AVC employs intra prediction which refers to the reconstructed neighboring pixels of the current block. In case of ‘Intra 4x4’, for example, 13 pixels are utilized to predict the current block with eight directions except for ‘DC mode’. After predicting, the encoder performs further process such as mode decision, transformation and quantization. Figure 1 describes nine intra prediction modes of ‘Intra 4x4’. Every intra prediction mode utilizes the reconstructed neighboring pixels. ‘A’ to ‘M’ represents the reconstructed pixels. ‘Intra 8x8’ uses similar prediction method with nine prediction modes. On the other hand, ‘Intra 16x16’ mode has only four prediction modes: vertical, horizontal, DC, and plane. After predicting the current block with prediction modes, the encoder calculates the residual data between the original value and the predicted value. Then it is transformed and quantized.

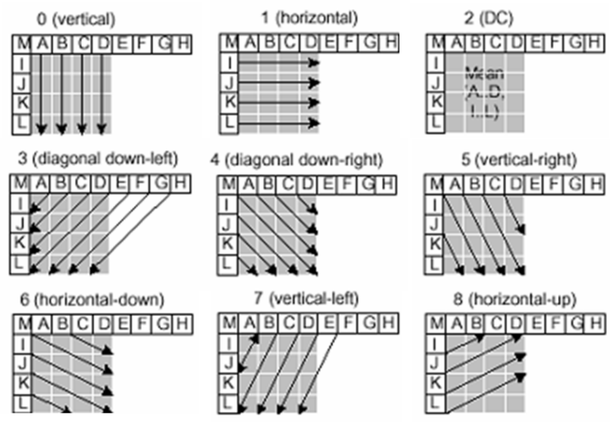


Fig. 1. Intra prediction modes of ‘Intra 4x4’

2.2. Intra Prediction for Depth Video Coding

We considered two problems of intra prediction on depth video coding. The first problem is efficiency of reference pixels. The reference pixels for intra prediction contain coding errors due to the block-based coding method. As the errors are added, the prediction accuracy is become lower.

The second problem is the use of the interpolation filter. Differently from the color image, the pixel values in a depth image changes abruptly around depth discontinuity. Hence, the typical interpolation method such as the 3-tap interpolation filter is not efficient. The conventional intra prediction mode predicts the target block using the 3-tap interpolation filter as described in Eq. (1).

$$P(x_i) = \left\lfloor \frac{x_{i-1} + 2x_i + x_{i+1}}{4} \right\rfloor \quad (1)$$

where the description *i-1* means the left in the above blocks or upper pixels in the left blocks, and *i+1* means the right in the above blocks or lower pixels in the left blocks. This interpolation method for intra prediction is efficient for the color video coding, but it is not for the depth video coding since the depth values change abruptly around the object boundaries. If the depth values of two adjacent objects are much different, the interpolated values consist of a lot of transition values. Consequently, those values can induce a lot of residual data.

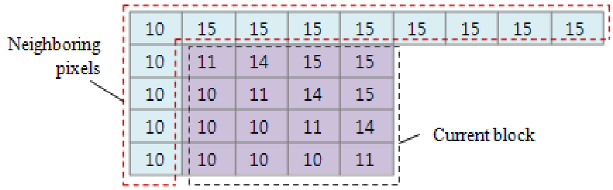


Fig. 2. Predicted pixel values using interpolation filter

Figure 2 shows the resultant predicted values using the conventional interpolation filter. If the neighboring pixels are reconstructed and consist of ‘10’ and ‘15’, the ‘Diagonal-down-right’ mode generates 16 pixel values of the current block. Since the predictor uses the interpolation filter, intermediate pixel values such as ‘11’ and ‘14’ are generated. These intermediate values induce residual data. If we can design a better prediction filter using this characteristic, we can reduce residual data.

3. PROPOSED INTRA PREDICTION

The purpose of this work is to improve performance of depth video coding by modifying the intra prediction method. Figure 3 describes the overall procedure of the proposed method. At first we determine the reference pixels. Then, we predict the current block using modified directional prediction modes.

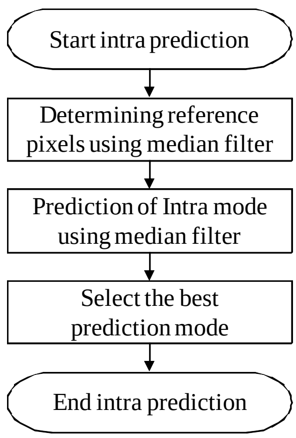


Fig. 3. Proposed intra prediction method

3.1. Determining Reference Pixels

As we mentioned above, reference pixels contain coding errors. High level of coding error can decrease the prediction accuracy in the intra prediction. Therefore, we need to reduce coding errors from the neighboring pixels.

Assuming that most of the reconstructed pixels in the neighboring blocks are close to the original depth value and coding errors are distinguishable in the reference pixels. We regard them as noise and we disregard from the reference pixels. In order to remove the noise from the reference pixels, we use a median filter as Eq. (2).

$$\mathrm{MEDIAN}(x,y,z) = x+y+z - \mathrm{Min}(x,y,z) - \mathrm{Max}(x,y,z) \quad (2)$$

Figure 4 shows an example to determine the reference pixel 'A' using the 3-tap median filter. If the value of 'A' is not a middle value, it is changed with the real median value among three values. In the case of the value of 'H' or 'L', we do not use the median filter. In this manner, all pixels are processed and saved in the buffer. We do not change the reconstructed values with the filtered values. We can also use the 5-tap median filter to extract the noise more efficiently. For the pixels of 'H' and 'L', we use the higher value between them.

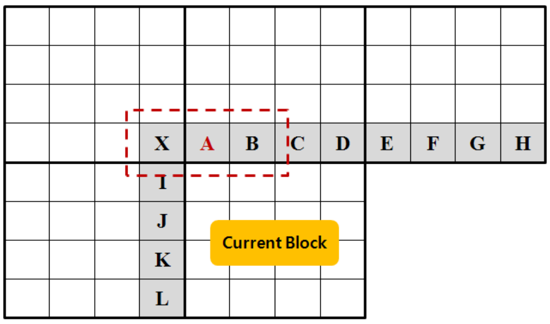


Fig. 4. Noise filtering of neighboring pixels

3.2. Intra prediction using Median Filter

In order to predict the current block, we use the median filter as described in Eq. (3) instead of the interpolation filter. This equation is just a representative formula. For the lack of space, we do not write all prediction formula of each prediction mode.

$$P(x_i) = MEDIAN(x_{i-1}, x_i, x_{i+1}) \quad (3)$$

where the description *i-1* means the left or upper pixels for the current pixel, and *i+1* means the right or lower pixels.

Figure 5 shows an example of the proposed method. If the reconstructed neighboring pixels are the same with that of Fig. 2, the proposed method generates the current pixel values as shown in the figure. As we expected, the predicted pixels consists of only two reference values. We applied these two methods to both 'intra 4x4' and 'intra 8x8' modes except for three modes such as 'vertical', 'horizontal' and 'DC'. These three modes do not use the 3-tap interpolation but use the neighboring values directly.

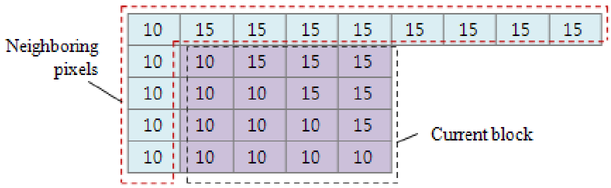
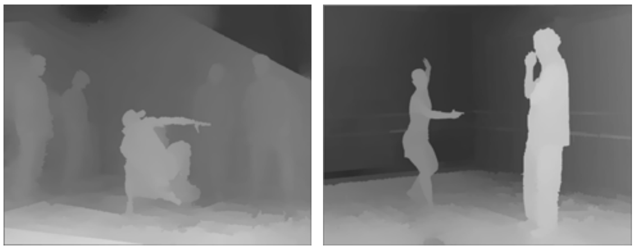


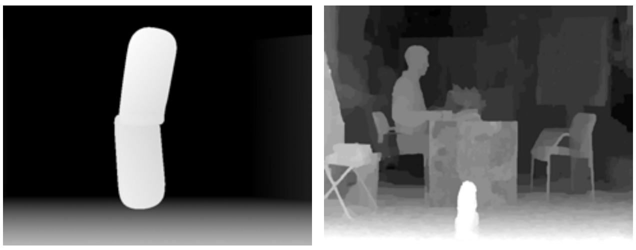
Fig. 5. Predicted values using the median filter

4. EXPERIMENTAL RESULTS

In order to evaluate the proposed method, we used various multi-view sequences having valid depth data or estimated depth video data. 'Breakdancers' and 'Ballet' sequences consist of eight views with both color and depth videos as shown in Fig. 6(a). Those depth sequences are provided by Microsoft Research (MSR) and generated by segment based depth estimation. Other two sequences are provided by MPEG 3DV group. The depth data of 'Mobile' sequence is generated by CG (computer graphics), and the depth video of 'Book Arrival' is estimated by the depth estimation software called DERS (depth estimation reference software) 5.0 provided by MPEG 3DV group [5]. Fig. **6**(b) shows their depth data.



(a) MSR: 'Breakdancers' (left), 'Ballet'(right)



(b) MPEG 3DV: 'Mobile'(left), 'Book_arrival'(right)

Fig. 6. Depth videos for experiments

We implemented the proposed algorithm at the reference software of H.264/AVC tagged as 'Jm14.2'. We tested 100 frames for each sequence with intra only prediction structure because this proposed algorithm is designed for the intra prediction method. Three intra prediction modes are enabled. We used four coding parameters, 17, 22, 27, and 32. The testing QP set is aimed at the high bitrate because the proposed method is effective if the reference pixels values are close to the original values.

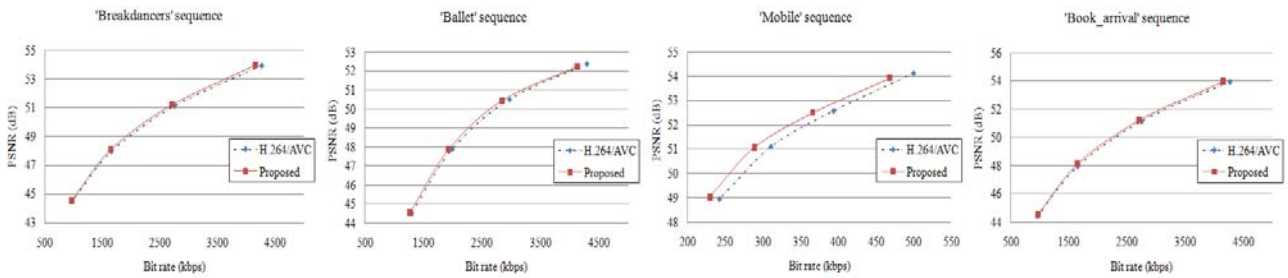


Fig. 7. Experimental results: rate-distortion curves

Table 1. Coding results of depth video coding

Test Data	PSNR (dB)								Bit rate (kbps)							
	QP17		QP22		QP27		QP32		QP17		QP22		QP27		QP32	
	H.264	Prop.	H.264	Prop.	H.264	Prop.	H.264	Prop.	H.264	Prop.	H.264	Prop.	H.264	Prop.	H.264	Prop.
Break	54.0	54.0	51.1	51.2	48.0	48.1	44.5	44.5	4267.6	4159.3	2751.1	2710.0	1653.5	1642.6	961.6	964.7
Ballet	52.4	52.2	50.5	50.4	47.9	47.9	44.5	44.6	4295.0	4133.0	2971.5	2849.0	1993.9	1935.8	1299.1	1271.2
Mobile	54.1	53.9	52.6	52.5	51.1	51.8	49.0	49.0	500.2	468.9	394.0	366.3	311.0	288.6	242.5	229.3
Book	54.1	54.2	49.9	49.9	46.6	46.6	43.7	43.7	4755.3	4674.2	3150.1	3101.3	1862.7	1843.4	1095.4	1087.6

Table 1 presents coding results of four test sequences. While PSNR values of reconstructed images are almost the same, overall bitrates are reduced considerably. ‘Mobile’ sequence showed the best coding performance. It is because that the foreground object is far from the background. In other words, Since the depth discontinuity is obvious, coding errors in the reference pixels are removed efficiently and the predicted values are accurate. In the same manner, ‘Ballet’ sequence showed good coding performance.

Table 2. Improvements of coding performance

Test Data	BDPSNR (dB)	BDBR (%)
Breakdancers	0.159	-2.45
Ballet	0.176	-2.57
Mobile	0.422	-5.80
Book_arrival	0.116	-1.59

Table 2 contains the values of BDBR (Bjonteggaard Delta BitRate) and BDPSNR (Bjonteggaard Delta PSNR) to evaluate the coding performance [6]. The PSNR values of reconstructed images are improved by 0.1 to 0.4dB, or the bit rates are reduced as much as 1.6 to 5.8 %. The coding performance is efficient at high bit rate.

5. CONCLUSION

In this paper, we proposed an efficient intra prediction method for depth video coding using the median filter. Since depth video is relatively simple compared to color video, depth discontinuity region is coded by intra prediction modes in general. In order to improve coding performance, we selected the reference pixels using the median filter. Then, we predicted the pixels of the current block using the same filter according to prediction directions. By simulating this proposed method, we confirmed that the bitrate of depth video coding has been reduced as much as 1.6 to 5.8 %.

6. ACKNOWLEDGEMENT

This research was supported in part by MKE under the ITRC support program supervised by the NIPA (NIPA-2010-(C1090-1011-0003)).

7. REFERENCES

[1] A. Smolic, K. Müller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards," *IEEE International Conference on Multimedia & Expo*, July 2006.

[2] ISO/IEC JTC1/SC29/WG11, "Vision on 3D Video," N10357, Feb. 2009.

[3] C. Fehn, "Depth-Image-Based Rendering (DIBR), Compression and Transmission for a New Approach on 3D-TV, " *Proc. SPIE Stereoscopic Displays and Virtual Reality System XI*, Jan. 2004.

[4] ITU-T Rec. H.264 / ISO/IEC 11496-10, "Advanced Video Coding," Final Committee Draft, Doc. JVT-F100, Dec. 2002.

[5] ISO/IEC JTC1/SC29/WG11 "Reference Software of Depth Estimation and View Synthesis for FTV/3DV," M15836, Oct. 2008.

[6] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," ITU-T Q.6/16, VCEG-M33, March 2001.