

Fast Mode Decision Algorithm for Depth Coding in 3D Video Systems Using H.264/AVC

Da-Hyun Yoon and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Korea
{yoon,hoyo}@gist.ac.kr

Abstract. Complexity of multiview coding is proportional to the number of cameras. It makes difficult to implement multiview sequences in real applications. Thus, we propose a fast mode decision algorithm for both intra and inter prediction to reduce the computational complexity of H.264/AVC for depth video coding. By analyzing the depth variation, we classify the depth video into depth-continuity and depth-discontinuity regions. We determine a threshold value for classifying these regions by experiments. Since the depth-continuity region has an imbalance in the mode distribution, we limit the mode candidates. Experimental results show that our proposed algorithm reduces the encoding time by up to 78% and 84% for the intra and inter frames, respectively, with negligible PSNR loss and slight bit-rate increase, compared to JMVC 8.3.

Keywords: depth video coding, macroblock mode decision, depth compression.

1 Introduction

The development of 3DTV has realized the human dream of viewing scenes as if in the real world owing to advances in three-dimensional (3D) display technologies. Via 3DTV, the interactive selection of viewpoint and direction becomes possible within a certain operational range; this is referred to as free viewpoint TV (FTV). FTV has been widely used because it transmits and records all spatiotemporal information from the real world [1]. Multiview plus depth (MVD) is another framework used to represent 3D scenes; MVD has been used to synthesize intermediate views from captured images and depth maps. In a recent Moving Picture Experts Group (MPEG) meeting, MVD has received increased attention and has been discussed as a next-generation FTV format [2]. However, since the amount of data and complexity of MVD is proportional to the number of cameras, an efficient encoding method for 3D video must be developed. In order to utilize limited bandwidth and storage capacity efficiently, depth map sequences also have to be compressed in the same manner as texture videos.

In order to select the optimal coding mode, we use a rate-distortion optimization (RDO) technique [3] in JMVC, where J is the RD cost of the current mode, D denotes the distortion between the original and reconstructed macroblock (MB), R stands for the total bits of the MB header, motion vectors, and DCT coefficients, and λ is the Lagrange multiplier. The rate-distortion (RD) cost is calculated as follow

$$J=D+\lambda \cdot R \quad (1)$$

JMVC 8.3[4] is then used to select the best mode among 14 different macroblock modes: SKIP, Direct, Inter16×16, Inter16×8, Inter 8×16, Inter 8×8, Inter 8×8 Frext, Inter 8×4, Inter 4×8, Inter 4×4, Intra16×16, Intra 8×8, Intra4×4, and Intra PCM. The full search algorithm uses all modes to determine the optimal macroblock mode in terms of the RD cost, and the mode having the minimum RD cost is subsequently selected as the best mode. Unfortunately, the full search algorithm is time consuming, making it difficult to implement MVD in real applications. In this paper, we propose a fast mode decision scheme to reduce the complexity.

Depth sequences represent the distance between an object and a camera as a gray scale image. The image has a continuous area at an object and background and a discontinuous area at the boundaries between an object and background. Because the characteristics of depth sequences are very different from those of texture sequences, efficient coding algorithms specializing in depth sequences have been introduced. For example, Oh [5] re-uses motion information of the corresponding texture sequences in order to reduce the complexity in a motion estimation process and the bitrate for motion vector coding in the depth sequences. The fast mode decision method presented by Shin [6] uses a region analysis to reduce the complexity. Then, based on the RD cost correlation between neighboring views and the RD cost of different texture segmentation regions, a pre-decision of the SKIP mode was introduced by Whu [7]. Lee [8] skipped some blocks in the depth image to reduce the depth coding bitrate according to introducing an inter-view correlation of the texture image.

Approaches for fast depth video coding can be classified into two groups. One group exploits the correlation between color and depth video, and shares common information to reduce redundancy. The other group focuses only on the unique properties of depth video. Most fast depth coding algorithms belong to the first group. However, the second group is generally more available to various 3DTV applications because their algorithms only use unique properties of the depth video itself and is independent of the experimental framework.

The proposed algorithm consists of two parts: (1) the SKIP and intra modes are searched in the depth-continuity regions, whereas all modes are searched in the depth-discontinuity regions in the same manner as in JMVC; and (2) vertical, horizontal, DC, and diagonal down-left modes are implemented in the Intra 4×4 prediction of the depth-continuity regions. A threshold value is used to determine whether the region is homogeneous or not, and it is calculated adaptively according to the quantization parameter (QP).

2 Proposed Algorithm

2.1 Edge Classification

Depth sequences have different properties from those of texture sequences. Since the Intra 16×16, Intra 4×4, and SKIP modes are frequently selected as the best modes, motion vector search and the mode decision process can be skipped in homogeneous

regions. Before applying our proposed algorithm, we first separate a macroblock into continuous and discontinuous regions. For this task, the degree of variation of the depth value in a macroblock is defined by

$$f(x,y) = \frac{1}{16^2} \sum_{i=1}^{16} \sum_{j=1}^{16} (r(i,j) - m_{x,y})^2 \quad (2)$$

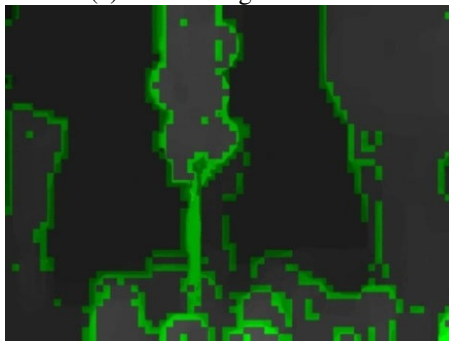
where coordinate (x,y) is the position of the current macroblock, $r(i,j)$ is the depth value at a relative coordinate (i,j) in the current macroblock, and $m_{x,y}$ is the mean value of the current macroblock. If the depth values drastically change, the value of $f(x,y)$ is large; thus, the value of $f(x,y)$ is large in discontinuity regions. Conversely, in continuity regions, where the depth values are almost fixed, the value of $f(x,y)$ is small. Therefore, we use a threshold value (T) to determine whether the current macroblock is located at a boundary or not. In Fig.1, macroblocks shown in green represent depth-discontinuity regions with $f(x,y) > T$. Compared to color sequence, depth-discontinuity that is filled with green macroblock regions are adequately detected. So, using variation is easy way to separate depth-continuity regions.



(a) color image of "Balloon"



(b) color image of "Book Arrival"



(c) depth image of "Balloon"



(d) depth image of "Book Arrival"

Fig. 1. Segmentation of discontinuity regions ($T = 30$)

2.2 Analysis of Mode Selection

The macroblock mode distribution of the depth-continuity regions is different from that in depth-discontinuity regions. Figure 1 compares the mode distribution between the depth-continuity and depth-discontinuity regions. In the depth-continuity regions, there is a severe imbalance in the mode distribution as most macroblocks are encoded by the SKIP and intra modes. However, in the depth-discontinuity regions the mode is balanced; therefore, we use this property to design a fast mode decision algorithm. Table 1 presents the encoding configurations for Fig. 2 and Fig. 3.

Table 1. Encoding configurations

Parameter	Setting
<i>Sequence</i>	Balloon
<i>Threshold</i>	30
<i>QP</i>	32
<i>View</i>	1
<i>Time</i>	0

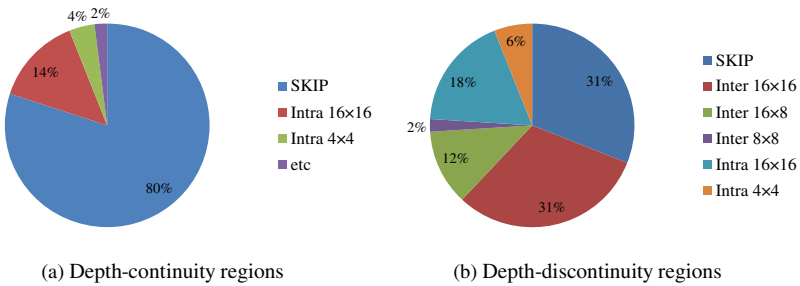


Fig. 2. Macroblock mode distribution

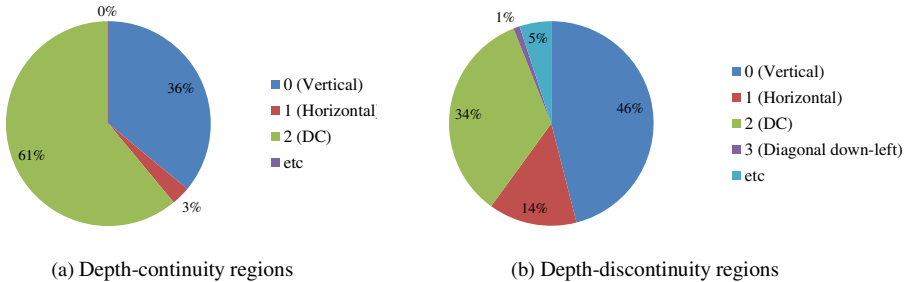


Fig. 3. Intra prediction mode distributions

Since mode 0, mode 1, mode 2, and mode 3 comprise close to 100% of the depth-continuity regions, as shown in Fig. 3, we consider these modes among the nine prediction modes selected as candidate modes set in the Intra 4x4 prediction. However, since the mode distributions for all four modes in the Intra 16x16 prediction are similar,

we do not need to consider the candidate mode set in the Intra 16×16 prediction. Therefore, we calculate the SKIP mode, the candidate mode set in the Intra 4×4 prediction, and all four modes in the Intra 16×16 prediction in order to determine the best mode in the depth-continuity regions. Since the distribution for all modes varies in the depth-discontinuity regions, we use a conventional mode decision method in JMVC.

2.3 Fast Depth Coding Algorithm

The algorithm developed by Peng [9] empirically indicates that the threshold value that discriminates between the depth-continuity and depth-discontinuity should be set to 30. If $f(x,y)$ is less than or equal to the threshold value, a macroblock is determined to be a continuous region. Since the SKIP and intra modes are used as candidate modes in continuous regions, we can reduce the complexity by skipping the inter mode in the mode decision. If $f(x,y)$ is larger than a threshold value, a macroblock is determined to be a discontinuous region, and all macroblock mode decisions including the inter mode decision are implemented. However, in low QPs, Peng's algorithm does not keep detailed boundary regions because it uses a fixed threshold value. To solve this problem, we use adaptive threshold values for the intra and inter predictions.

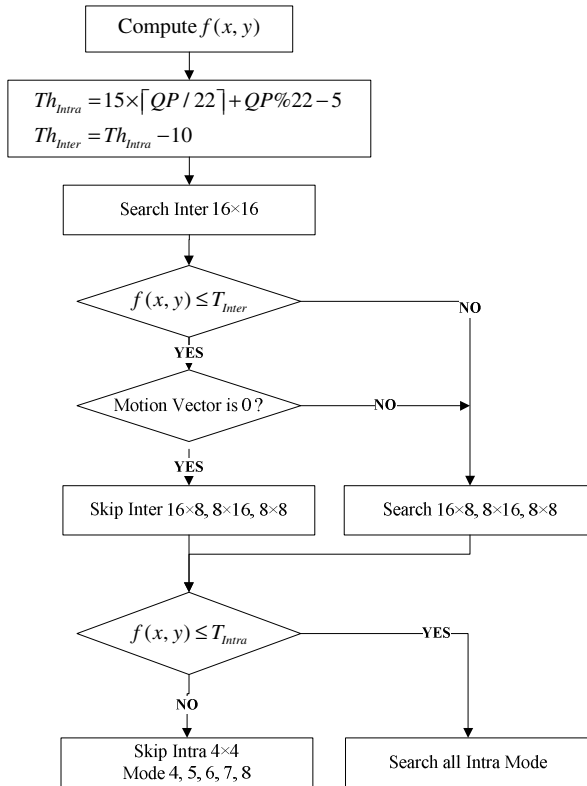


Fig. 4. Flowchart of the proposed algorithm

Figure 4 presents a flowchart of the proposed algorithm. After $f(x,y)$ is calculated at each macroblock, the threshold values for the intra and inter predictions are calculated according to the QP. In order to guarantee the coding efficiency at all QP ranges, we assign small and large threshold values in the low and high QP values, respectively. Since the SKIP of the inter mode decision induces additional bits from the wrong determination, the threshold value for the inter mode (T_{Inter}) is lower than that for the intra mode (T_{Intra}). After SKIP and inter 16×16 are performed, If $f(x,y) < T_{Inter}$ and the motion vector of inter 16×16 equals to 0, calculating inter 16×8, inter 8×16, and inter 8×8 to find the best mode of the current macroblock is skipped. Otherwise, we calculate inter 16×8, inter 8×16, inter 8×8. Next, if $f(x,y) < T_{Intra}$, mode decisions for mode 0, mode 1, mode 2, and mode 3 in the Intra 4×4 are performed. Otherwise, all modes are calculated in order to find the best intra prediction mode.

3 Experimental Results

In order to evaluate the efficiency of the proposed algorithm, we performed experiments on several depth sequences having 1024×768 resolutions. All test sequences have 50 frames, and we implemented our proposed algorithm on reference software JMVC 8.3. The detailed encoding parameters for the reference software are summarized in Table 2.

Table 2. Experimental conditions

Parameter	Setting
<i>Reference software</i>	JMVC 8.3
<i>Profile</i>	FRExt
<i>Depth sequences</i>	Book Arrival, Love Bird, and Newspaper
<i>Resolution</i>	1024×768
<i>Number of encoded frames</i>	50
<i>Search range</i>	±32
<i>QP</i>	22, 27, 32, 37
<i>Symbol mode</i>	1 (CABAC)
<i>GOP size</i>	Intra frame : 1 Inter frame : 8
<i>Intra period</i>	Intra frame : 1 Inter frame : 16

Table 3 shows that there are bitrate increases within ±0.04% and PSNR reductions up to -0.13 dB. Compared to the full search algorithm, up to 78.73% of the encoding time is saved; sufficiently compensating for the significant bitrate increase and decreased PSNR. Because the “Love Bird” depth sequence has many continuous background regions, it shows better performance compared to the other sequences.

Table 4 and Table 5 compare Peng’s algorithm and the proposed algorithm. There is a negligible bitrate increment and PSNR reduction of up to 1.57% and 0.26 dB in the proposed algorithm. The average PSNR difference is decreased by 0.12 dB and

Table 3. Comparison of intra coding performance in terms of Δ PSNR, Δ BR, and Δ TS

Sequences	QP	Δ PSNR (dB)	Δ BR (%)	Δ TS (%)	BDBR (%)	BDPSNR (dB)
Book Arrival	22	-0.11	-0.04	-8.41	0.00	-0.01
	27	-0.13	0.00	-11.06		
	32	-0.04	0.00	-11.47		
	37	0.01	0.00	-14.28		
Love Bird	22	-0.05	0.00	-78.73	-0.05	0.003
	27	-0.06	0.00	-70.03		
	32	-0.05	0.00	-74.46		
	37	-0.00	0.00	-76.43		
Newspaper	22	-0.03	-0.01	-27.43	0.12	-0.007
	27	-0.09	-0.01	-32.50		
	32	0.03	0.00	-29.54		
	37	-0.01	0.00	-29.58		
Average		-0.05	-0.01	-38.66	0.02	0.00

Table 4. Inter coding performance in terms of Δ PSNR, Δ BR, and Δ TS for Peng's algorithm

Sequences	QP	Δ PSNR	Δ BR	Δ TS	BDBR	BDPSNR
Book Arrival	22	-0.32	-1.42	-48.40	5.23	-0.174
	27	-0.41	-3.96	-52.64		
	32	-0.30	-3.98	-52.42		
	37	-0.26	-0.85	-40.37		
Love Bird	22	-0.28	-2.08	-92.52	2.5	-0.092
	27	-0.20	-1.39	-85.49		
	32	-0.16	-2.21	-85.32		
	37	-0.14	-0.65	-85.80		
Dog	22	-0.76	-6.03	-87.52	8.54	-0.298
	27	-0.80	-10.81	-86.68		
	32	-0.51	-9.28	-85.76		
	37	-0.29	-2.40	-84.44		
Kendo	22	-0.61	2.25	-83.99	17.75	-0.95
	27	-0.72	2.02	-84.45		
	32	-0.76	3.73	-84.75		
	37	-1.29	1.26	-84.14		
Pantomime	22	-0.59	5.80	-94.10	42.31	-1.665
	27	-1.09	8.21	-94.22		
	32	-0.94	22.04	-94.23		
	37	-3.11	4.23	-94.19		
Average		-0.22	-0.68	-83.79	15.27	-0.65

0.22 dB in the proposed algorithm and Peng's algorithm, respectively; the bitrates are reduced by an average of 0.37% and 0.68% in the two algorithms. In order to measure the overall improvement of the proposed method, the Bjontegaard delta bitrate (BDBR) and PSNR (BDPSNR) were used. BDBR and BDPSNR mean bitrate and PSNR under equivalent PSNR and bitrate, respectively.

Table 5. Inter coding performance for the proposed algorithm

Sequences	QP	Δ PSNR	Δ BR	Δ TS	BDBR	BDPSNR
Book Arrival	22	-0.22	1.57	-76.30	4.08	-0.156
	27	-0.20	0.51	-79.01		
	32	-0.12	-0.65	-68.95		
	37	-0.09	-0.77	-66.61		
Love Bird	22	-0.13	0.40	-81.57	0.53	-0.023
	27	-0.05	-0.79	-63.95		
	32	-0.01	-0.64	-62.07		
	37	-0.01	0.09	-66.33		
Dog	22	-0.17	-0.13	-84.0	2.26	-0.083
	27	-0.12	-0.49	-83.35		
	32	-0.05	0.10	-81.35		
	37	-0.04	-0.17	-82.65		
Kendo	22	-0.26	0.55	-79.77	2.6	-0.174
	27	-0.19	0.16	-78.90		
	32	-0.14	-0.80	-76.72		
	37	-0.07	-0.81	-74.89		
Pantomime	22	-0.19	-1.18	-79.61	1.78	-0.062
	27	-0.14	-2.19	-78.66		
	32	-0.09	-1.41	-76.68		
	37	-0.03	-0.81	-74.38		
Average		-0.12	-0.37	-75.79	2.25	-0.10

$$\Delta PSNR = PSNR_{proposed} - PSNR_{original} \quad (3)$$

$$\Delta TS = \frac{Time_{proposed} - Time_{original}}{Time_{original}} \times 100 \quad (4)$$

$$\Delta BR = \frac{Bitrate_{proposed} - Bitrate_{original}}{Bitrate_{original}} \times 100 \quad (5)$$

In terms of BDBR and BDPSNR, the proposed algorithm generates low bitrate and PSNR degradation compared to Peng's algorithm, which used a fixed threshold value of 30. Since low QP sequences are more sophisticated than high QP sequences, the inaccurate prediction process induced from skipping modes caused the quality degradation. Because Peng's algorithm skips many modes regardless of the macroblock properties, a bitrate increase and PSNR loss occur. In terms of BDBR and BDPSNR, Peng's algorithm represents BDBR increase and BDPSNR loss up to 42% and -1.6

dB respectively. This results show that Peng's algorithm is inefficient. However, experimental results show that the proposed algorithm is improved about 2.25% and -0.10 dB in terms of BDBR and BDPSNR respectively. Thus, experimental results in terms of BDBR and BDPSNR mean that the proposed algorithm maintains the quality of depth sequence coding and bitrate value compared to the Peng's algorithm. The proposed method reduces the encoding time an average of 75.79%. Figure 5 shows that BDBR and BDPSNR values are sensitive to T_{inter} and that the smaller the T_{inter} the smaller the bit increase and PSNR degradation.

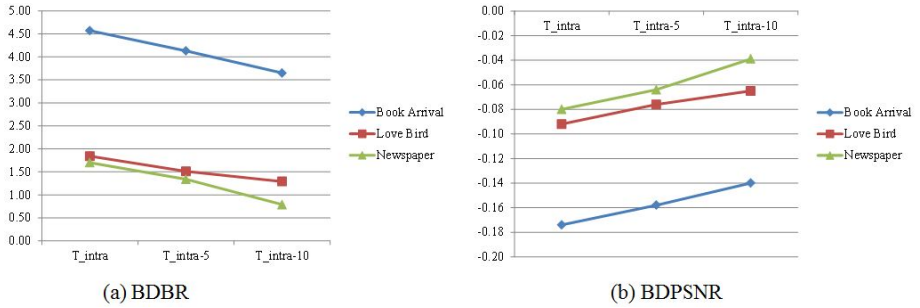


Fig. 5. BDBR and BDPSNR change according to T_{inter} value

To synthesize the intermediate virtual view, we used the 4th and 6th views in Newspaper and “Love Bird”, and the 7th and 9th views in “Book Arrival”

The results of PSNR differences between the synthesized view and original sequences compared to full search algorithm are shown in Table 7; VSRS rendering software was used for this experiment [10]. Both Peng's and the proposed algorithms have negligible differences in the range from -0.02 dB to 0.06 dB, indicating that both algorithms maintain rendering quality. Figure 6 shows that there is no significant rendering quality degradation between the full search and the proposed algorithm.



(a) Full search algorithm

(b) Proposed algorithm

Fig. 6. Synthesized results of “Book Arrival sequence”

Table 6. Δ PSNR of the synthesized result

	QP	22	27	32	37
Peng's algorithm	Book Arrival	0.01	0.01	0.06	-0.01
	Love Bird	0.01	0.00	0.00	-0.01
	Newspaper	0.00	0.00	0.00	0.00
Proposed algorithm	Book Arrival	0.02	0.02	0.06	-0.02
	Love Bird	0.01	0.01	-0.01	0.00
	Newspaper	0.00	-0.01	0.00	0.00

4 Conclusions

Although MVC coding is time consuming, it is still the most effective way to represent 3D scenes. The property of depth images is different from that of texture images. Since depth-discontinuity regions have imbalanced macroblock mode distributions, we propose a fast depth video coding algorithm using a threshold value that determines the depth-continuity or depth-discontinuity adaptive to the QP. If the variation is lower than the inter threshold value and motion vector is equal to zero, the proposed algorithm uses the SKIP and intra modes with no motion estimation or compensation; if the variation is smaller than the intra threshold value, the proposed algorithm performs the intra prediction using only mode 0, mode 1, mode 2, and mode 3. Our algorithm reduce the encoding time up to 78% and 84% for the intra and inter frames, respectively, with no significant degradation for the PSNR and rendering quality, or bitrate increment.

Acknowledgments. This research was supported by the MKE(Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency). (NIPA-2011-(C1090-1111-0003)).

References

1. Benzie, P., Watson, J., Surman, P., Rakkolainen, I., Hopf, K., Urey, H., Sainov, V., Kopylow, C.V.: A Survey of 3DTV displays: techniques and technologies. *IEEE Transactions on Circuits and Systems for Video Technology* 17(7), 1647–1658 (2007)
2. Mori, Y., Fukushima, N., Fujii, T., Tanimoto, M.: View generation with 3D warping using depth information form FTV. In: 3DTV Conference, pp. 229–232 (2008)
3. Sullivan, G.J., Wiegand, T.: Rate-distortion optimization for video compression. *IEEE Signal Process* 15, 74–90 (1998)
4. Joint multiview coding (JMVC) 8.3, <http://garcon.ient.rwth-aachen.de>
5. Oh, H., Ho, Y.-S.: H.264-Based Depth Map Sequence Coding Using Motion Information of Corresponding Texture Video. In: Chang, L.-W., Lie, W.-N. (eds.) *PSIVT 2006. LNCS*, vol. 4319, pp. 898–907. Springer, Heidelberg (2006)

6. Shin, K., Chun, S., Chung, K.: A fast mode prediction of multi-view video coding using region analysis. *Digital Content, Multimedia Technology and its Applications*, 87–90 (2010)
7. Whu, W., Jiang, W., Chen, Y.: A fast inter mode decision for multiview video coding. *Consumer Electronics*, 689–694 (2011)
8. Lee, J., Wey, H., Park, D.: A fast efficient multiview depth image coding method based on temporal and inter-view correlations of texture images. *IEEE Transactions on Circuits and Systems for Video Technology* 99, 1–4 (2011)
9. Peng, Z., Yu, M., Jiang, G., Si, Y., Chen, F.: Virtual view synthesis oriented fast depth video encoding algorithm. In: *International Conference on Industrial and Information Systems*, pp. 204–207 (2010)
10. ISO/IEC JTC1/SC29/WG11 M16090: View synthesis algorithm in view synthesis reference software 2.0 (VSRS 2.0) (2009)