Fast Depth Video Coding Method using Adaptive Edge Classification

Da-Hyun Yoon and Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Republic of Korea
Email: {yoon, hoyo}@gist.ac.kr

Abstract—**In this paper, we propose a fast mode decision algorithm for both the intra prediction and inter prediction in the depth video sequence. The proposed algorithm reduces the complexity of the depth video coding. According to the depth variation, depth video can be classified into depth-continuity and depth-discontinuity regions. From experiments, we determine a threshold value for classifying these regions. Since the depth-continuity region has an imbalance in the mode distribution, we limit the mode candidates to reduce the complexity of the mode decision process. Experimental results show that our proposed algorithm reduces encoding time up to 31% and 98% for the intra and inter frames, respectively, compared to the H.264/AVC standard with negligible PSNR loss and bit rate increase.**

I. Introduction

Due to the advance in three-dimensional (3D) display technologies, 3DTV realized the human dream of seeing the scene as the real world. Even if we changed viewpoint, it makes possible us to interactive selection of viewpoint and direction within a certain operation range. It is called as free viewpoint TV (FTV). FTV has been widely utilized because it transmits and records all spatiotemporal information from the real world [1]. Multiview plus depth (MVD) is another framework used to represent 3D scene and it was 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 is considered as a next-generation FTV format [2]. Since an amount of data and complexity of MVD are proportional to the number of cameras, we have to encode 3D video scenes efficiently to transmit and store them for practical use. To reduce the complexity of depth video sequences, a research reuses motion information of the corresponding texture sequences in order to reduce the complexity [3].

H.264/AVC selects the best macroblock mode among eight different macroblock modes: SKIP, Inter16×16, Inter16×8, Inter8×16, P8×8, Intra16×16, Intra8×8, and Intra4×4. The full search algorithm uses all modes to determine the best macroblock mode in terms of the rate distortion (RD) cost. A mode having the minimum RDcost is determined as the best mode. However, unfortunately, the full search algorithm is time consuming. It makes hard to put MVD into applications. In this paper, we proposed a fast mode decision scheme to reduce the complexity.

The proposed algorithm consists of two parts: 1) the SKIP and intra modes are considered only in the depth-continuity regions, whereas all modes are searched in the depth-discontinuity regions in the same manner as H.264/AVC. 2) Vertical, horizontal, DC, and diagonal down-left modes are implemented in the Intra4 × 4 prediction of the depth-continuity regions. A threshold value is determined whether the region is homogeneous or not, and is calculated adaptively according to the quantization parameter (QP). The experimental results show that the proposed algorithm reduces the computational complexity without significant degradations of bit rate, PSNR, and rendering quality.

II. Mode Decision in H.264/AVC

H.264/AVC supports five inter modes and three intra modes. The inter modes are composed of SKIP, Inter16×16, Inter16×8, Inter8×16, and P8×8 modes [4]. The P8×8 mode is divided into one of 8×8, 8×4, 4×8, and 4×4 modes. The intra modes consist of Intra4×4, Intra8×8, and Intra16×16 modes. The Intra8×8 is supported in the FRExt profile only. The intra prediction used reconstructed pixels of the adjacent blocks.

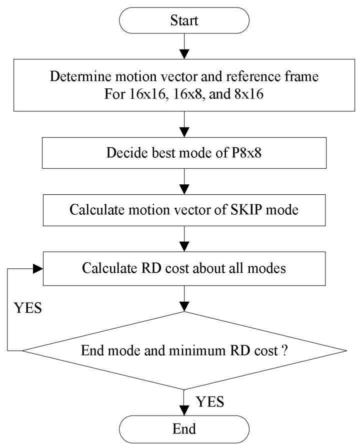


Fig. 1. Flowchart of mode decision in the H.264/AVC

In order to determine the best motion vector and the reference frame for the Inter16×16, Inter16×8, and Inter8×16, the H.264/AVC standard calculates J_{motion} based on the sum of absolute differences (SAD) and the bit rates for the motion vector and reference frame with λ_{motion}. λ_{motion} represents the Lagrangian multiplier that depends on QP. Next, the best P8×8 mode is determined using J_{motion} and J_{mode}. J_{mode} is calculated according to the sum of squared differences (SSD) and encoding bit rates with λ_{mode}. Next, the best intra mode is selected among the Intra4×4, Intra8×8, and Intra16×16 predictions. Finally, among these prediction modes, the mode, which has minimum J_{mode} cost, is selected as the best mode finally.

$J_{motion}\ (MV,REF \mid \lambda_{motion})$

$$=SAD(s,r(MV,REF))+\lambda_{motion}\cdot R(MV,REF) \qquad (1)$$

$SAD(s,r(MV,REF)$

$$=\sum_{x\in H,y\in H}^{H,V} |S(x,y)-r(x-x_m,y-y_m)| \qquad (2)$$

$$J_{mode}(S,R,M|\ \lambda_{mode}) = SSD(S,R,M) + \lambda_{mode}\cdot R(S,R,M) \qquad (3)$$

$SSD(s,r,M)$

$$=\sum_{x\in H,y\in H}^{H,V} (S(x,y)-r(x-x_m,y-y_m))^2 \qquad (4)$$

Fig. 1 shows the procedure of the mode decision of H.264/AVC. First, the motion vector and the reference frame were determined using J_{motion} for the Inter16×16, Inter16×8, and Inter8×16. Next, the best sub macroblock mode was decided. Finally, after J_{mode} for all modes except for P8×8 is calculated, the best mode is determined.

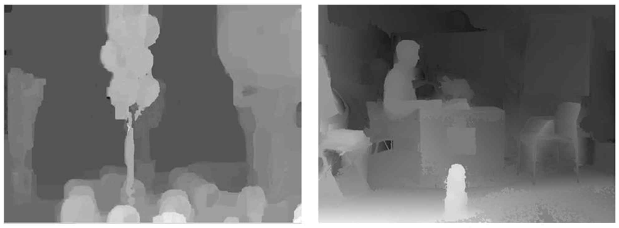
III. Proposed Method

A. Edge Classification

Depth sequences have different properties from those of texture sequences. Since Intra16×16, Intra4×4, and SKIP modes are selected frequently as the best mode, motion vector search and mode decision processes can be skipped in the homogeneous regions [5]. Before applying our proposed algorithm, we classify a macroblock into continuous and dis continuous regions. In order to separate continuous and discontinuous regions, the variation degree of depth value in a macroblock is defined by

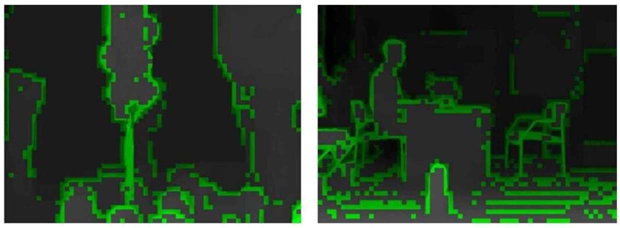
$$f(x,y)=\frac{1}{16}\sum_{i=1}^{16}\sum_{j=1}^{16}(r(i,j)-m_{x,y})^2 \qquad (5)$$

where coordinate (x,y) is the position of the current macroblock and $r(i,j)$ is the depth value at a relative coordinate (i,j) in the current macroblock. $m_{x,y}$ is the mean value of the current macroblock. If the depth values change drastically, the value of $f(x,y)$ is large. Thus the value of $f(x,y)$ is large in the discontinuity regions. On the contrary, in the 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 the boundaries or not. In the Fig. 2, we show that the original depth sequences and in the Fig. 3, we show the discontinuity regions which was determined by a threshold value, where the macroblocks with green color are depth discontinuity-regions with $f(x,y)>T$.



(a) (b)

Fig. 2.Original depth sequence (a) Balloon, (b) Book Arrival.



(a) (b)

Fig. 3. Segmentation of discontinuity-regions (QP=37 and T=30) (a) Balloon, (b) Book Arrival.

B. Analysis of Mode Selection

The macroblock mode distribution of the depth-continuity regions is different from that of depth-discontinuity regions. Fig. 4 shows the comparison of the mode distribution between depth-continuity and depth-discontinuity regions in balloon sequence. In the depth-continuity regions, mode distribution is severely imbalance. Most macroblocks are encoded as SKIP and the intra modes. However, in the depth-discontinuity regions, the mode is balanced. Therefore, we use this property to design a fast mode decision algorithm.

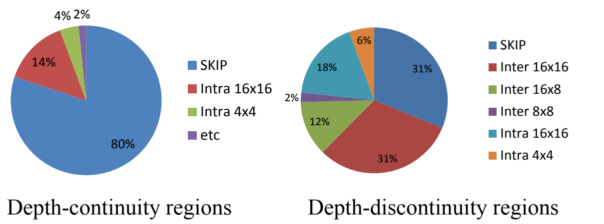
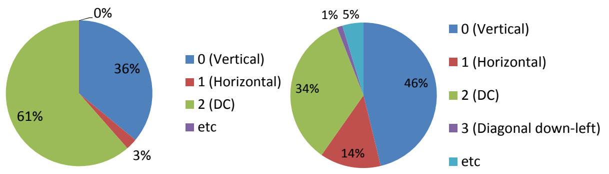


Fig. 4. Macroblock mode distribution

Since modes 0, mode 1, mode 2, and mode 3 take up to almost 100% in the depth-continuity regions, as shown in Fig. 5, we just consider the modes 0, mode 1, mode 2, and mode 3 among nine prediction modes as the candidate mode set in the Intra4×4. However, since the mode distribution for all four modes in the Intra16×16 is similar, we do not consider the candidate mode set in the Intra16×16 prediction. Therefore, we calculate SKIP, four candidate modes in the Intra4×4, and all modes in the Intra16×16 to determine the best mode in the depth-continuity regions.



Depth-continuity regions Depth-discontinuity regions

Fig. 5. Intra prediction mode distribution of Intra4×4

Since, in the depth-discontinuity regions, the distribution for all modes is various, we use the conventional mode decision method in H.264/AVC

C. Fast Depth Coding Algorithm

Peng's algorithm indicates that a threshold value, which discriminates depth-continuity and depth-discontinuity, is set to 30 [6]. If $f(x,y)$ is less than the threshold value or equal to the threshold value, a macroblock is determined as the continuous region. Since, SKIP and the intra modes are used as the candidate modes in the continuous regions, the motion estimation and mode decision processes are not performed. Therefore, we can reduce the complexity by skipping the inter mode in the mode decision. If $f(x,y)$is larger a threshold value, a macroblock is determined as discontinuous region, all macroblock mode decisions including the inter mode decision are implemented.

However, the Peng's algorithm has a problem. In the low QP, the algorithm does not fine the detailed boundary regions because it uses a fixed threshold value. To solve the problem, we use the adaptive threshold values for the intra and inter prediction empirically.

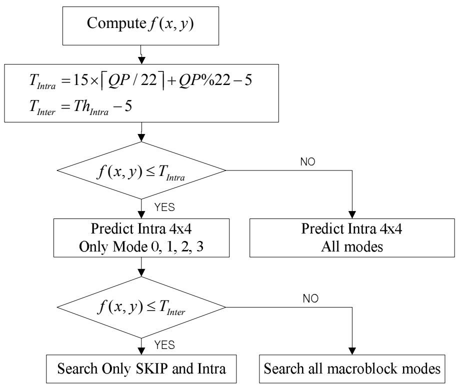


Fig. 6. Flowchart of the proposed algorithm

Fig. 6 shows the 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 coding efficiency at all QP ranges, we assign the small and large threshold values in low QP and high QP, respectively.

Since the skip of the inter mode decision induces the additional bits from wrong determination, the threshold value for the inter mode (T_{Inter}) is less than that for the intra mode (T_{Intra}). If $f(x,y) \leq T_{Intra}$, mode decision for the modes 0, mode 1, mode 2, and mode 3 in the intra4×4 is performed. Otherwise, all modes are considered to find the best intra prediction mode. Next, if $f(x,y) \leq T_{Inter}$, SKIP and intra are calculated to find the best mode of the current macroblock. Otherwise, we calculate all modes.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In order to evaluate efficiency of the proposed algorithm, we performed experiments on several depth sequences with 1024×768 resolutions. All test sequences have 50 frames. We implemented our proposed algorithm on the H.264/AVC reference software version JM12.4. The encoding parameters for the reference software are summarized in Table I.

TABLE I. EXPERIMENTAL CONDITIONS

Depth Sequences	**Book Arrival, Love Bird, and Newspaper**
Resolution	1024×768
The Number of Encoding Frame	50
Profile	Main profile
Search Range	32
Frame Structure	1) III …. I 2) IPP…. P
QP	22, 27, 32, 37
Symbol Mode	CAVLC

Table II shows bit rate increment exceeding ±1%, and PSNR reduction from -0.04dB to 0.01dB. The encoding time is saved up to 31%, compared to the full search algorithm. It compensates for the significant bit rate increment and decreased PSNR. Because Love Bird depth sequence has many continuous background regions, it shows better performance compared to other sequences.

TABLE II. COMPARISON OF THE INTRA CODING PERFORMANCE

Sequences	**QP**	**ΔPSNR (dB)**	**ΔBR (%)**	**ΔTS (%)**
Book Arrival	22	-0.02	0.35	18.58
	27	0.00	0.10	20.31
	32	0.00	0.08	21.52
	37	0.01	0.03	22.69
Love Bird	22	0.00	0.11	30.88
	27	0.01	0.47	31.32
	32	0.00	0.58	31.30
	37	-0.01	-0.19	31.33
Newspaper	22	0.00	0.11	30.88
	27	0.01	0.47	31.32
	32	0.00	0.58	31.30
	37	-0.01	-0.19	31.33
Average		-0.01	0.16	24.80

$$\Delta PSNR = PSNR_{proposed} - PSNR_{original} \quad (6)$$

$$\Delta BR = \frac{Bitrate_{original} - Bitrate_{proposed}}{Bitrate_{original}} \times 100 \quad (7)$$

$$\Delta TS = \frac{Time_{original} - Time_{proposed}}{Time_{original}} \times 100 \quad (8)$$

Table III shows comparison between Peng's and the proposed algorithm. There is negligible bit rate increment up to 1.13% and 5.07% in proposed algorithm and Peng's algorithm, which used the fixed threshold value as 30 entirely. Moreover, PSNR difference is decreased by an average of -0.06 dB and -0.09 dB in the proposed algorithm and Peng's algorithm. In terms of the RD cost, the performance of the proposed algorithm is better than Peng's algorithm. Since we applied the adaptive threshold values depending on QP, our proposed algorithm does not induce immense bit rate increment in low QP. Although Peng's algorithm reveals good performance in high QP, the bit rate increment is over 5% in high QP. However, our proposed algorithm does not reveal great bit rate increment because of adaptive threshold values depending on QP.

TABLE III. COMPARISON OF THE INTER CODING PERFORMANCE

Sequences	QP	Peng's algorithm			Proposed algorithm		
		ΔPSNR (dB)	ΔBR	ΔTS	ΔPSNR (dB)	ΔBR	ΔTS
Book Arrival	22	-0.22	3.22	77.33	-0.15	1.13	57.58
	27	-0.10	2.85	77.30	-0.06	0.66	63.67
	32	-0.10	0.80	77.44	-0.06	0.23	69.55
	37	-0.02	0.62	77.58	-0.03	0.11	72.88
Love Bird	22	-0.12	4.46	96.47	-0.14	-1.10	95.60
	27	-0.07	3.57	96.50	-0.07	0.82	96.41
	32	-0.05	0.28	96.51	-0.03	-0.65	98.36
	37	0.04	-0.64	96.51	0.08	0.69	98.51
Newspaper	22	-0.20	5.07	82.05	-0.09	1.12	61.41
	27	-0.13	2.62	82.25	-0.09	0.64	71.44
	32	-0.07	-0.82	82.25	-0.09	-1.01	75.31
	37	-0.06	-0.41	82.25	-0.04	-0.99	78.26
Average		-0.09	1.80	85.37	-0.06	0.14	78.25

TABLE IV. SELECTED VIEWS FOR EXPERIMENT

Depth Sequences	Left view	Right view	Virtual view
Newspaper and Love Bird	4	6	5
Book Arrival	7	9	8

To synthesize the intermediate virtual view, we used 4th and 6th views in Newspaper and Love Bird, and 7th and 9th views in Book Arrival, as shown in Table IV.

Table V shows the result of PSNR difference between the synthesized view and original sequences. In order to make the synthesized view, we used view synthesis reference software (VSRS) [7]. Both Peng's and the proposed algorithms have negligible differences in range from -0.04dB to 0.01dB. This result shows that both algorithms maintain rendering quality. Fig. 7 shows that there is no significant rendering quality degradation between full search and the proposed algorithm.

TABLE V. ΔPSNR OF THE SYNTHESIZED RESULT

Method	Sequences	QP			
		22	27	32	27
Peng's algorithm	Book Arrival	0.00	-0.04	0.02	0.03
	Love Bird	0.00	0.01	0.00	-0.02
	Newspaper	-0.01	0.01	-0.01	0.00
Proposed algorithm	Book Arrival	0.01	-0.04	0.01	0
	Love Bird	0.00	0.00	-0.01	-0.02
	Newspaper	0.00	0.00	0.00	0.00

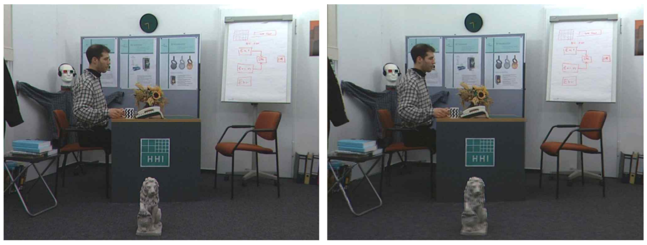


Fig. 7. Synthesized result of Book Arrival sequence (a) Full search algorithm, (b) Proposed algorithm

V. CONCLUSIONS

Although encoding multiview video plus depth (MVD) is time consuming, MVD is the effective method to represent three-dimensional (3D) scenes. Since depth-discontinuity regions have imbalanced macroblock mode distributions, we proposed a fast depth video coding algorithm using a threshold value that determines the depth-continuity or depth-discontinuity regions based on quantization parameter (QP). If the variation is smaller than the inter threshold value, the proposed algorithm uses the SKIP and intra modes with no motion estimation and 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. The threshold value for inter is lower than that for intra because skipping the motion estimation process is proportional to the bitrate increment induced from the mismatch. Since the motion estimation time accounts for most of the encoding time, our proposed algorithm reduces the encoding time up to 98% without significant degradation for the peak signal-to-noise ratio (PSNR) and rendering quality and increase for bit rate.

REFERENCES

This research was supported in part by MKE under the ITRC support program supervised by NIPA (NIPA-2011-(C1090-1111-0003)).

REFERENCES

[1] P. Benzie, J Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. V. Kopylow, "A Survey of 3dTV displays: techniques and technologies," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 7, pp. 1647-1658, Nov. 2007.

[2] Y.Mori, N. Fukushima, T.Fujii, and M. Tanimoto, "View generation with 3D warping using depth information form FTV," in Proc of 3DTV conference, pp. 229-232, May 2008.

[3] H. Oh and Y. Ho, "H.264/AVC-Based depth map sequence coding using motion information of corresponding texture video," Lecture Notes in Computer Science, vol 4319, Dec. 2006.

[4] http://iphome/hhi.de/shehring/tml/download/old_jm/jm12.4.zip, Joint Video Team, Reference Software Version 12.4.K. Elissa, "Title of paper if known," unpublished.

[5] I.E.G. Richardson, H.264 and MPEG-4 Video Compression, WILEY, 2003.

[6] Zongju Peng, Mei Yu, Gangyi Jiang, Yuehou Si, and Fen Chen, "Virtual view synthesis oriented fast depth video encoding algorithm," International Conference on Industrial and Information Systems, pp. 204-207, July 2010.

[7] M. Tanimoto, T. Fujii, and M. K. Suzuki, "View Synthesis Algorithm in View Synthesis Reference Software 2.0 (VSRS 2.0)," ISO/IEC JTC1/SC29/WG11 MPEG2009/M16090, Feb. 2009.