Circuits Syst Signal Process (2013) 32:803–813
DOI 10.1007/s00034-012-9501-3

Fast Depth Video Coding Method Using Adaptive Edge Classification

Da-Hyun Yoon · Yo-Sung Ho

Received: 15 February 2011 / Revised: 18 September 2012 / Published online: 9 November 2012

Abstract In this paper, we propose a fast-mode decision algorithm for both intra and inter prediction in depth videos. The proposed algorithm can reduce the computational 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 the encoding time by up to 78 % and 97 % for the intra and inter frames, respectively, compared to JMVC 8.3, with negligible PSNR loss and bitrate increase.

Keywords Depth video coding · Macroblock mode decision · Depth compression

1 Introduction

Due to advances in three-dimensional (3D) display technologies, the development of 3DTV has realized the human dream of viewing scenes as if in the real world. 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 utilized because it transmits and records all spatiotemporal information from the real world [1]. Multiview plus depth (MVD) is another framework used

D.-H. Yoon (✉) · Y.-S. Ho
Gwangju Institute of Science and Technology (GIST), 261 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, Korea
e-mail: yoon@gist.ac.kr

Y.-S. Ho
e-mail: hoyo@gist.ac.kr
url: http://vclab.gist.ac.kr

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 [10]. Since the amount of data and complexity of MVD is proportional to the number of cameras, an efficient encoding method for 3D video scenes must be developed.

An efficient multiview video coding (MVC) scheme is needed to transmit and store multiview contents. The MPEG 3D Audio/Video (3DAV) group is working on the standardization of MVC. The joint MVC (JMVC) was released by the MPEG 3DAV group as the reference software.

JMVC uses a rate-distortion optimization (RDO) technique [8] to select the optimal coding mode, where the rate-distortion (RD) cost is calculated by

$$J = D + \lambda R \tag{1}$$

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. JMVC 8.3 [3] is then used to select the best mode among 14 different macroblock modes: SKIP, Direct, Inter 16×16, Inter 16×8, Inter 8×16, Inter 8×8, Inter 8×8 Frext, Inter 8×4, Inter 4×8, Inter 4×4, Intra 16×16, Intra 8×8, Intra 4×4, 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 objects 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 or object. 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 [7] 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 [11]. Lee [4] skipped some blocks in the depth image to reduce the depth coding bitrate according by 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, focusing on depth video will be benefit if sharing information between is carried out, although depth video is always companied with color video for the free-view point application.

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 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 is calculated adaptively according to the quantization parameter (QP). Our experimental results show that the fast-mode decision scheme reduces the computational complexity without incurring any noticeable coding loss.

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

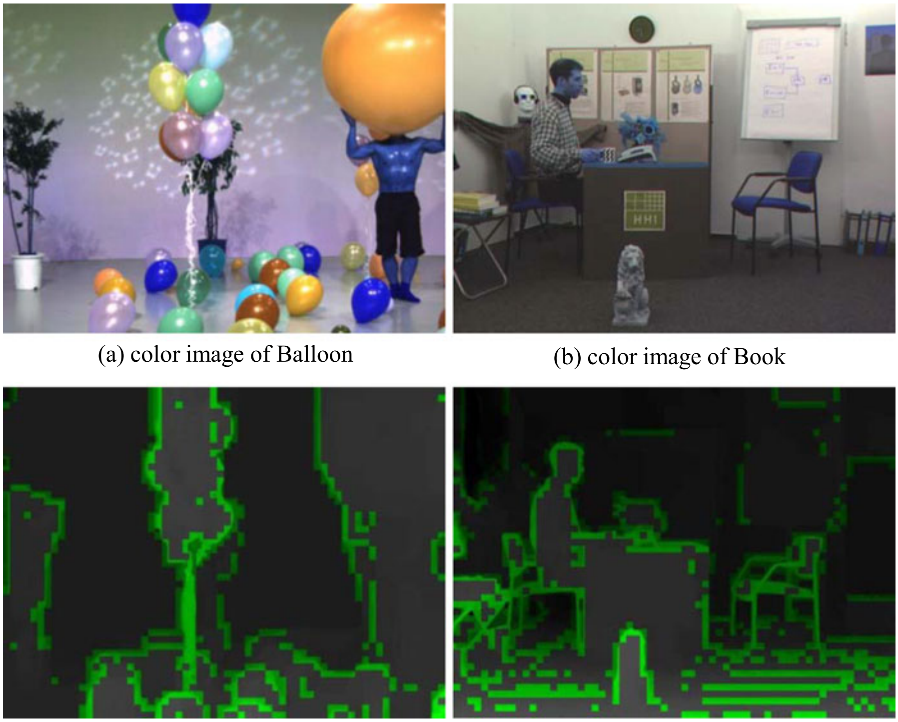
$$f(x, y) = \frac{1}{16^2} \sum_{i=1}^{16} \sum_{j=1}^{16} \bigl(r(i, j) - m_{x,y}\bigr)^2 \tag{2}$$

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 $f(x, y) > T$.

2.2 Analysis of Mode Selection

The macroblock mode distribution of the depth-continuity regions is different from that in depth-discontinuity regions. Figure 2 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 Figs. 2 and 3.

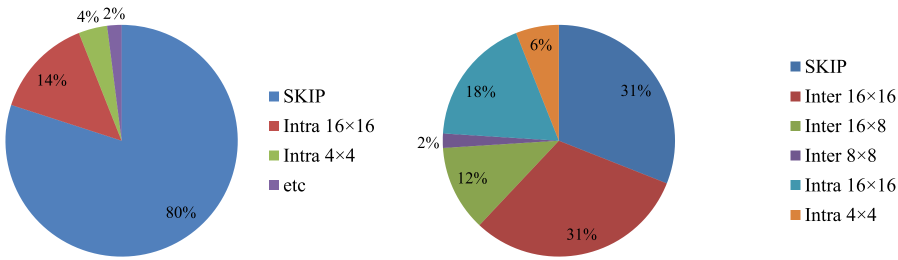
Figure 3 shows that Intra 4×4 prediction mode distribution in depth-continuity and discontinuity regions. Since Mode 0, Mode 1, Mode 2, and Mode 3 comprise



(a) color image of Balloon (b) color image of Book

(c) depth image of Balloon (d) depth image of Book Arrival

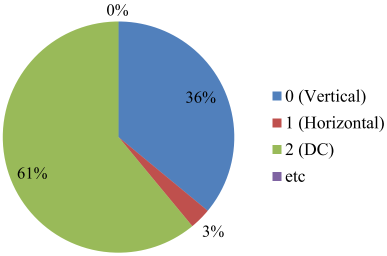
Fig. 1 Segmentation of discontinuity regions ($T = 30$)



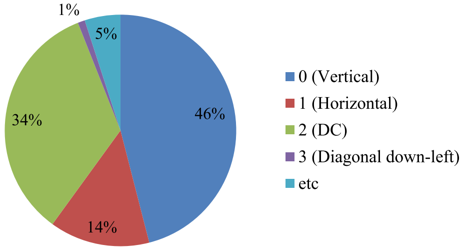
(a) Depth-continuity regions (b) Depth-discontinuity regions

Fig. 2 Macroblock mode distribution

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 4×4 prediction. However, since the mode distributions for all four modes in the Intra 16×16 prediction are similar, we consider the candidate mode set in the Intra 4×4 prediction only. 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 to



(a) Depth-continuity regions



(b) Depth-discontinuity regions

Fig. 3 Intra prediction mode distributions

Table 1 Encoding configurations

Parameter	Setting
Sequence	Balloon
Threshold	30
QP	32
View	1
Time	0

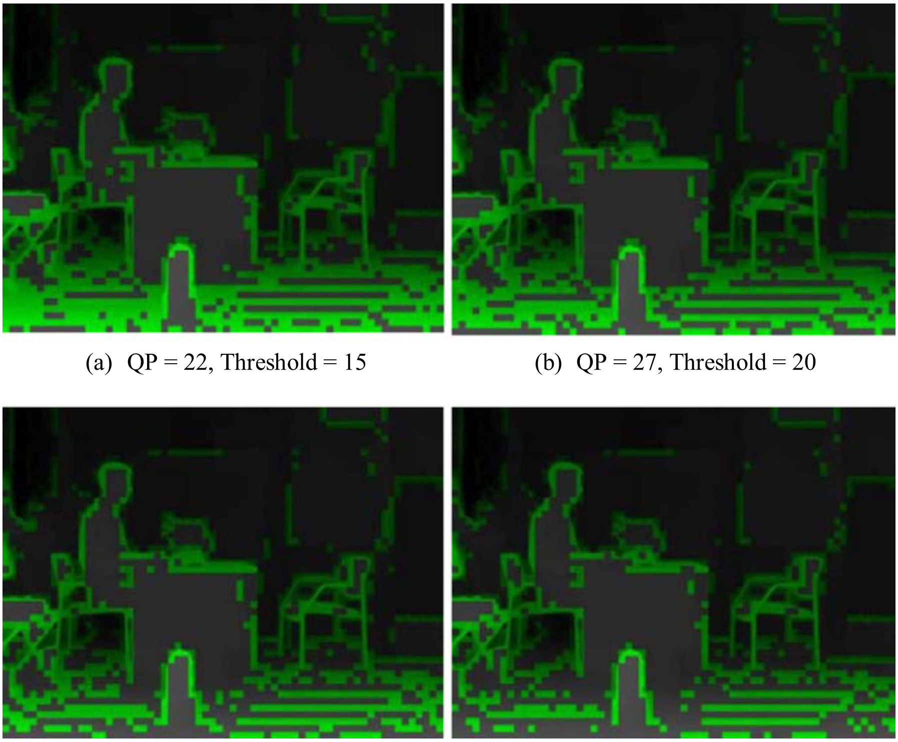
determine the best mode in the depth-continuity regions. Since the distribution for all modes varies in the depth-discontinuity regions, we use a conventional mode decision method in JMVC.

2.3 Fast Depth Coding Algorithm

The algorithm developed by Peng [6] 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, the complexity can be reduced 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, Peng's algorithm has a problem. In low QPs, the 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.

Figure 4 shows that depth segment results using dynamic threshold adaptive to QP. In low QP, there is low threshold and edge segmentation is sensitive. In high QP, there is high threshold and edge segmentation is rough. Less green color represents that skipping inter mode decision has more often occurred. Thus, accurate mode decision is performed in low QP.

Figure 5 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



(a) QP = 22, Threshold = 15 (b) QP = 27, Threshold = 20

(c) QP = 32, Threshold = 25 (d) QP = 37, Threshold = 30

Fig. 4 Intra prediction mode distributions

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}). If $f(x, y) < T_{Inter}$, the SKIP and intra modes are calculated to find the best mode of the current macroblock. Otherwise, we calculate all modes. 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 Result

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.

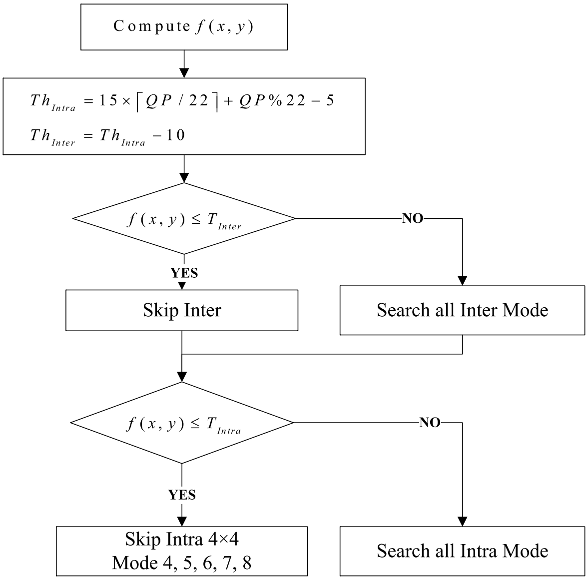


Fig. 5 Flowchart of the proposed algorithm

Table 2 Experimental conditions

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

Table 3 shows experimental result of intra frame coding performance. 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. To compare the coding performance, PSNR difference, percentage of time saving and bitrate is used.

$$\Delta PSNR = PSNR_{proposed} - PSNR_{original}, \tag{3}$$

Table 3 Comparison of intra coding performance in terms of $\Delta PSNR$, ΔBR, and ΔTS

Sequences	QP	$\Delta PSNR$ (dB)	ΔBR (%)	ΔTS (%)	$BDBR$ (%)	$BDPSNR$ (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

$$\Delta TS = \frac{Time_{proposed} - Time_{original}}{Time_{original}} \times 100, \tag{4}$$

$$\Delta BR = \frac{Bitrate_{proposed} - Bitrate_{original}}{Bitrate_{original}} \times 100. \tag{5}$$

Tables 4 and 5 compare Peng's algorithm and the proposed algorithm. There is a negligible bitrate and PSNR reduction of up to 3.53 % and 0.16 dB in the proposed algorithm. The average PSNR difference is decreased by 0.16 dB and 0.24 dB in the proposed algorithm and Peng's algorithm, respectively. The bitrates are reduced by −1.53 % and −1.91 % in the two algorithms on average, respectively.

In order to measure the overall improvement of the proposed method, we have employed the Bjontegaard delta bitrate (BDBR) and PSNR (BDPSNR) values [2]. 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. Figure 6 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. The proposed method reduces the encoding time by an average of 76.71 %.

To synthesize the intermediate virtual view, we used the fourth and sixth views in Newspaper and Love Bird, and the seventh and ninth views in Book Arrival, as shown in Table 6.

Table 7 shows the results of PSNR differences between the synthesized view and original sequences. VSRS rendering software was used for this task [9]. Both Peng's

Table 4 Inter coding performance in terms of $\Delta PSNR$, ΔBR, and ΔTS for Peng's algorithm

Sequences	QP	$\Delta PSNR$	ΔBR	ΔTS	$BDBR$	$BDPSNR$
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		
Newspaper	22	−0.35	−1.37	−78.94	2.34	−0.08
	27	−0.27	−2.76	−79.05		
	32	−0.14	−1.92	−79.64		
	37	−0.07	−0.36	−80.66		
Average		−0.24	−1.91	−71.77	3.35	−0.143

Table 5 Inter coding performance in terms of $\Delta PSNR$, ΔBR, and ΔTS for the proposed algorithm

Sequences	QP	$\Delta PSNR$	ΔBR	ΔTS	$BDBR$	$BDPSNR$
Book Arrival	22	−0.16	−0.01	−79.60	3.65	−0.14
	27	−0.28	−2.66	−54.10		
	32	−0.21	−2.43	−58.95		
	37	−0.22	−0.46	−61.31		
Love Bird	22	−0.20	−3.53	−97.11	1.29	−0.065
	27	−0.14	−1.48	−82.17		
	32	−0.11	−1.22	−83.31		
	37	−0.11	−0.62	−84.25		
Newspaper	22	−0.16	−1.09	−70.82	0.79	−0.039
	27	−0.14	−2.25	−79.42		
	32	−0.11	−1.48	−83.20		
	37	−0.11	−0.08	−86.29		
Average		−0.16	−1.53	−76.71	1.91	−0.0813

and the proposed algorithm have negligible differences in the range from −0.01 dB to 0.06 dB, indicating that both algorithms maintain rendering quality. Figure 7 shows that there is no significant rendering quality degradation between the full search and the proposed algorithm.

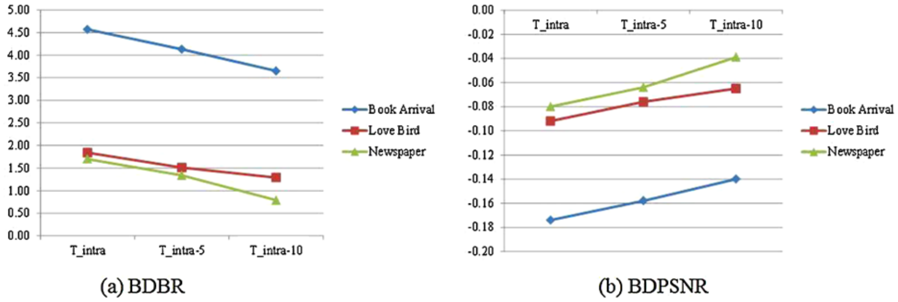


Fig. 6 BDBR and BDPSNR change according to T_{inter} value

Table 6 Selected views for experiment

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

Table 7 $\Delta PSNR$ 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	0.01	0.03	0.05	−0.01
	Love Bird	0.00	0.01	0.00	−0.01
	Newspaper	0.00	−0.01	0.01	0.00

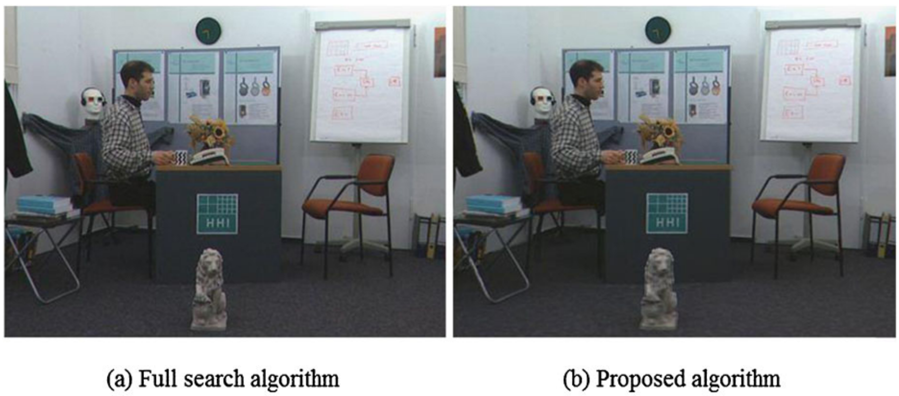


Fig. 7 Synthesized results of Book Arrival sequence ($QP = 22$)

4 Conclusions

Although MVD coding is time consuming, it is still the most effective way to represent 3D scenes. The property of depth sequences is different from that of texture sequences. 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 adaptive to the QP. If the variation is lower than the inter threshold value, 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. This threshold value for inter is lower than that for intra due to the fact that skipping the motion estimation process is proportional to the bitrate increment induced from the mismatch. As such, since the motion estimation time accounts for most of the encoding time, our algorithm reduces the encoding time by up to 78 % and 97 % for the intra and inter frames, respectively, with no significant degradation for the PSNR and rendering quality, or bitrate increment

Acknowledgement This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0030822).

References

1. P. Benzie, J. Watson, P. Surman, C. Kopylow, H. Urey, I. Rakkolainen, K. Hopf, V. Sainov, A survey of 3DTV displays: techniques and technologies. IEEE Trans. Circuits Syst. Video Technol. **17**(7), 1647–1658 (2007)
2. G. Bjontegaard, Calculation of average PSNR differences between RD-curves, Doc. VCEG-M33, April 2001
3. Joint Multiview Coding (JMVC) 8.3, garcon.ient.rwth-aachen.de, May 2010
4. J. Lee, H. Wey, D. Park, A fast efficient multiview depth image coding method based on temporal and inter-view correlations of texture images. IEEE Trans. Circuits Syst. Video Technol. **99**, 1–4, (2011)
5. H. Oh, Y. Ho, H.264/AVC-based depth map sequence coding using motion information of corresponding texture video, in *Lecture Notes in Computer Science*, vol. 4319 (Springer, Berlin, 2006), pp. 898–907
6. Z. Peng, G. Jiang, M. Yu, Y. Si, F. Chen, Virtual view synthesis oriented fast depth video encoding algorithm, in *International Conference on Industrial and Information Systems*, July 2010, pp. 204–207
7. K. Shin, K. Chung, S. Chun, A fast mode prediction of multi-view video coding using region analysis, in *6th International Conference on Digital Content, Multimedia Technology and Its Applications*, August 2010, pp. 87–90
8. G. Sullivan, T. Wiegand, Rate-distortion optimization for video compression. IEEE Signal Process. Mag. **15**, 74–90 (1998)
9. M. Suzuki, M. Tanimoto, T. Fujii, View synthesis algorithm in view synthesis reference software 2.0 (VSRS 2.0), ISO/IEC JTC1/SC29/WG11 MPEG2009/M16090 (2009)
10. M. Tanimoto, N. Fukushima, T. Fujii, Y. Mori, View generation with 3D warping using depth information form FTV, in *3DTV Conference*, May 2008, pp. 229–232
11. W. Whu, W. Jiang, Y. Chen, A fast inter mode decision for multiview video coding. Consumer Electronics, 689–694 (2011)