

A Fast Mode Decision Algorithm Based On MVD

Huimin Zhao¹ and Xiuhua Jiang² and Caihong Wang³

Abstract There are a lot of researches on acceleration the mode decision process in MVD video coding. The main goal of these researches lies in speeding-up the encoding process with negligible coding efficiency loss. Our work reveals the usage of depth information for speeding up video encoding process. We proposed a novel fast mode decision algorithm based on depth information. The comparative experimental results show that the proposed algorithm reduces the computational complexity of MVD coding while improves the coding performance, and the video quality maintain rate-distortion performance at a level which is similar to the original algorithm based on the JMVC.

Keywords: the video plus depth, multi-view video plus depth, depth mode decision algorithm

1 Introduction

With the advance of 3D technologies and the increasing requirement of watching, three-dimensional television (3DTV) technologies which are also developing rapidly have gradually become a focus in the field of consumption in recent years. For coding and display purposes, the captured sequences need to be converted from the production into the transport format by post-processing. There are so many formats in 3D video, but all the 3D video formats can be divided into two main classes [1][2]: video-only formats and depth-enhanced formats. All of the three formats: classical stereo video (CSV) with 2 views, mixed resolution stereo video (MRS) with one view spatially sub-sampled and multi-view video

¹ Huimin Zhao (✉)

College of Information Engineering, Communication University of China, Beijing, China
e-mail:huimin_0316@163.com

² Xiuhua Jiang

College of Information Engineering, Communication University of China, Beijing, China

³ Caihong Wang

College of Information Engineering, Communication University of China, Beijing, China

(MVV) with more than 2 views are Video-only formats. Video plus depth (V+D), multi-view video plus depth (MVD), and layered depth video (LDV) are depth enhanced formats. There are some problems of the video-only formats. One basic problem of it is which they do not support the adaptation of 3D video content to the actual display conditions. And with the increasing of views, the number of data will more and more large. The 3D impression varies with the viewing position, display resolution, distance to the screen, etc. In contrast to video-only formats, these requirements are supported by depth-enhanced formats. The depth or disparity information included in the representation allows for adaptation to different displays by rendering or synthesizing virtual views. Therefore current activities concentrate on developing new technologies and standards for depth enhanced formats.

In this paper, we focus the encoding of Multi-view video plus depth. The research issues of depth video coding are different from the normal 2D color video coding which have been addressed in the works [9], [10], [11]. The HEVC [3] which finalized by VCEG (Visual Coding Experts Group) and MPEG (Moving Pictures Experts Group) have proposed the multi-view plus depth (MVD) format. In the extension of HEVC, in addition to there are 4 modes for depth encoding, HEVC also usage the motion vector information of the texture. The motion vector sharing technique has been addressed in [12] by investigating the similarity of the motion vectors between color videos and depth videos. Because the MVD format data contains multiple texture video sequences, which are viewing the same scene from different camera position. And the depth video is not represented on TV directly. It can be rendered to a new view video at consumer side by depth-image-based-rendering (DIBR) techniques [4] with the texture video. So the video plus depth can apply to more condition than other formats. And each depth video provides a depth value for per pixel of one view represented by the corresponding texture video sequence. The regions of depth video are usually smooth, which can be compressed effectively. As prescribed, the video plus depth 3D video gains the attention for its relatively low storage and transmission bandwidth requirements as compared with stereo and multi-view videos.

No matter what 3D format it is, the information of 3D video are larger than traditional 2D video. Therefore, how to code effectively for 3D video is one of the key technologies for 3D television transmission. In this paper, we focused on 3D video encoding, and proposed a novel fast mode decision algorithm based on depth information. Section 2 briefly describes the nature of color plus depth-map 3D video data, and the proposed fast mode decision algorithm is illustrated. Experimental results are shown and discussed in Section 3. Section 4 concludes this work.

2 The Proposed Fast Mode Decision Algorithm

In this section, the nature of color plus depth-map 3D video data, and the proposed fast mode decision algorithm will be discussed.

MVD is a kind of 3D video representation formats. Fig 1 shows the structure of the MVD. Such a format is the combination of texture video and associated depth maps. Although a depth video was different from a texture video, but both the texture video and depth video in one view are shot from a same angel. So there is a connection between them: the object boundary and scene movement are similar in both sequences. Essentially, MB mode correlation between texture and depth map sequence can be used to speed up the mode decision in joint coding.

In the MVC reference software JMVC [5][6], there are different kinds of mode sizes including 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 and 4x4. The original encoding algorithm in JMVC need test the entire MB mode to decide which the best it is. The amount of calculation is large and it will spend much time. In this paper, a fast mode size decision based on depth map is proposed to reduce MVD computational complexity.

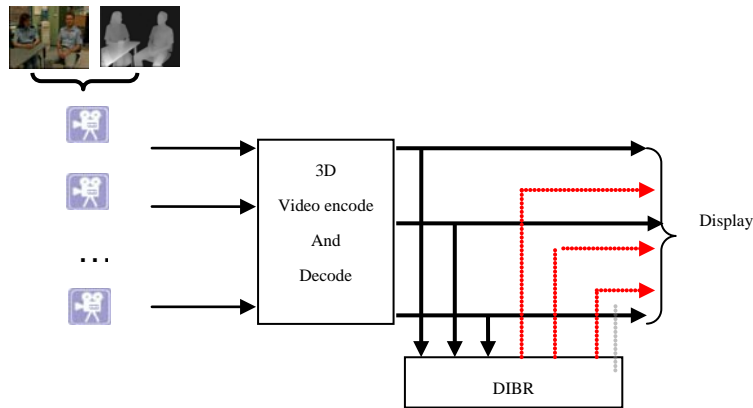


Fig. 1 the structure of the MVD

Because a depth map represents a relative distance from a camera to an object in the 3D space, and it is represented as a gray image which the pixel value set from 0 to 255. There are much smooth regions in the depth map, which is different from texture video. So the edge information is easier to get. In this paper, we utilize the depth information to provide mode candidates for texture encoding, which decrease the number of operations and speed up the encoding time. We can classify the 3D signal space into three regions: Near region, Middle region and Far region according to the depth information. Base on the depth information, each macro-block can be classified into one of the three regions when the associated depth information satisfies one of the conditions listed in equations1-3.

$$depth_avg(MB) \geq thresholdnear, MB \in near\ region \quad (1)$$

$$thresholdfar \geq depth_avg(MB) \geq thresholdnear, \quad (2)$$

$$MB \in middle\ region$$

$$depth_avg(MB) \leq thresholdfar MB \in far\ region \quad (3)$$

$depth_avg(MB)$ is the average intensity value in the depth-map for the macro-block MB. We reference the paper [7] to select the $thresholdnear$ and $thresholdfar$. The $thresholdnear$ is 20, and the $thresholdfar$ is 220. We chose seven sequences for MB mode statistic. The detailed information of the test sequences is provided in Table 1. Figure 2 shows the average mode type distributions for each MB with different frame type based on depth map. It is observed that for MB in near regions, the probability of choosing size 16x16 is about 90% for P frame, and the size 8x16 is about 95 for B frame. For MB in Middle region mode, nearly all the size have been chosen for P frame, and the most probability of choosing size is 16x16 for B frame. The percentage of 4x4 and 16x16 are nearly for I frame. In Far region mode, the average probabilities of 16x16 are larger than other sizes. So the depth map can be used to classify the motion activities of objects by this property. Object in the near region and in the Far region are less than that in the Middle region. Also people will give much attention to objects in Middle region. And large sizes are usually selected for the MB in regions with smooth region, while small sizes are selected for the MB with more details. The edge image has much high frequency so that small MB size may be more suitable than large MB size. Above all, the mode candidates for three regions are shows in table2

Table 1 Test Sequence information

Sequence	Resolution	Frame rate	GOP	Views
Kendo	1024x768	25	8	1-3-5
Balloons	1024x768	25	8	1-3-5
Lovebird1	1024x768	25	8	4-6-8
Newspaper	1024x768	25	8	2-4-6
Poznan_Hall2	1920x1088	25	8	7-6-5
Poznan_Street	1920x1088	25	8	5-4-3
Cafe	1920x1088	25	8	2-3-4

A flow chart of the proposed fast mode size decision is given in Fig 3. In the proposed algorithm, the task of step 1 which is the first stage of the proposed algorithm is to find these areas. We select the variance of depth MB as a judge standard for edge detection. Because it is easy to computer, and result of depth map edge detection are as well as using a Sobel edge detector. Then step 2 begins by detecting the region of the MB of the depth in order to decide what mode candidate will be used to predict the mode of texture MB. The candidate of

different MB is made as follows: for non-edge MB with near region, size 16x16 and 8x16 are tested, and for edge MB with near region, size 16x16, 8x8 and 4x4 are tested; for MB with middle region, all sizes are tested; for non-edge MB with far region, size 16x16 are tested, and for edge MB with far region, size 16x16, 8x8 and 4x4 are tested.

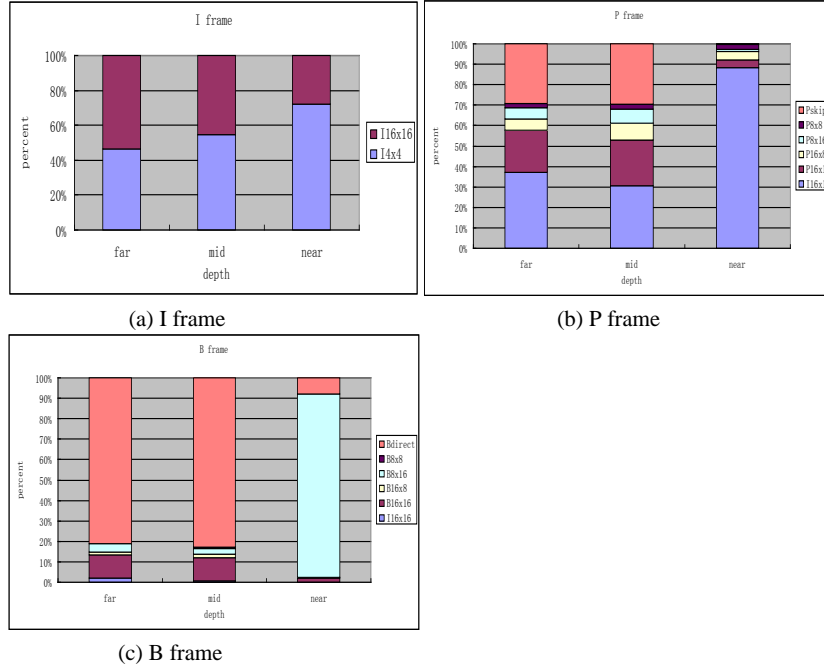


Fig. 2 The average mode type distributions for each MB with different frame type base on depth map

Table 2 The mode candidates for three regions

Region	Intra(edge)	Intra(not edge)	inter(edge)	inter(not edge)
Near region	Intra16x16, Intra4x4, IPCM	Intra16x16, IPCM	Inter8x8, SKIP	Inter16x16, Inter8x16, SKIP
Middle region		Intra16x16, Intra4x4, IPCM	All inter modes	All inter modes
Far region		Intra16x16, IPCM	Inter8x8, SKIP	Inter16x16, SKIP

3 Experiment Results and Analysis

The experiments were performed on the MVC reference software JMVC 8.3. We give the test results of four sequences released by MPEG [8], which were Kendo, Balloons, Lovebird1, and Newspaper in 3-view configuration. In the experiments, We adopt I-B-P view structure for all sequences, test 25 frames for each sequence, and the GOP size is 8. The coding structure use hierarchical B-frames and the search range is 64, and basis QP set 24, 28 and 32. The goal of the algorithm proposed in this paper is the reduction of the computational burden of the ME and MD stages in the encoding process of the depth maps. Therefore, in order to evaluate the reduction complexity, the experimental analysis is also oriented to the evaluation of the differences in terms of coding time between the proposed fast method and a traditional MVC encoder. Fig. 4 shows the RD performance comparison between the proposed fast mode size decision algorithm and the original JMVC 8.3. $view_x_original$ represent the sequence of view x encoding by JMVC8.3. And $view_x_original$ represent the sequence of view x encoding by the algorithm we proposed. Bitrate represents the total bitrates for multi-view texture coding; PSNR represents the average PSNR gain for each view. Each RD curves of the proposed algorithm and the original algorithm in MVD coding is near.

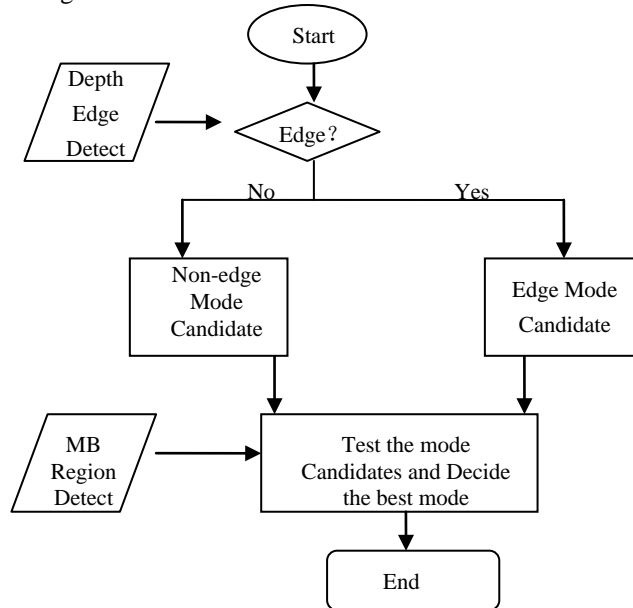


Fig. 3 Flowchart of the proposed fast mode size decision algorithm

The complexity comparisons among the proposed algorithm are lower than the original. The proposed algorithm reduces the MVD coding encoding time by about 46% on average with limited bitrates increment and PSNR loss.

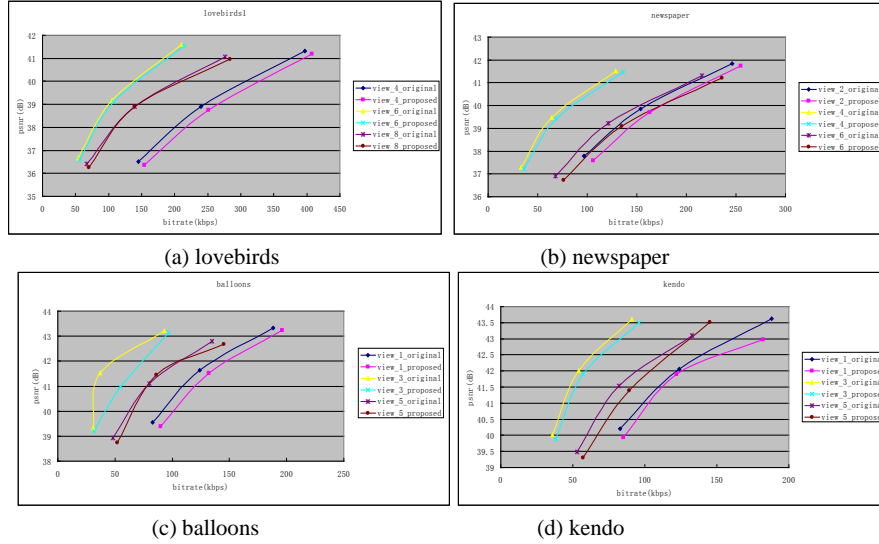


Fig. 4 RD curves of the proposed algorithm and the original algorithm in MVD coding

5 Conclusions

This paper presents a low-complexity algorithm to reduce the computational complexity of the MVD encoder. Based on the texture and depth map correlation, the proposed algorithm utilizes the depth information for fast mode size decision in texture coding. The comparative experimental results show that the proposed algorithm reduces the computational complexity of MVD coding while improve the coding performance, and the video quality maintain RD performance at a level which is similar to the original algorithm. We should jointly consider the above properties of texture and depth map, so it is not enough for the proposed algorithm only use the depth information for coding. How to utilize the texture information in MVD coding to increase the coding efficiency will be main directions of our future work.

References

1. P. Merkle, K. Muller and T. Wiegand, "3D Video: Acquisition, Coding and Display" Digest of Technical Papers, IEEE International Conference on Consumer Electronics, pp. 127-128, January 2010.
2. Philipp Merkle, Karsten Müller, Thomas Wiegand, "3D Video Coding - An Overview of Present and Upcoming Standards." Proc. IEEE Visual Communications and Image Processing, VCIP'10, Huang Shan, An Hui, CN, July. 2010
3. G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 12, pp. 1648–1667, Dec. 2012
4. C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3d-tv," SPIE Stereoscopic Displays Virtual Reality Syst. XI, vol. 5291, no. 1, pp. pp. 93–104, 2004.
5. D. Tian, P. Pandit, P. Yin, and C. Gomila, "Study of MVC Coding Tools," JVT Doc. JVT—Y044, Shenzhen, China, Oct. 2007.
6. Houjie Bi. "The new standard of video compression coding—H.264/AVC"[M](in Chinese). Beijing: people's posts and telecommunications of publishing company, 2009.
7. Lin YH, Wu JL, "A depth information based fast mode decision algorithm for color plus depth-map 3D videos." IEEE Trans. Broadcast 2011,57(2): 542–550.
8. ISO/IEC JTC1/SC29/WG11, "Call for Proposals on 3D Video Coding Technology," Doc. w12036, Mar. 2011.
9. D. De Silva and W. Fernando, "Intra mode selection for depth map coding to minimize rendering distortions in 3d video," IEEE Trans. Consum. Electron., vol. 55, no. 4, pp. pp. 2385–2393, Nov. 2009.
10. D. De Silva, W. Fernando, and H. Arachchi, "A new mode selection technique for coding depth maps of 3d video," in IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), Mar. 2010, pp. 686–689.
11. B. Kamolrat, W. Fernando, M. Mrak, and A. Kondo, "3d motion estimation for depth image coding in 3d video coding," IEEE Trans. Consum. Electron., vol. 55, no. 2, pp. 824–830, May 2009.
12. S. Grewatsch and E. Muller, "Sharing of motion vectors in 3d video coding," in Int. Conf. Image Process. (ICIP), Oct. 2004, vol. 5, pp. pp.3271–3274.