J Real-Time Image Proc
DOI 10.1007/s11554-013-0328-3

ORIGINAL RESEARCH PAPER

Fast mode decision based on texture–depth correlation and motion prediction for multiview depth video coding

Zhaoqing Pan · Yun Zhang · Sam Kwong

Received: 20 September 2012 / Accepted: 8 February 2013

Abstract The motion estimation and disparity estimation are used to remove the temporal and inter-view redundancies in multiview plus depth video coding, however, the variable block-size ME and DE make the computational complexity increase dramatically. This drawback limits it to be applied in real-time applications. In this paper, based on the mode correlations between depth video and its corresponding texture video, motion prediction and coded block pattern, we propose a fast mode decision algorithm to reduce the computational complexity of multiview depth video coding. Experimental results show that the proposed algorithm can achieve 67.18 and 69.90 % encoding time saving for even and odd views, respectively, while maintaining a comparable rate-distortion performance. In addition, with the dramatic encoding time reduction, the proposed algorithm becomes more suitable for real-time applications.

Keywords Three-dimensional video · Fast mode decision · Multiview depth video coding · Video coding

Z. Pan · Y. Zhang · S. Kwong (✉)
Department of Computer Science,
City University of Hong Kong,
Kowloon, Hong Kong
e-mail: cssamk@cityu.edu.hk

Z. Pan
e-mail: zqpan3@student.cityu.edu.hk

Y. Zhang
e-mail: yunzhang@cityu.edu.hk

Y. Zhang
Shenzhen Institutes of Advanced Technology,
Chinese Academy of Sciences, Shenzhen, China

1 Introduction

As the demand for real-world visual perception increases, three-dimensional (3D) video is becoming more and more popular. Multivew video plus depth (MVD), which consists of multiview texture video and corresponding depth video, is an advanced 3D video representation format for the 3D applications, such as free-viewpoint television (FTV), three-dimensional television (3DTV) broadcasting, immersive teleconference and so on. The multiview texture video is captured simultaneously by multiple cameras from different viewpoints, and the depth maps provide the geometrical information for their corresponding texture video. At last, the 3D video is generated by image-based rendering techniques [1]. However, as the number of capturing cameras increases, the volume of raw texture and depth video data increases rapidly. To efficiently encode MVD, multiview video coding (MVC) is developed as an extension of H.264/AVC standard to exploit spatial, temporal and inter-view redundancies [2, 3]. Due to the variable block-size motion estimation (ME) [4] and disparity estimation (DE), the computational complexity of MVC is quite high.

To address the high computational complexity of variable block-size ME, a number of fast mode decision algorithms have been proposed for H.264. Based on the distribution of motion activity in each frame, a fast mode decision was proposed for H.264/AVC [5]. Hu et al. [6] proposed a fast inter mode decision for H.264/AVC, based on rate-distortion (R-D) cost characteristics. Zhao et al. [7] proposed an adaptive fast mode decision algorithm, which projects all candidate modes into a 2-D map, then the mode decision is performed according to a priority-based mode candidate list. Based on the optimal stopping theory and all-zero block detection, they also proposed a fast mode

Published online: 03 March 2013

decision algorithm for H.264/SVC [8]. These algorithms can efficiently reduce the computational complexity in H.264. However, they are not efficient enough when used directly in MVC, because the characteristics of hierarchical B frames and the correlations between inter-views are not considered. Further, they also do not consider the coding correlations between depth video and texture video.

To use the correlations among inter-views, many researchers have devoted their efforts to develop fast mode decision for MVC. Based on the macroblock (MB) motion homogeneity, a selective DE and variable size ME algorithm was proposed in Ref. [9]. In that algorithm, the motion homogeneity is measured by the motion vectors of spatial neighboring MBs and inter-view collocated MBs. Based on the coding mode complexity, they also proposed an early SKIP mode decision for MVC [10]. These two algorithms can reduce the computational complexity of MVC efficiently. However, in these two algorithms, the inter-view collocated MBs are obtained by the global disparity vectors (GDVs). Whereas GDVs are based on global displacement and measured in MB level, they are not accurate enough for videos which are with large depth-of-fields (DOFs) and captured by toed-in camera arrangement [11]. Especially, they can only be used to odd (inter-view) views, and the coding correlations between depth video and its corresponding texture video are also not considered by these two algorithms. In addition, these algorithms were proposed for multiview texture video coding and not efficient for depth video coding due to different characteristics between the depth and texture.

To encode multiview depth video efficiently, several methods have been proposed. In Ref. [12], based on H.264/AVC coding structure and depth video variation, a fast mode decision algorithm was proposed for depth video coding. Based on the encoded information from the corresponding MB in the base view, Micalle et al. [13] proposed a fast inter-mode decision for multiview video plus depth coding. By considering the coding correlations between depth video and texture video, Peng et al. [14] proposed a fast MB mode algorithm for multiview depth video coding, which is based on mode prediction and object boundary discriminating method. Zhang et al. proposed a SKIP mode decision for depth video video coding [15]. In that algorithm, current depth MB directly uses the SKIP mode as the best mode selection when its corresponding MB in texture video is encoded as SKIP mode. However, this will lead to the R-D performance degrade dramatically when the mode prediction is not accurate.

In this paper, we propose a fast mode decision for multiview depth video coding, based on the best mode correlations between depth and texture video, motion prediction and coded block pattern (CBP). The rest of this paper is organized as follows. The statistical analyses and motivations are presented in Sect. 2. Then, the details of the proposed fast mode decision algorithm are presented in Sect. 3. Experiment results are shown in Sect. 4. Finally, Sect. 5 concludes this paper.

2 Statistical analyses and motivations

Multiview depth video is the associated per-pixel depth of its texture video, the mode between depth video and texture video may be quite similar, since they have some common object structures [14]. Based on this characteristic, we propose a fast mode decision algorithm for multiview depth video coding (MDVC) using the coding information of multiview texture video coding (MTVC). A joint multiview depth and texture video coding structure is illustrated in Fig. 1, where the joint coding structure includes two parts, MTVC and MDVC; MTVC adopts the MVC hierarchial B picture (HBP) prediction structure [16]; the texture videos are encoded before depth videos, then the coding information (such as best mode selection) is recoded, and used in encoding depth videos. In Fig. 1, the length of group of pictures (GOP) equals to 8; S_n represents the nth camera's view; T_n denotes the nth frame in temporal direction; the dotted lines represent the best mode of texture MBs, which is used for encoding their corresponding depth MBs. Figure 2 gives an example of depth MB and its associated MB in texture video. To exploit the mode selection correlations between depth video and its corresponding texture video, three multiview depth videos and their associated texture videos (Balloons, Cafe and Kendo) are tested. The test conditions are tabulated in Table 1. The statistical results of the MBs in depth video which have the same optimal mode with their corresponding MBs in texture video are listed in Table 2.

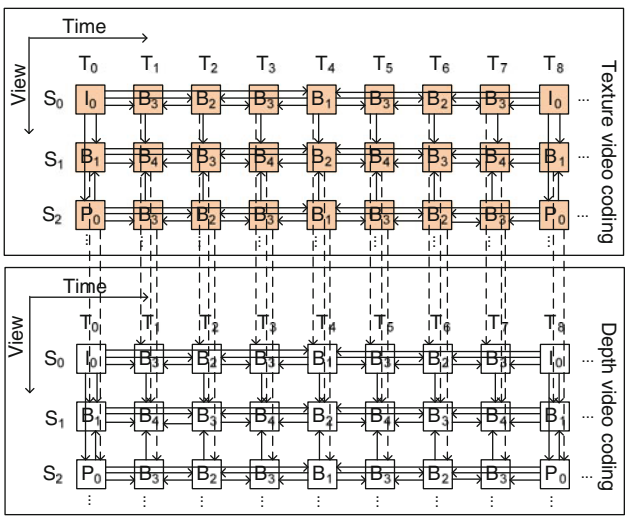

Fig. 1 Coding structure of the joint MVD coding method

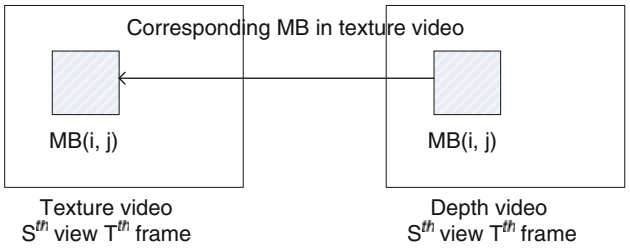

Fig. 2 Illustration of current MB in depth video and its corresponding MB in texture video

Table 1 Test conditions

Basis quantization parameter (bQP)	28, 32, 36, 40
GOP size	12
Number of reference frames	2
Search range	64
Max no. of Iterations bi-prediction search	4
Search range for iterations	8
Search mode	Fast
Number of frames to be encoded	49
Views to be encoded	0, 1, 2

Table 2 Statistical results of depth MBs which have the same best mode with their corresponding texture MBs (%)

bQP	Balloons	Cafe	Kendo	Average
28	65.20	50.26	50.33	55.26
32	72.55	53.87	56.64	61.02
36	77.92	56.76	63.28	65.99
40	81.27	59.58	69.67	70.17
Average	74.24	55.12	59.98	63.11

From Table 2, it can be observed that there are a larger number of MBs in depth video, which have the same optimal mode with their corresponding MBs in texture video. Statistical results show that the optimal modes in texture and depth match are from 50.33 to 81.27 %, 63.11 % on average. More than 60 % MBs in depth video have the same optimal mode with their corresponding MBs in texture video. If these matched MBs are determined early, significant encoding time will be saved.

Take a further step, if two MBs have similar motion activity in encoding process, they will have large probability of selecting the same optimal mode. Usually, if one MB is static, it will be encoded as DIRECT mode; in contrast, if one MB moves fast, it has a large probability to be encoded as B8 × 8 or INTRA mode; if the MB moves quite slow or is with medium motion activity, it will be encoded as B16 × 16, B16 × 8 or B8 × 16 [18, 19]. Based on these characteristics, motion activity can also be considered in the optimal mode selection to improve the mode prediction accuracy.

3 Proposed fast mode decision algorithm

3.1 Early mode decision based on the correlation between depth video and texture video

In natural video sequence, the motion trace of a moving object is usually successive [17]. The aim of ME is to find the best-matching block in the reference frame. Therefore, if the movement between current block and its best-matching block in the reference frame is larger, it means the MB moves fast. Otherwise, it indicates that this MB is with slow or medium motion activity. In this paper, the movement is defined by the initial search point and the final best search point of 16 × 16 MB in ME/DE process, and computed as:

$$d = \frac{1}{r} \sum_{(i,j),(m,n)\in\psi} [(i-m)^2 + (j-n)^2]^{1/2}, \quad (1)$$

where r represents the number of the best motion vectors for the current 16 × 16 MB; (i, j) denotes the motion vector (MV) of initial search point of 16 × 16 MB in ME/DE process; (m, n) is the MV of the final best search point of 16 × 16 MB in ME/DE process; ψ denotes the best reference frame in each reference frame list. We denote the movement of current MB in depth video as d_{d16}, and the movement of its corresponding MB in texture video as d_{t16}. Hence, the current depth MB is encoded as the mode of its corresponding MB in texture video if

$$d_{d16} = d_{t16}. \quad (2)$$

To evaluate the efficiency of the proposed algorithm, determination rate (DR) and hit rate (HR) are adopted, which are defined as

$$\begin{cases} S_{DR}(\mathbf{B}|\mathbf{A}) = N(\mathbf{B}|\mathbf{A})/N(\mathbf{A}) \times 100\,\%, \\ T_{HR}(\mathbf{A}|\mathbf{B}) = N(\mathbf{A}|\mathbf{B})/N(\mathbf{B}) \times 100\,\%, \end{cases} \quad (3)$$

where $S_{DR}(\mathbf{B}|\mathbf{A})$ and $T_{HR}(\mathbf{A}|\mathbf{B})$ denote the DR and HR, respectively; $N(\cdot)$ represents the number of total MBs of corresponding event, and the event **A** represents the selected mode of the encoded MB, **B** denotes the early mode decision condition. These two events are defined in the following specific early mode decision algorithm. **B**|**A** and **A**|**B** are two conditional events. DR represents the computational time saving of the proposed algorithm. If DR is large, more coding complexity could be reduced. HR indicates the best MB mode prediction accuracy by the proposed algorithm. If HR is large and close to 100 %, it means the best mode is correctly predicted and almost no R-D degradation would be caused.

To evaluate the efficiency of the proposed texture–depth correlation based mode decision strategy, **A** is the event that the MB in depth video has the same best mode selection as its corresponding MB in texture video, which is denoted as $M_d = M_t$; **B** represents the event that d of the MB in depth video is equal to d of its corresponding MB in texture video. In other words, **B** is the early mode decision condition which is $d_{d16} = d_{t16}$. Three multiview depth videos (Balloons, Kendo and Cafe) are tested. The test conditions are tabulated in Table 1. The detailed results of $S_{DR}(d_{d16} = d_{t16}|M_d = M_t)$ and $T_{HR}(M_d = M_t|d_{d16} = d_{t16})$ are listed in group P1 of Table 3.

From Table 3, it can be observed that DR is from 46.87 to 85.10 %, 67.17 % on average, which indicates 67.17 % depth MBs have the same mode as their corresponding MBs in texture video. Thus, these MBs can be early determined by the condition $d_{d16} = d_{t16}$. The HR holds from 87.86 to 96.84 %, 93.23 % on average. In other words, 93.23 % of the early determined depth MBs have correctly selected the best mode using the proposed mode decision strategy. These values demonstrate that the proposed texture–depth correlation based mode decision algorithm can work efficiently.

3.2 Fast mode decision based on motion prediction and CBP

For these depth MBs that do not satisfy to the texture–depth correlation based early mode decision condition, their optimal modes can be determined by the following strategies. It is well known that if one MB is in slow-motion area, it has a large probability to be encoded as DIRECT mode. Furthermore, in encoded video stream, there is a syntax element named CBP, which has six bits to indicate whether six 8×8 blocks (including four luma blocks and two chroma blocks) have non-zero coefficients [20, 21]. Figure 3 shows the correspondence between CBP bits and luma/chroma blocks in an MB. In the figure, each rectangle indicates an 8×8 block. Thus, there are 4 luma 8×8 blocks and 2 chroma blocks in an MB while its color format is YUV 4:2:0. Each bit in CBP indicates whether the corresponding 8×8 block in current MB has non-zero coefficients or not. If yes, the corresponding CBP bit is 1 (binary), otherwise, it is 0 (binary). In such case, all-zero block is with no residual and usually indicates the block is well predicted. Thus, if the CBP of current depth MB is zero after checking DIRECT mode, it indicates that the current depth MB is already well predicted and might not necessary to check other time-consuming modes. It is especially true for the static or regular moving regions. Therefore, the mode of the current depth MB, M_d, can be determined by

$$M_d = \begin{cases} DIRECT & if\ d_{d16} = 0\ \&\&\ CBP_{DIR} = 0, \\ Non - DIRECT & otherwise, \end{cases} \quad (4)$$

where d_{d16} represents the movement after B16 $\times$ 16 block ME/DE, and it is computed as Eq. (1), CBP_{DIR} denotes the CBP value of DIRECT of the current depth MB, && is an add operation and it means that the optimal mode of the depth MB is DIRECT when it can simultaneously meet these two requirements, $d_{d16} = 0$ and $CBP_{DIR} = 0$.

To evaluate the efficiency of the proposed early DIRECT mode decision algorithm, DR and HR defined in Eq. (3) are also used, where **A** is the event that the MB selects the DIRECT as its best mode, i.e., $M_d = DIRECT$; **B** represents

Table 3 Statistical results of determination rate and hit rate

Algorithm	bQP	Determination rate (%)				Hit rate (%)			
		Balloons	Kendo	Cafe	Average	Balloons	Kendo	Cafe	Average
P1	24	69.69	61.86	46.87	59.47	90.15	87.86	94.14	90.72
	28	74.41	67.38	52.92	64.90	94.26	89.88	96.15	93.43
	32	79.70	72.33	58.80	70.28	94.57	91.33	94.63	93.51
	36	85.10	75.64	61.28	74.01	96.84	92.14	96.80	95.26
	Average	77.23	69.30	54.97	67.17	93.96	90.30	95.43	93.23
P2	24	91.41	86.13	82.13	86.56	93.83	90.12	96.08	93.34
	28	92.66	84.54	85.65	87.62	96.54	90.79	97.29	94.87
	32	94.70	85.41	88.12	89.41	98.14	91.90	97.65	95.90
	36	96.36	86.25	90.29	90.97	99.03	94.82	97.50	97.12
	Average	93.78	85.58	86.55	88.64	96.89	91.91	97.13	95.31
P3	24	97.23	90.93	93.50	93.89	99.02	96.24	94.36	96.54
	28	98.61	95.15	93.99	95.92	99.32	96.41	95.85	97.19
	32	99.53	97.77	94.41	97.24	99.73	96.25	96.38	97.45
	36	99.85	99.10	95.48	98.14	99.82	98.14	96.69	98.22
	Average	98.81	95.74	94.35	96.30	99.47	96.76	95.82	97.35

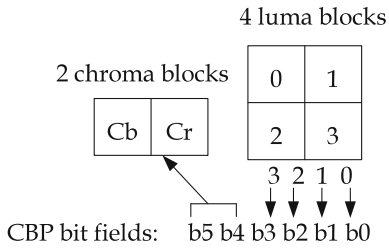

Fig. 3 An illustration of CBP bit format

the condition $d_{d16} = 0$ and $CBP_{DIR} = 0$. The test results of $S_{DR}((d_{d16} = 0 \;\&\&\; CBP_{DIR} = 0)|M_d = DIRECT)$ and $T_{HR}(M_d = DIRECT|(d_{d16} = 0 \;\&\&\; CBP_{DIR} = 0))$ are tabulated in group P2 of Table 3.

From group P2 of Table 3, it can be observed that there are from 82.13 to 96.36 %, 88.64 % on average, MBs which select DIRECT as their best mode can be determined early. We can also see that the HR of proposed early DIRECT mode decision algorithm is from 90.12 to 99.03 %, 95.31 % on average, which indicates sufficient high mode determination accuracy of using the proposed strategy. These values show that the proposed early DIRECT mode decision algorithm can work efficiently with intact R-D degradation.

In depth video coding, depth video is regarded as the *Y*-component and encoded by the color codec. The pixel value of chroma component is filled with consistent 128 and no residual will be caused. Thus, CBP bit values of two chroma blocks are always equal to 0, since Cb and Cr blocks are always all-zero block. Hence, when CBP value is equal to the value from the set $\Omega(\Omega = \{1, 2, 4, 8\})$, it represents that only one 8×8 luminance block has non-zero coefficients. Most blocks in the current depth MB can be well predicted by DIRECT, thus, the current MB is usually located in slow or simple motion area. In this situation, this depth MB has a large probability to be encoded as the mode from the set $\Theta = \{\mathbf{B}16 \times 16, \mathbf{B}16 \times 8, \mathbf{B}8 \times 16\}$. Hence, the mode from Θ can be selected as the best mode for the current depth MB if

$$CBP_{DIR} \in \Omega. \tag{5}$$

Three multiview depth video sequences are encoded to analyze the DR and HR defined in Eq. (3), where $\mathbf{A}$ is the event that the current depth MB is encoded as one mode of Θ, denoted as $M_d \in \Theta$; $\mathbf{B}$ is the mode decision condition that CBP of DIRECT belongs to Ω, i.e., $CBP_{DIR} \in \Omega$. The statistical results of $S_{DR}(CBP_{DIR} \in \Omega | M_d \in \Theta)$ and $T_{HR}(M_d \in \Theta | CBP_{DIR} \in \Omega)$ are tabulated in the group P3 of Table 3.

From group P3 of Table 3, it can be observed that there are from 90.93 to 99.85 %, 96.30 % on average, MBs which select $\mathbf{B}16 \times 16, \mathbf{B}16 \times 8$ or $B8 \times 16$ as their best mode can be determined early. We can also see that the proposed algorithm achieves the mode decision accuracy from 94.36 to 99.82 %, 97.35 % on average. These values demonstrate that this mode decision condition works perfectly.

If CBP_{DIR} is neither 0 nor Ω, it is for sure that two or more luma blocks have non-zero coefficient. It refers that the current depth MB can hardly be well predicted by large block-size mode. Thus, smaller block-size mode, such as $\mathbf{B}8 \times 8$, shall be further checked. Since INTRA modes consume quite little time in mode decision process, INTRA modes will be performed after checking set of $\mathbf{B}16 \times 16, \mathbf{B}16 \times 8, \mathbf{B}8 \times 16$ and $\mathbf{B}8 \times 8$ to trade off the computational complexity and R-D performance.

3.3 The overall algorithm

Based on the above analyses, the proposed fast mode decision algorithm is summarized and illustrated step-by-step as follows.

Step 1. Encode the current depth MB as $\mathbf{B}16 \times 16$, and compute the d according to Eq. (1), which is denoted as d_{d16}. If $d_{d16} = d_{t16}$, the current depth MB is encoded as the mode of its corresponding texture MB, go to Step 6; otherwise, go to Step 2.

Step 2. Encoded the current depth MB as DIRECT, and get its CBP as CBP_{DIR}. If $CBP_{DIR} = 0$ and $d_{d16} = 0$, the current depth MB is encoded as DIRECT mode, go to Step 6; otherwise, go to Step 3.

Step 3. If $CBP_{DIR} \in \Omega$, check the current depth MB with $\mathbf{B}16 \times 16, \mathbf{B}16 \times 8, \mathbf{B}8 \times 16$, and go to Step 5; otherwise, go to Step 4.

Step 4. All $\mathbf{B}8 \times 8$ modes are performed for the current depth MB, go to Step 5.

Step 5. All INTRA modes are performed for the current depth MB, go to Step 6.

Step 6. Choose the best mode among all tested modes based on R-D cost comparison. Go back to Step 1 to process the next depth MB.

4 Experimental results

To evaluate the efficiency of the proposed algorithm, MVC reference software JMVC8.0 [22] is used as the software platform. The test conditions are listed in Table 1. The hardware platform is Intel Core 2 Duo CPU E5800 @ 3.16GHz and 3.17GHz, 4.00GB RAM with Microsoft Windows 7 64-bit operating system.

We compare the coding performance of the proposed algorithm with two recent fast mode decision algorithms, Shen [9] or Peng [14], in terms of peak signal-to-noise ratio (PSNR), bit rate (BR) and total encoding CPU time. Six

Table 4 Summary of encoding results, even views

Sequence	bQP	Shen vs. JMVC	Peng vs. JMVC			Proposed vs. JMVC		
		ΔPSNR/ΔBR/TS	ΔPSNR	ΔBR	TS	ΔPSNR	ΔBR	TS
Balloons	28		–0.325	1.22	–70.89	–0.113	0.19	–68.54
	32		–0.201	–0.35	–69.39	–0.078	–0.20	–71.96
	36	0/0/0	–0.106	–0.72	–66.73	–0.073	–0.22	–69.61
	40		–0.057	–1.18	–66.61	–0.038	–0.34	–73.24
	Average		–0.172	–0.26	–68.41	–0.076	–0.14	–70.84
	BDPSNR/ BDBR	0/0/0	–0.149/3.76			–0.068/1.65		
Breakdancer	28		–0.140	0.97	–60.49	–0.043	0.55	−50.87
	32		–0.078	–0.09	–59.10	–0.061	–0.05	−56.96
	36	0/0/0	–0.045	–0.40	–49.69	–0.056	–0.24	−59.94
	40		–0.022	–0.26	–47.29	–0.037	–0.10	−64.53
	Average		–0.071	0.06	–54.14	–0.049	0.04	−58.08
	BDPSNR/ BDBR	0/0/0	–0.064/1.34			–0.052/1.07		
Cafe	28		–0.315	3.39	–63.91	–0.076	0.93	–64.25
	32		–0.257	2.35	–61.36	–0.056	0.42	–66.93
	36	0/0/0	–0.267	1.31	–59.61	–0.013	0.76	–63.95
	40		–0.214	1.04	–58.17	–0.216	0.52	–69.20
	Average		–0.263	2.02	–60.76	–0.090	0.66	–66.08
	BDPSNR/ BDBR	0/0/0	–0.414/4.34			–0.124/1.19		
Champange	28		–0.066	–0.12	–75.76	–0.031	–0.02	–74.55
	32		–0.044	–0.59	–74.09	–0.041	–0.51	–76.54
	36	0/0/0	–0.019	–0.45	–73.68	–0.043	–0.53	–77.01
	40		–0.006	–0.32	–75.57	–0.069	–0.53	–78.86
	Average		–0.034	–0.37	–74.78	–0.046	–0.40	–76.74
	BDPSNR/ BDBR	0/0/0	–0.023/1.05			–0.034/1.47		
Doorflowers	28		–0.194	1.62	–64.82	–0.050	0.25	–63.35
	32		–0.134	–0.25	–66.38	–0.049	–0.18	–66.57
	36	0/0/0	–0.066	–0.70	–66.87	–0.033	–0.46	–71.24
	40		–0.028	–0.74	–64.63	–0.023	–0.34	–72.45
	Average		–0.106	–0.02	–65.68	–0.039	–0.18	–68.40
	BDPSNR/ BDBR	0/0/0	–0.097/3.71			–0.041/1.48		
Kendo	28		–0.188	–0.11	–73.36	–0.131	–0.02	–57.81
	32		–0.130	–0.61	–69.89	–0.236	–0.07	–61.22
	36	0/0/0	–0.073	–0.34	–66.93	–0.254	–0.68	–64.51
	40		–0.016	–0.64	–64.10	–0.279	–1.01	–68.20
	Average		–0.102	–0.43	–68.57	–0.225	–0.45	–62.94
	BDPSNR/ BDBR		–0.069/1.11			–0.210/3.39		
Average ΔPSNR/ΔBR/TS		0/0/0	–0.125/0.17/–64.26			–0.087/–0.08/–67.18		
Average BDPSNR/BDBR		0/0/0	–0.136/2.55			–0.088/1.71		

multiview depth video sequences (Balloons, Breakdancer, Cafe, Champange, Doorflowers and Kendo) are tested. The experimental results are compared and summarized in Tables 4 and 5. In these two tables, △PSNR, △BR and TS represent the PSNR, BR and total encoding time change, respectively. They are defined as

Table 5 Summary of encoding results, odd views

Sequence	bQP	Shen vs. JMVC			Peng vs. JMVC			Proposed vs. JMVC		
		△PSNR	△BR	TS	△PSNR	△BR	TS	ΔPSNR	△BR	TS
Balloons	28	–0.233	–0.74	–60.72	–0.341	2.05	–74.23	–0.138	0.19	–72.01
	32	–0.254	–2.36	–70.16	–0.251	–0.05	–72.42	–0.132	0.27	–70.09
	36	–0.204	–2.69	–72.29	–0.112	–0.42	–68.45	–0.355	–0.14	–68.84
	40	–0.148	–3.55	–75.86	–0.014	–0.55	–64.77	–0.064	–0.94	–70.91
	Average	–0.210	–2.34	–69.76	–0.180	0.26	–69.97	–0.172	–0.16	–70.46
	BDPSNR/ BDBR	–0.089/1.80			–0181/3.69			–0.207/4.12		
Breakdancer	28	–0.181	3.01	–22.34	–0.255	5.28	–67.71	–0.036	2.49	–60.57
	32	–0.234	–4.37	–29.35	–0.157	1.55	–60.88	–0.058	1.23	−61.64
	36	–0.326	6.04	–44.39	–0.115	–0.16	–56.47	–0.053	1.43	−65.34
	40	–0.393	9.28	–54.77	–0.054	0.07	–54.3	–0.026	0.49	−69.77
	Average	–0.284	3.49	–37.71	–0.145	1.69	–59.84	–0.043	1.41	–64.33
	BDPSNR/ BDBR	–0.540/11.68			–0.204/4.52			–0.113/2.44		
Cafe	28	–0.131	–1.46	–34.91	–0.328	4.73	–62.76	–0.109	1.10	–67.66
	32	–0.120	–3.49	–40.65	–0.364	0.79	–59.30	–0.068	–0.80	−68.55
	36	–0.267	–3.80	–41.75	–0.356	0.07	–59.06	–0.100	0.79	–69.24
	40	–0.401	–4.40	–48.99	–0.301	0.05	–60.20	–0.211	1.37	–71.56
	Average	–0.230	–3.29	–41.58	–0.337	1.41	–60.33	–0.122	0.62	–69.25
	BDPSNR/ BDBR	0.191/−1.64			–0.490/4.47			–0.139/1.34		
Champange	28	–0.100	0.69	–63.19	–0.051	0.46	–79.74	–0.031	0.20	–77.12
	32	–0.160	0.17	–71.32	−0.029	–0.04	–77.68	–0.046	–0.23	–78.23
	36	–0.236	–1.58	–77.04	–0.032	0.42	–75.53	–0.068	–0.78	–78.87
	40	–0.255	–1.81	–78.38	–0.014	–0.18	–75.08	–0.141	–1.52	–78.28
	Average	–0.188	–0.63	–72.48	–0.032	0.17	–77.01	–0.072	–0.58	–78.13
	BDPSNR/ BDBR	–0.170/8.24			–0.034/1.61			–0.051/2.32		
Doorflowers	28	–0.121	–1.32	–51.89	–0.183	1.82	–74.59	–0.054	0.59	–65.65
	32	–0.154	–2.34	–61.69	–0.132	0.91	–63.56	–0.054	0.51	–68.62
	36	–0.191	–4.6	–72.11	–0.047	0.07	–59.57	–0.004	0.63	–70.01
	40	–0.169	–4.09	–78.5	–0.072	0.75	–55.68	–0.033	1.62	–75.21
	Average	–0.159	–3.09	–66.05	–0.109	0.89	–63.35	–0.036	0.84	–69.87
	BDPSNR/ BDBR	–0.085/3.52			–0.115/4.88			–0.049/2.03		
Kendo	28	–0.163	–0.25	–39.3	–0.187	–0.79	–71.80	–0.180	–0.40	–65.53
	32	–0.239	–0.36	–49.81	–0.098	–0.90	–68.94	–0.185	–0.39	–67.42
	36	–0.292	–1.79	–59.29	–0.084	–0.64	–68.39	–0.198	0.07	–69.14
	40	–0.385	–3.46	–69.54	–0.020	–1.72	–63.94	–0.288	–0.28	–67.41
	Average	–0.270	–1.47	–54.49	–0.097	–1.01	–68.27	–0.213	–0.25	–67.38
	BDPSNR/ BDBR	–0.155/1.96			–0.021/0.35			–0.230/3.16		
Average △PSNR/△BR/TS		–0.223/–1.22/–57.01			–0.150/0.57/–65.46			–0.110/0.31/–69.90		
Average BDPSNR/BDBR		–0.141/4.26			–0.174/3.25			–0.132/2.57		

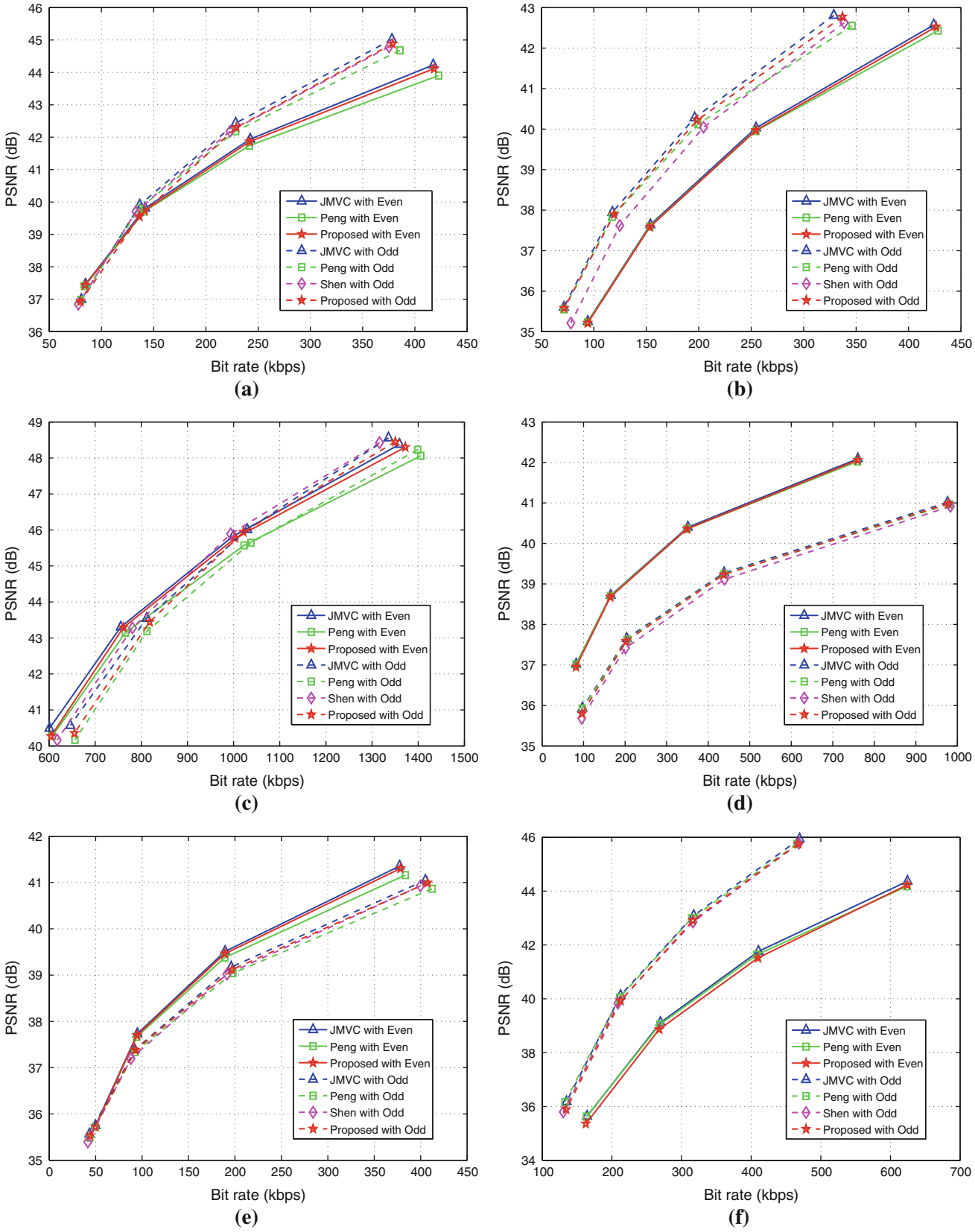

Fig. 4 R-D Curves **a** Balloons. **b** Breakdancers. **c** Cafe. **d** Champange. **e** Doorflowers. **f** Kendo

$$\begin{cases} \triangle PSNR = PSNR_p - PSNR_o(dB), \\ \triangle BR = \frac{BR_p - BR_o}{BR_o} \times 100\,\%\,(\%), \\ TS = \frac{T_p - T_o}{T_o} \times 100\,\%\,(\%), \end{cases} \tag{6}$$

where the subscript p represents our proposed algorithm, Shen and Peng. o denotes the JMVC8.0. BDPSNR and BDBR are computed according to [23].

Table 4 shows the summary encoding results of even views. It can be observed that Shen's method do not optimize the even views. Peng's method can reduce the computational complexity from 47.29 to 75.76 %, 64.26 % on average; meanwhile, the PSNR degrades from 0.006 to 0.325 dB, 0.125 dB on average; and the BR changes from −1.18 to 3.39 %, 0.17 % on average. The average BDPSNR and BDBR between Peng's method and original JMVC8.0 are −0.136 dB and 2.55 %, respectively. The proposed algorithm can save the encoding time from 50.87 % to 78.86 %, 67.18 % on average, when the PSNR degrades from 0.013 to 0.279 dB, 0.087 dB on average, and BR changes from −1.01 to 0.93 %, −0.08 % on average. The average BDPSNR and BDBR between the proposed method and original JMVC8.0 are −0.088 dB and 1.71 %, respectively. Compared to Peng's method, the proposed algorithm has a better coding performance, PSNR increases 0.038 dB, BR decreases 0.25 %, and total encoding time saves about 2.92 % more; BDPSNR increases 0.048 dB, BDBR decreases 0.84 %.

The summary encoding results of odd views are tabulated in Table 5, it can be observed that the Shen's method reduces the computational complexity from 22.34 to 78.38 %, 57.01 % on average; meanwhile, the PSNR degrades from 0.100 to 0.393 dB, 0.223 dB on average, and the BR changes from −4.37 to 9.28 %, −1.22 % on average. The average BDPSNR and BDBR between Shen's method and original JMVC8.0 are −0.141 dB and 4.26 %, respectively. However, the PSNR of Breakdancer and Kendo degrades dramatically. This is because, these two sequences are with fast motion and the collocated MBs in neighboring views are obtained by GDVs, which is not accurate. Peng's method can reduce the encoding time from 54.30 to 79.74 %, 65.46 % on average, meanwhile, the PSNR degrades from 0.014 to 0.364 dB, 0.150 dB on average, and the BR changes from −1.72 to 5.28 %, 0.57 % on average. The average BDPSNR and BDBR between Peng's method and original JMVC8.0 are −0.174 dB and 3.25 %, respectively. The proposed fast mode decision algorithm can reduce the computational complexity from 60.57 to 78.87 %, 69.90 % on average. The PSNR degrades from 0.004 to 0.355 dB, 0.110 dB on average. The BR changes from −0.94 to 2.49 %, 0.31 % on average. The average BDPSNR and BDBR between the proposed method and original JMVC8.0 are −0.132 dB and 2.57 %, respectively. It can be observed that the proposed algorithm achieves the best R-D performance. Compared to Shen's method and Peng's method, 12.89 and 4.44 % computational complexity are further reduced by the proposed algorithm.

To demonstrate the overall R-D performance of the proposed algorithm, we give the R-D curves of all six multiview depth video sequences (Balloons, Breakdancer, Cafe, Champange, Doorflowers and Kendo) with even views and odd views in Fig. 4. It can be observed that the proposed algorithm achieves similar R-D performance as compared with the JMVC and Peng's method, and it is better than the Shen's method.

In addition to HBP prediction structure, low-latency prediction structure, IPPP structure, is also implemented to verify the adaptation of the proposed algorithm. Since JMVC8.0 does not support the IPPP structure, H.264/AVC reference software JM 14.1 [24] is adopted as the software platform. Four QPs (28, 32, 36 and 40) and four sequences (Balloons, Breakdancer, Champange and Doorflowers) are tested. Similar to HBP prediction structure, similar results can be found for IPPP structure. It indicates that the proposed algorithm is not only effective for high-efficiency HBP prediction structure, but also efficient for low-latency IPPP structure.

5 Conclusion

In this paper, we propose a fast mode decision algorithm for multiview depth video coding, which is based on the mode selection correlations between depth video and its corresponding texture video, motion prediction and CBP. Experimental results show that the proposed algorithm can achieve a quite promising coding performance in terms of R-D performance and computational complexity saving. In addition, the proposed algorithm can be applied not only to odd views but also to even views. Compared to other two algorithms, the proposed algorithm is more suitable for real-time applications.

Acknowledgments This work was supported in part by the Natural Science Foundation of China under Grants 61272289, 61102088 and in part by the Guangdong Provincial Nature Science Foundation under Grant S2012010008457, Shenzhen Emerging Industries of Strategic Basic Research Project under Grant JCYJ201206171 51719115.

References

1. Kauff P., Atzpadin N., Fehn C., Müller M., Schreer O., Smolic A., Tanger R.: Depth map creation and image-based rendering for advanced 3DTV services provding interoperability and scalability. Signal Process. Image Commun. **22**(2), 217–234 (2007)

2. Mueller, K., Merkel, P., Smolic, A., Wiegand, T.: Multiview coding using AVC, ISO/IEC JTC1/SC29/WG11, Document M12945, Bangkok, Thailand (2006)
3. Vetro A., Wiegand T., Sullivan G.J.: Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. Proc. IEEE **9**(4), 626–664 (2011)
4. Pan, Z., Kwong, S., Xu, L., Zhang, Y., Zhao, T.: Predictive and distribution-oriented fast motion estimation for H.264/AVC. J. Real Time Image Process. doi:10.1007/s11554-012-0264-7
5. Nieto M., Salgado L., Cabrera J., García N.: Fast mode decision on H.264/AVC baseline profile for real-time performance. J. Real Time Image Proc. **3**(1–2), 61–75 (2008)
6. Hu S., Zhao T., Wang H., Kwong, S.: Fast inter-mode decision based on rate-distortion cost characteristics. Proc. PCM 10 **2**, 145–155 (2010)
7. Zhao T., Wang H., Kwong S., Kuo C.-C. Jay: Fast mode decision based on mode adaptation. IEEE Trans. Circuits Syst. Video Technol. **20**(5), 697–704 (2010)
8. Zhao T., Kwong S., Wang H., Kuo C.-C. J.: H.264/SVC mode decision based on optimal stopping theory, IEEE Trans. Image Process. **21**(5), 2607–2618 (2012)
9. Shen L., Liu Z., Liu S., Zhang Z., An, P.: Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding. IEEE Trans. Broadcast. **55**(4), 761–766 (2009)
10. Shen L., Liu Z., Yan T., Zhang Z., An, P.: Early SKIP mode decision for MVC using inter-view correlation. Signal Process. Image Commun. **25**(2), 88–93 (2010)
11. Zhang, Y., Kwong, S., Jiang, G., Wang, X., Yu, M.: Statistical early termination model for fast mode decision and reference frame selection in multiview video coding. IEEE Trans. Broadcast. **58**(1), 10–23 (2012)
12. Yoon, D.-H., Ho, Y.-S.: Fast mode decision algorithm for depth codingin 3D video systems using H.264/AVC. LNCS (7088) 25–35 (2012)
13. Micallef, B.W., Debono, C.J., Farrugia, R.A.: Fast inter-mode decision in multi-view video plus depth coding, Proc. PCS'12, 113–116 (2012)
14. Peng, Z., Yu, M., Jiang, G., Shao, F., Zhang, Y., Yang, Y.: Fast macroblock mode selection algorithm for multiview depth video coding. Chinese Optics Lett. **8**(2), 151–154 (2010)
15. Zhang, Q., An, P., Zhang, Y., Shen, L., Zhang, Z.: Low complexity multiview video plus depth coding. IEEE Trans. Consum. Electr. 1857–1865 (2011)
16. Merkle, P., Smolic, A., Müller, K., Wiegand, T.: Efficient prediction structure for multi-view video coding, IEEE Trans. Circuits Syst. Video Technol. **17**(11), 1461–1473 (2007)
17. Chen, Z., Xu, J., He, Y., Zheng, J.: Fast integer-pel and fractional-pel motion estimation for H.264/AVC. Journal of visual communication and image representation **17**(2), 264–290 (2006)
18. Pan, Z., Kwong, S.: A fast Inter-Mode decision scheme based on luminance difference for H.264/AVC. Proc. ICSSE'11, 260–263 (2011)
19. Zeng, H., Ma, K.-K., Cai, C.: Fast mode decision for multiview video coding using mode correlation. IEEE Trans. Circuits Syst. Video Technol. **21**(11), 1659–1666 (2011)
20. ITU-T and ISO/IEC JTC 1: Advanced video coding for generic audiovisual services. ITU-T Recommedation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), (2010)
21. Chen, B.-Y., Yang, S.-H.: Using H.264 Coded block patterns for fast inter-mode selection. Proc. ICME'08, 721–724 (2008)
22. Vetro, A., Pandit, P., Kimata, H., Smolic, A., Wang, Y.-K.: Joint Draft 8.0 on Multiview Video Coding. Document JVT-AB204, 28th Meeting, Hannover, DE (2008)
23. Bjontegaard, G.: Calculation of average PSNR differences between RD-curves. Document VCEG-M33, Thirteenth Meeting, Austin, Texas, USA (2001)
24. JVT H.264/AVC reference software version JM14.1. http://iphome.hhi.de/suehring/tml/download/. Accessed 17 Dec 2012

Author Biographies

Zhaoqing Pan received his B.S. degree in Computer Science and Technology with honors from Yancheng Normal University, Yancheng, China, in June 2009. Currently, he is a Ph.D. candidate in the Department of Computer Science at the City University of Hong Kong, Kowloon, Hong Kong. His research interests include motion estimation and mode decision in video coding.

Yun Zhang received the B.S. and M.S. degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. From 2009 to 2012, he was a Postdoc Researcher and Visiting Researcher with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. In 2010, he joined in the Shenzhen Institutes of Advanced Technology (SIAT), CAS, where he serves as an Associate Professor since 2012. His research interests are 3D video coding, 3D video perception and content-based video processing.

Sam Kwong received the B.S. and M.S. degrees in electrical engineering from the State University of New York at Buffalo in 1983, the University of Waterloo, Waterloo, ON, Canada, in 1985, and the Ph.D. degree from the University of Hagen, Germany, in 1996. From 1985 to 1987, he was a Diagnostic Engineer with Control Data Canada. He joined Bell Northern Research Canada as a Member of Scientific Staff. In 1990, he became a Lecturer in the Department of Electronic Engineering, City University of Hong Kong, where he is currently a Professor in the Department of Computer Science. His research interests are video and image coding and evolutionary algorithms.