J Real-Time Image Proc (2016) 12:575–582
DOI 10.1007/s11554-015-0527-1

ORIGINAL RESEARCH PAPER

Adaptive mode decision for multiview video coding based on macroblock position constraint model

Yue Li[1] · Gaobo Yang[1] · Yapei Zhu[2] · Can Liu[2] · Kai Liu[3]

Received: 8 April 2015 / Accepted: 15 August 2015 / Published online: 28 August 2015

Abstract Multiview video coding (MVC) exploits mode decision, motion estimation and disparity estimation to achieve high compression ratio, which results in an extensive computational complexity. This paper presents an efficient mode decision approach for MVC using a macroblock (MB) position constraint model (MPCM). The proposed approach reduces the number of candidate modes by utilizing the mode correlation and rate distortion cost (RD cost) in the previously encoded frames/views. Specifically, the mode correlations both in the temporal-spatial domain and the inter-view are modeled with MPCM. Then, MPCM is exploited to select the optimal prediction direction for the current encoding MB. Finally, the inter mode is early determined in the optimal prediction direction. Experimental results show that the proposed method can save 86.03 % of encoding time compared with the exhaustive mode decision used in the reference software of joint multiview video coding, with only 0.077 dB loss in Bjontegaard delta peak signal-to-noise ratio (BDPSNR) and 2.29 % increment of the total Bjontegaard delta bit rate (BDBR), which is superior to the performances of state-of-the-art approaches.

Keywords Multiview video coding · Mode decision · Macroblock position constraint model · H.264/AVC

✉ Gaobo Yang
yanggaobo@hnu.edu.cn

[1] School of Information Science and Engineering, Hunan University, Changsha 410082, China

[2] Faculty of Physics and Electronic Information Science, Hengyang Normal University, Hengyang 421002, China

[3] North University of China, Taiyuan 030051, China

1 Introduction

Multi-view video refers to a set of temporally synchronized videos captured at the same scene by multiple cameras from different viewpoints [1]. Compared with the single-view video, multi-view video provides more interactivity and realistic experience for viewers, which has great potential in new video applications such as Free-viewpoint Television (FTV) and Three-dimensional Television (3DTV) [2, 3]. To facilitate the research of multi-view video coding (MVC), Joint Video Team (JVT), which was composed of experts from both ISO/IEC MPEG and ITU-T Video Coding Experts Group (VCEG) [4, 5], developed reference software of Joint Multiview Video Coding (JMVC) on the basis of H.264/AVC video coding standard. In JMVC, hierarchical B picture (HBP) structure achieves higher coding efficiency compared with the straightforward solution of independently encoding each view with H.264/AVC. Figure 1 shows the HBP architecture in JMVC, where the arrow denotes the direction of reference frame. All the views are divided into two classes: even views and odd views. The even views (V0, V2, V4 and V6) use variable block-size motion estimation (ME) technique to exploit the spatial-temporal correlation. Meanwhile the odd views (V1, V3, V5 and V7) adopt a new variable block-size disparity estimation (DE) technique which exploits the inter-view correlation to improve the coding efficiency. Because the process of ME and DE is separately and repeatedly performed for each MB, the computational complexity of mode decision is very intensive.

To reduce the computational complexity, several fast mode decision approaches are presented in the literature. They can be categorized into two classes. The first class is to early terminate the SKIP/DIRECT mode decision process [6–8]. If the SKIP/DIRECT mode is considered as the

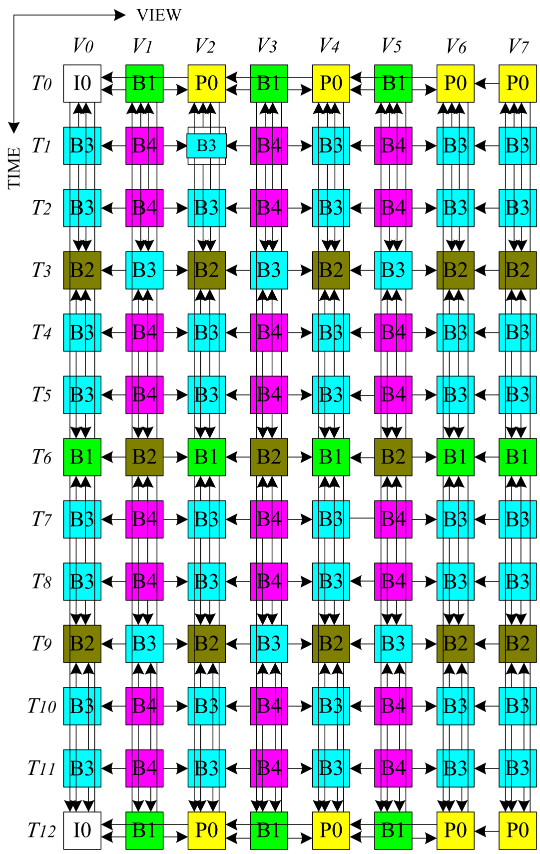

Fig. 1 Architecture of hierarchical B picture (HBP) in JMVC

best mode, the remaining modes will not be checked. This class of approaches is effective for video without motion or just slow-motion content. In this case, the best mode of most MBs in the motionless or slow-motion video is actually SKIP/DIRECT mode. The second class reduces the number of candidate modes to be checked in the mode decision process [9–15]. Specifically, for those RD cost-based approaches [9–11], the mode decision is early terminated by threshold, where the threshold is computed from the RD cost of previously coded frames. Furthermore, the motion vector difference based motion cost is exploited to reduce the computation of Inter mode decision process [12]. In addition, low-complexity mode decision methods are proposed for MVC based on inter-view mode correlation [13–15]. It is apparent that the mode correlations in the temporal and spatial domain are not fully exploited in these approaches. To further speed up the process of mode decision, an adaptive mode decision approach is proposed for MVC in this paper. It is based on a novel macroblock position constraint model, which exploits the temporal-spatial and inter-view mode correlations. The proposed approach uses the mode correlation and the RD cost inferred from the previously coded MBs simultaneously, which leads to a significant reduction of candidate modes for each MB.

The rest of the paper is organized as follows: The motivations and preliminary analysis are presented in Sect. 2 for candidate modes. The proposed fast mode decision algorithm is presented Sect. 3. The experimental results and analysis are given in Sect. 4. We conclude this paper in Sect. 5.

2 Motivations and preliminary analysis

2.1 Observation and motivation

JMVC supports 11 candidate modes, which are the same with H.264/AVC. These prediction modes are SKIP/DIRECT, inter16 $\times$ 16, inter16 $\times$ 8, inter8 $\times$ 16, inter8 $\times$ 8, inter8 $\times$ 4, inter4 $\times$ 8, inter4 $\times$ 4, intra16 $\times$ 16, intra8 $\times$ 8 and intra4 $\times$ 4. To facilitate the following mode decision process, all these candidate modes are firstly divided into seven sub-classes, and each class is assigned with a value from 0 to 6. The classification is summarized in Table 1.

In order to get the mode values of MBs in B frames of auxiliary view (view 1 in Fig. 1), extensive experiments have been conducted. The test sequences used for experiments are listed in Table 2. And the test conditions for each sequence are as follows: hierarchical B picture (HBP) structure is used, and RDO and CABAC entropy coding are enabled. The GOP length, quantization parameter (QP) and search range are 12, 26 and 96, respectively.

Firstly, in order to explore the mode correlation in the spatial domain, the mode value of the current encoding MB is defined as V_C, and V_M as the maximum mode value of its four adjacent MBs. The absolute difference V_{AD} between V_C and V_M is computed as follows:

$$V_{\mathrm{AD}} = |V_C - V_M| \tag{1}$$

From Table 1, it can be observed that $V_{\mathrm{AD}} \subseteq \{0, 1, \ldots, 6\}$. The number of MBs with the same V_{AD} in each test sequence is described with a set A, and $A = \{N_{V_{\mathrm{AD}}=0}, N_{V_{\mathrm{AD}}=1}, N_{V_{\mathrm{AD}}=2}, N_{V_{\mathrm{AD}}=3}, N_{V_{\mathrm{AD}}=4}, N_{V_{\mathrm{AD}}=5}, N_{V_{\mathrm{AD}}=6}\}$.

Then, the percentage of each element in set A is calculated as:

$$P_{V_{\mathrm{AD}}=i} = \frac{N_{V_{\mathrm{AD}}=i}}{\sum_{i=0}^{6} N_{V_{\mathrm{AD}}=i}} \times 100\,\% \tag{2}$$

The experimental results of $P_{V_{\mathrm{AD}}=i}$ are shown in Table 3. The larger the $P_{V_{\mathrm{AD}}=0}$, the stronger the mode correlations in the spatial domain. It can be observed from Table 3 that $P_{V_{\mathrm{AD}}=0}$ are 68.30 % in average, which is the largest one.

Table 1 Modes and their corresponding values

Values (n)	0	1	2	3	4	5	6
Mode(s)	DIRECT	Inter16 × 16	Inter16 × 8 Inter8 × 16	Inter8 × 8	Inter8 × 4 Inter4 × 8	Inter4 × 4	Intra16 × 16 Intra8 × 8 Intra4 × 4

Table 2 Multi-view video sequences

Sequences	Resolution	Frame rate (fps)	Features
Flamenco1	320 × 240	30	Slow motion
Race1	640 × 480	30	Global fast motion
Exit	640 × 480	25	Slow motion
Vassar	640 × 480	25	Larger static background
Ballroom	640 × 480	25	Complex texture
Door flowers	1024 × 768	16.7	Slow motion
Dog	1280 × 960	30	Slow motion

Table 3 Statistical analysis of the correlation between the current MB and its adjacent MBs

Sequences	$P_{V_{AD}=0}$	$P_{V_{AD}=1}$	$P_{V_{AD}=2}$	$P_{V_{AD}=3}$	$P_{V_{AD}=4}$	$P_{V_{AD}=5}$	$P_{V_{AD}=6}$
Flamenco1	52.41	35.52	7.95	0.83	1.36	1.21	0.72
Exit	72.90	19.16	3.18	0.24	1.72	1.71	1.09
Race1	54.83	39.39	3.13	0.35	0.93	0.70	0.67
Vassar	75.94	16.26	2.06	0.01	2.40	2.04	1.29
Ballroom	62.29	31.09	3.74	0.06	1.07	1.08	0.67
Door flowers	82.43	13.88	2.07	0.14	0.48	0.43	0.57
Dog	77.33	17.62	3.78	0.09	0.44	0.36	0.38
Average	68.30	24.70	3.70	0.25	1.20	1.08	0.77

That means the absolute difference between the current MB's mode and the maximum value in its four adjacent MBs' modes is more likely to be 0. For example, if the maximum value in these four adjacent MBs's modes is 1, the value of the current MB' mode is more likely to be smaller or equal to 1. That is, the mode decision of the current MB would be early terminated after checking DIRECT and Inter16 × 16. In this way, the process of mode decision can be early terminated to save the encoding time.

Secondly, in the multi-view video sequences, there exists motion-less or slow-motion scenes which results in high correlation in the temporal direction. And as two neighboring cameras capture the same content at different viewpoints, there exists correlation between inter-views. Motivated by these facts, the correlation between the current MB and its corresponding MBs in the reference temporal and inter-view directions is analyzed. The percentage of current MB is defined as P_S which has the same mode to its corresponding MBs in the reference temporal and inter-view directions. The distribution of P_S from experiments is shown in Table 4. It is apparent that P_S occupies 87.34 % on average. It implies that the current MB tends to choose the same mode to its corresponding MBs in the reference temporal or inter-view direction. In this way, it is possible to speed the entire mode decision process.

Table 4 Percentage of current MB which has the same mode with its corresponding MBs in the test frames of V1

Sequences	P_S	$1 - P_S$
Flamenco1	75.75	24.25
Exit	87.27	12.73
Race1	81.67	18.33
Vassar	95.45	4.55
Ballroom	84.56	15.44
Door flowers	95.76	4.24
Dog	90.94	9.06
Average	87.34	12.66

Thirdly, since DIRECT mode occupies less encoding time compared with other modes, it is expected that most MBs in the test sequence will choose DIRECT modes as the optimal mode. Table 5 shows the distribution of the optimal mode in the B frames of V1 from experiments. It is shown that the DIRECT mode occupies 72.14 % on

Table 5 Distribution of the optimal mode resulted from the test frames of V1

Sequences	DIRECT	Other modes
Flamenco1	62.27	37.73
Exit	76.83	23.17
Race1	55.30	44.7
Vassar	79.48	20.52
Ballroom	64.84	35.16
Door flowers	87.69	12.31
Dog	78.59	21.41
Average	72.14	27.86

average. Especially for the *Door Flowers* sequence with less motion, the percentage of DIRECT mode is about 87.69 %. It implies that the DIRECT mode should be checked firstly to provide a chance for early terminating the process of mode decision.

2.2 Macroblock position constraint model

According to the analysis in Sect. 2.1, the mode of an MB in current frame/view is similar to the mode of the adjacent MBs and the corresponding MBs in the reference frame/view. Motivated by this observation, we establish the MB position constraint model (MPCM) to predict the optimal mode of the current encoding MB. The MPCM is illustrated in Fig. 2. Note that the corresponding MB of the current encoding MB locates at the same position in the nearest reference frame.

In Fig. 2, the modes of the current encoding MB and its corresponding MBs in the temporal and inter-view are donated by four variables: (1) $M_{V,T}$ denotes the mode of the current encoding MB at the time T in the view V. (2) $M_{V,T-1}$ denotes the mode of the corresponding coded MB at the time $T-1$ in the view V. (3) $M_{V-1,T}$ denotes the mode of the corresponding coded MB at the time T in the view $V-1$. (4) $M_{V-1,T-1}$ denotes the mode of the corresponding coded MB at the time $T-1$ in the view $V-1$. Based on these four variables, two functions are defined as Eqs. (3) and (4):

$$\Delta M_V = M_{V,T-1} - M_{V-1,T-1} \tag{3}$$

$$\Delta M_T = M_{V-1,T} - M_{V-1,T-1} \tag{4}$$

where ΔM_V is the mode change between the MB in the current view V and the corresponding coded MB in the reference view $V-1$. And ΔM_T is the mode change between the MB at the current time T and the corresponding coded MB at the time $T-1$.

Make comparison between ΔM_V and ΔM_T. If ΔM_V is larger than ΔM_T , that is, the mode change in the inter-view is more obvious than that in the temporal direction, the current MB at time T in view V tends to code with the similar mode as the corresponding MB at time $T-1$ in the view V. And the mode of the current MB $M_{V,T}$ is constrained by the MPCM in the temporal direction, which is defined as Eq. (5):

$$M_{V,T} = M_{V,T-1} + \Delta M_T \tag{5}$$

Otherwise, the mode of the current MB $M_{V,T}$ is constrained by the MPCM in the inter-view direction, which is defined as Eq. (6):

$$M_{V,T} = M_{V-1,T} + \Delta M_V. \tag{6}$$

2.3 Best mode prediction direction for MB in MPCM

From Table 4, it can be seen that the current MB tends to choose the same mode to its corresponding MBs in the reference temporal or inter-view direction. To early determine the encoding mode without exhaustive search for the current MB, it is necessary to determine the optimal mode prediction direction.

To determine the mode prediction direction of the current encoding MB in MPCM, the mode of the corresponding coded MB and its eight neighbor MBs at time $T-1$ or T in the view $V-1$ or V are used. The mode similarities in the temporal and inter-view (T_{ms}, V_{ms}) are respectively defined as

$$T_{\mathrm{ms}} = \sum_{i=0}^{8} \left| M_{V-1,T_{\mathrm{Mb_i}}} - M_{V-1,T-1_{\mathrm{Mb_i}}} \right| \tag{7}$$

$$V_{\mathrm{ms}} = \sum_{i=0}^{8} \left| M_{V,T-1_{\mathrm{Mb_i}}} - M_{V-1,T-1_{\mathrm{Mb_i}}} \right| \tag{8}$$

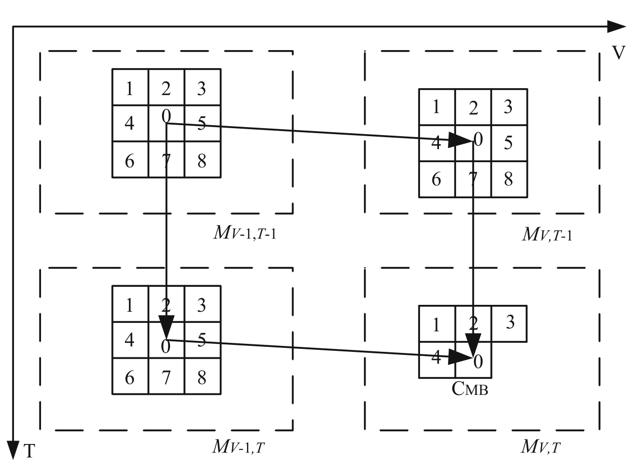

Fig. 2 Macroblock position constraint model

where $M_{V-1,T_{\mathrm{Mb_i}}}$, $M_{V,T-1_{\mathrm{Mb_i}}}$ and $M_{V-1,T-1_{\mathrm{Mb_i}}}$ in which $i \subseteq \{0,1,2,\ldots8\}$ represent the mode of the coded MB and its corresponding MBs at time $T-1$ or T in the view $V-1$ or V, respectively.

For the current MB, if T_{ms} is smaller than V_{ms}, that is, the mode change in the inter-view is more obvious than the mode change in the temporal direction, and the temporal direction is considered to be the optimal mode prediction direction. Otherwise the inter-view direction is selected as the optimal mode prediction direction.

3 Proposed adaptive early mode decision based on the MPCM

3.1 Early DIRECT mode using two-stage threshold

Based on the analysis in the Sect. 2, it is confirmed that DIRECT mode should be checked firstly to provide a chance for terminating the process of mode decision. In this section, two-stage threshold are proposed to early terminate the DIRECT mode decision.

For the first-stage threshold, determine the optimal mode prediction direction in MPCM for the current encoding MB firstly. Then, choose the maximum RD cost of the DIRECT mode among the corresponding MB and its eight adjacent MBs, and set it as the threshold T_{DIRECT}.

$$T_{\mathrm{DIRECT}} = \max\{K_i \cdot \mathrm{RD}_{\mathrm{Mb_i}}\}, \quad i \subseteq \{0,1,2,\ldots,8\} \tag{9}$$

where $\mathrm{RD}_{\mathrm{Mb_i}}$ is the RD costs of Mb_i computed at the DIRECT mode. K_i is equal to 1 when the DIRECT mode is the optimal mode of Mb_i. Otherwise, K_i is equal to 0. Lastly, compare the RD cost of the DIRECT mode in the current encoding MB with T_{DIRECT}, and determine whether to early terminate the mode decision process or not.

The mode correlation in the spatial domain is strong inferred from Table 3. If the adjacent MBs choose DIRECT mode as the optimal mode, the region are more likely to locate in the background or motion-less regions. And the mode of the current MB may be early decided as DIRECT mode. Donate ATh_M as the second-stage threshold by adopting modes of the adjacent MBs in MPCM, which is calculated as follows

$$ATh_M = \max\{M_{\mathrm{Mb_i}}\}, \quad i \subseteq \{1,2,3,4\}. \tag{10}$$

where $M_{\mathrm{Mb_i}}$ is the modes of adjacent MBs.

Then, the first-stage and the second-stage thresholds are used to early determine the mode of the current encoding MB. If the RDcost of the DIRECT mode of the current coding MB is smaller than T_{DIRECT} and ATh_M is equal to 0, choose the DIRECT mode as the optimal mode for the current coding MB. Otherwise, the process of mode decision is performed by adopting inter and intra mode early terminate strategy described in Sect. 3.2.

3.2 Inter and Intra mode early terminate strategy

As described in Sect. 2, the optimal mode of the current encoding MB is highly correlated with its co-located and adjacent MBs in the temporal or inter-view direction. Two kinds of mix early termination (ET) threshold are proposed to terminate the process of mode decision. The basic one uses the mode of the corresponding MB in the optimal mode prediction direction, and the adaptive threshold is defined as

$$CTh_M = M_{\mathrm{MB_0}} \tag{11}$$

where M_{MB_0} is the mode of the corresponding MB in the optimal prediction direction.

The second threshold is estimated based on the adjacent MBs' modes of the current encoding MB, the threshold is represented by Eq. 10. According to the analysis of Tables 3 and 4, the value of mode change between MB and its adjacent MBs which is <2 occupies 96.7 % on average. Therefore, the adaptive early termination threshold T_{INTER} is expressed as

$$T_{\mathrm{INTER}} = \begin{cases} CTh_M, & \text{if } ATh_M - CTh_M \le 0 \\ CTh_M + 1, & \text{if } ATh_M - CTh_M = 1 \\ CTh_M + 2, & \text{if } ATh_M - CTh_M \ge 2 \end{cases} \tag{12}$$

3.3 Proposed mode decision algorithm

Based on the above analysis, the proposed adaptive early mode decision based on the MPCM for current encoding MB is summarized as follow:

1. Start the mode decision of the current encoding MB.
2. MPCM: locate the adjacent MBs and the neighboring MBs of the current encoding MB by adopting the MPCM which is established in Sect. 2.2.
3. Optimal mode prediction direction: compute T_{ms} and V_{ms} of the current encoding MB based on Eqs. 7 and 8, respectively. If T_{ms} is larger than V_{ms}, choose Inter-view direction as the optimal prediction direction. Otherwise, choose temporal direction as the optimal direction.
4. Threshold calculation: compute T_{DIRECT}, ATh_M and T_{INTER} based on Eqs. 9, 10 and 12, respectively.
5. Early DIRECT mode decision: test DIRECT mode and calculate RDcost (RDcost_0). If RDcost_0 is smaller than T_{DIRECT} and ATh_M is equal to 0, choose DIRECT mode as the optimal mode and go to step 7.
6. Inter and Intra mode early terminate: according to the classification in Table 1, test the Inter and Intra modes

sequentially until the value of the tested mode is equal to T_{INTER}. And decide the optimal coding mode among all the examined DIRECT, Inter and Intra modes.

7. Perform the mode decision of the next MB.

4 Experiment results

The MVC reference software JMVC 8.3.1 with HBP prediction structure is used to demonstrate the proposed method coding performance, and the test sequences of multiview video have various motion activities, including 320 × 240 sequences (*Flomenco1* and *Golf1*), 640 × 480 sequences (*Ballroom*, *Exit*, *Race1*, *Vassar*), 1024 × 768 sequences (*Ballet*, *Door flowers*, *Jungle* and *Uli*), 1280 × 960 sequences (*Champagne tower* and *Dog*). The even views are used as the reference view (i.e., V0, V2, V4, V6 in Fig. 1), while the odd views (i.e., V1, V3, V5, V7 in Fig. 1) are utilized to test the proposed algorithm. Moreover, in order to demonstrate the coding efficiency of the proposed fast MVC method, the experimental results of three state-of-the-art algorithms, including Shen et al. [13], Zeng et al. [14], and Zhao et al. [15] are referenced. As the experimental results of Shen et al. [13], Zeng et al. [14] are derived from [15], the test configuration parameters of the proposed approach as shown in Table 6 are same to test the conditions in [15].

Table 6 Test configuration parameters

Search range	±96
Bi-prediction iteration	4
QP	20, 26, 32 and 38
RDO	Enabled
CABAC entropy	Used
GOP	12
Frames coded	2 GOPs × 8 views
Reference frames	2

We compare the coding efficiency of the proposed fast MVC method with three algorithms, Shen et al. [13], Zeng et al. [14] and Zhao et al. [15], in terms of BDPSNR, BDBR and time reduction (ΔT). The experimental results are compared in Tables 7 and 8. In these two tables, BDPSNR and BDBR are computed according to [16, 17], ΔT is defined as (13).

$$\Delta T = (T_{\mathrm{Proposed}} - T_{\mathrm{JMVC}})/T_{\mathrm{JMVC}} \times 100\,\% \tag{13}$$

where the subscript Proposed denotes the proposed algorithm, Shen et al. [13], Zeng et al. [14] and Zhao et al. [15]. JMVC represents the JMVC8.3.1.

Table 7 shows the performance comparison among Shen et al. [13], Zhao et al. [15] and the proposed algorithm on view 1, 3, 5 (mark with *). It can be observed that the approaches in Shen et al. [13] and Zhao et al. [15] reduce computational complexity by 77.52, 81.33 % on average, and the average BDPSNR are −0.220 dB, −0.094 dB, the average BDBR are 7.55, 3.18 %, respectively. The proposed algorithm save the encoding time from 71.70 to 95.15 % (86.47 % on average), the average BDPSNR and BDBR are −0.075 dB and 2.25 %, respectively. Compared to the method of Shen et al. [13], the proposed algorithm can achieve more computational complexity reduction while maintaining high coding performance. The total encoding time saves 8.95 % on average, BDPSNR increases 0.145 dB and BDBR decreases 5.3 %. Moreover,

Table 7 Comparison of multiview video sequences

Sequences	Shen et al. [13]			Zhao et al. [15]			Proposed*		
	Δ T (%)	BDPSNR (dB)	BDBR (%)	Δ T (%)	BDPSNR (dB)	BDBR (%)	Δ T (%)	BDPSNR (dB)	BDBR (%)
Flamenco1	−77.23	−0.195	4.66	−71.36	−0.130	3.12	−82.30	−0.118	2.51
Golf1	−81.70	−0.056	1.30	−92.24	−0.077	1.67	−95.15	−0.072	1.48
Ballroom	−70.92	−0.053	1.51	−74.17	−0.104	3.19	−86.03	−0.109	3.00
Exit	−75.79	−0.276	12.01	−79.25	−0.116	4.79	−85.43	−0.106	4.68
Race1	−73.05	−0.749	14.21	−77.06	−0.074	1.85	−84.75	−0.124	2.98
Vassar	−77.92	−0.014	0.65	−85.91	−0.037	1.93	−92.00	−0.035	1.58
Ballet	−81.73	−0.317	16.60	−82.39	−0.079	3.38	−87.96	−0.035	1.42
Door flowers	−87.29	−0.280	17.64	−89.63	−0.105	5.58	−93.34	−0.025	1.46
Jungle	−67.49	−0.267	7.87	−74.24	−0.141	4.08	−71.70	−0.061	1.66
Uli	−70.56	−0.206	6.13	−74.59	−0.126	3.65	−75.13	−0.103	2.87
Champagne tower	−88.31	−0.186	6.39	−91.03	−0.114	4.09	−93.06	−0.050	1.41
Dog	−78.24	−0.041	1.67	−84.06	−0.019	0.83	−90.82	−0.056	1.99
Average	−77.52	−0.220	7.55	−81.33	−0.094	3.18	−86.47	−0.075	2.25

* Views 1, 3, and 5

Table 8 Comparison of multiview video sequences

Sequences	Zeng et al. [14]			Zhao et al. [15]			Proposed		
	Δ T (%)	BDPSNR (dB)	BDBR (%)	Δ T (%)	BDPSNR (dB)	BDBR (%)	Δ T(%)	BDPSNR (dB)	BDBR (%)
Flamenco1	−78.05	−0.524	13.71	−71.34	−0.135	3.31	−81.48	−0.123	2.73
Golf1	−83.64	−0.032	0.69	−91.92	−0.104	2.22	−94.41	−0.085	1.73
Ballroom	−65.62	−0.297	8.63	−73.45	−0.123	3.57	−83.80	−0.119	3.25
Exit	−67.89	−0.334	14.19	−78.67	−0.126	5.18	−83.38	−0.103	4.31
Race1	−49.49	−0.346	9.27	−77.29	−0.106	2.70	−84.28	−0.119	2.98
Vassar	−79.35	−0.067	3.29	−85.42	−0.036	1.76	−91.38	−0.035	1.55
Ballet	−69.47	−0.213	9.26	−82.20	−0.117	5.04	−87.08	−0.043	1.73
Door flowers	−77.85	−0.077	4.13	−89.37	−0.098	5.27	−93.10	−0.024	1.37
Jungle	−65.59	−0.398	11.70	−73.75	−0.156	4.45	−70.51	−0.081	2.14
Uli	−67.97	−0.278	8.14	−74.16	−0.132	3.79	−75.24	−0.105	2.89
Champagne tower	−80.31	−0.108	3.56	−90.87	−0.142	4.73	−93.06	−0.050	1.41
Dog	−71.46	−0.191	7.51	−83.12	−0.026	1.00	−89.31	−0.053	1.86
Average	−71.39	−0.239	7.84	−80.96	−0.108	3.59	−85.59	−0.078	2.33

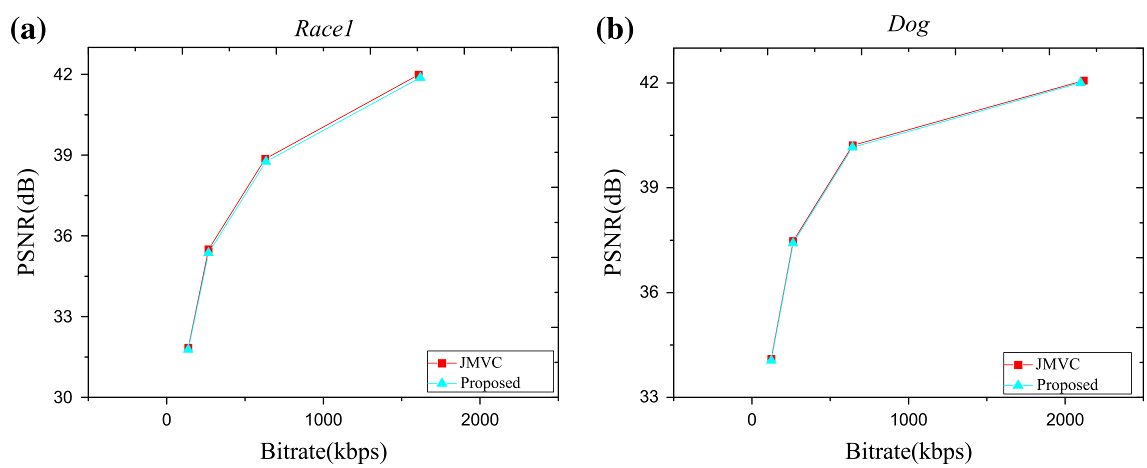

Fig. 3 The RD curves of two sequences. **a** RD curve of "*Race1*", **b** RD curve of "*Dog*"

compared to the method of Zhao et al. [15], the proposed algorithm achieves a better coding performance. The total encoding time saves 5.14 % on average, BDPSNR increases 0.019 dB and BDBR decreases 0.93 %.

Table 8 shows the encoding performance of all the odd views (i.e., views 1, 3, 5 and 7). It can be observed that the approaches in Zeng et al. [14] and Zhao et al. [15] reduce computational complexity by 71.39, 80.96 % on average, and the average BDPSNR are −0.239 dB, −0.108 dB, the average BDBR are 7.84, 3.59 %, respectively. The proposed algorithm save the encoding time from 70.51 to 94.41 % (85.59 % on average), the average BDPSNR and BDBR are −0.078 dB and 2.33 %, respectively. For video sequences such as *Golf1*, *Vassar*, *Door Flowers*, *Champagne tower* and *Dog*, the encoding time is saved 92.25 % on average, since the static background occupies a large areas of frames and DIRECT would be selected as the optimal mode to save more encoding time. For sequences such as *Flamenco1*, *Ballroom*, *Exit*, *Race1*, *Ballet*, *Jungle* and *Uli*, the average time saving is 80.82 %. Because these sequences contain lots of motion blocks, the mode correlation in spatial is not as stronger as other sequence (*Golf1* et al). Meanwhile, compared to the method of Zeng et al. [14], the proposed algorithm can achieve more computational complexity reduction while maintaining high coding performance. The total encoding time saves 14.2 % on average, BDPSNR increases 0.161 dB and BDBR decreases 5.51 %. Moreover, compared to the method of Zhao et al. [15], the proposed algorithm achieves a better coding performance. The total encoding time saves 4.63 % on average, BDPSNR increases 0.03dB and BDBR decreases 1.26 %.

In addition, the RD curves of two sequences (*Race1* and *Dog*) was given to demonstrate the overall RD performance of the proposed algorithm. From the Fig. 3, it is apparent that the proposed method has almost the same RD performance as compared with the JMVC.

5 Conclusion

In this paper, an adaptive early mode decision method for MVC, based on the MBs position constraint model (MPCM), is proposed to reduce the computational complexity. An experimental analysis is conducted to demonstrate that the probability is large, when the current MB has the similar mode to its corresponding MBs in adjacent, temporal and inter-view. Consequently, the probability is used to predict the best mode for the current MB. Experimental results have shown that the proposed fast mode decision can significantly reduce the total encoding time as compared with the exhaustive mode decision in JMVC8.3.1, while maintaining almost the same coding efficiency.

Acknowledgments This work is supported in part by the National Natural Science Foundation of China (61379143, 61232016, U1405254), the Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) under Grant 20120161110014 and the S&T Program of Xuzhou City (XM13B119) and the PAPD fund. The authors greatly appreciate Mr Moses Odero for his nice help in improving the English usages in this paper.

References

1. 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 **99**(4), 626–642 (2011)
2. Vetro, A., Tourapis, A.M., Muller, K., Chen, T.: 3D-TV content storage and transimission. IEEE Trans. Broadcast. **572**, 384–394 (2011)
3. Smolic, A., Muller, K., Stefanoski, N., Ostermann, J., Gotchev, A., Akar, G.B., Triantafyllidis, G., Koz, A.: Coding algorithms for 3DTV-A survey. IEEE Trans. Circ. Syst. Video Technol. **17**(11), 1606–1621 (2007)
4. Vetro, A., Pandit, P., Kimata, H.: Joint multiview video model (JMVM) 7.0. ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-Z207 (2008)
5. Chen, Y., Pandit, P., Yea, S.: WD 4 reference software for MVC (JMVC). ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-AD207 (2009)
6. Zhang, Y., Kwong, S., Xu, L., Jiang, G.Y.: Direct mode early decision optimization based on rate distortion cost property and inter-view correlation. IEEE Trans. Broadcast. **59**(2), 390–398 (2013)
7. Wang, F.S., Zeng, H.Q., Shen, Q.H., Du, S.D.: Efficient early direct mode decision for multiview video coding. Signal Process. Image Commun. **28**(7), 736–744 (2013)
8. Shen, L.Q., Liu, Z., Yan, T., Zhang, Z.Y., An, P.: Early SKIP mode decision for MVC using inter-view correlation. Signal Process. Image Commun. **25**(2), 88–93 (2010)
9. Zhu, W., Tian, X., Zhou, F., Chen, Y.W.: Fast inter mode decision based on textural segmentation and correlations for multiview video coding. IEEE Trans. Consum. Electron. **56**(3), 1696–1704 (2010)
10. Yeh, C.H., Li, M.F., Chen, M.J., Chi, M.C., Huang, X.X., Chi, H.W.: Fast mode decision algorithm through inter-view rate-distortion prediction for multiview video coding system. IEEE Trans. Ind. Inform. **10**(1), 594–603 (2014)
11. Zeng, H.Q., Ma, K.K., Cai, C.H.: Fast multiview video coding using adaptive prediction structure and hierarchical mode decision. IEEE Trans. Circ. Syst. Video Technol. **24**(9), 1566–1578 (2014)
12. Chan, C.C., Tang, C.W.: Coding statistics based fast mode decision for multiview video coding. J. Vis. Commun. Image Represent. **24**(6), 686–699 (2013)
13. Shen, L.Q., Liu, Z., An, P., Ma, R., Zhang, Z.Y.: Low-complexity mode decision for MVC. IEEE Trans. Circ. Syst. Video Technol. **21**(6), 837–843 (2011)
14. Zeng, H.Q., Ma, K.K., Cai, C.H.: Fast mode decision for multiview video coding using mode correlation. IEEE Trans. Circ. Syst. Video Technol. **21**(11), 1659–1666 (2011)
15. Zhao, T.S., Kwong, S., Wang, H.L., Wang, Z., Pan, Z., Kou, C.-C.J.: Multiview coding mode decision with hybrid optimal stopping model. IEEE Trans. Image Process. **22**(4), 1598–1609 (2013)
16. G. Bjontegaard. Calculation of average PSNR difference between RD-curves. Doc. VCEG-M33, Austin, TX (2001)
17. X. Li, M. Wien, J. R. Ohm. Rate-complexity-distortion evaluation for hybrid video coding. In: IEEE International Conference on Multimedia and Expo (ICME), pp. 685–190, Suntec City (2010)

Yue Li is a Ph.D student in Hunan University, China. He obtained his B.E. and M.E. degrees from Hengyang Normal University and Central South University in 2010 and 2013, respectively. His research interests are in the area of video coding, especially multiview video coding.

Gaobo Yang is a professor in the School of Information Science and Engineering, Hunan University, China. He obtained his Masters and Ph.D. degrees from East China Jiaotong University and Shanghai University in 2001 and 2004, respectively. From August 2010 to August 2011, he made an academic visit to the University of Surrey, UK. He has published more than 60 papers in international journals. His research interests include video information security and multimedia communication.

Yapei Zhu is a teaching assistant in Hengyang Normal University, China. She obtained her B.E. and M.E. degrees from Hengyang Normal University and Ningbo University in 2010 and 2013, respectively. Her research interests concentrate on image/video compression.

Can Liu is a lecturer in Hengyang Normal University, China. She obtained her B.E. and M.E. degree from North Center University in 2006 and 2009, respectively. Her research interests are in the area of image processing.

Kai Liu is a Masters student in North Center University, China. His research interests are in the area of image processing.