IET Image Processing

Research Article

Early DIRECT mode decision based on all-zero block and rate distortion cost for multiview video coding

ISSN 1751-9659
Received on 22nd December 2014
Revised on 9th June 2015
Accepted on 5th July 2015
doi: 10.1049/iet-ipr.2014.1018
www.ietdl.org

Zhaoqing Pan[1,2], Yun Zhang[3], Jianjun Lei[4] ✉, Long Xu[5], Xingming Sun[1,2]

[1]*Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing 210044, People's Republic of China*
[2]*School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, People's Republic of China*
[3]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, People's Republic of China*
[4]*School of Electronic Information Engineering, Tianjin University, Tianjin 300072, People's Republic of China*
[5]*Key Laboratory of Solar Activity, National Astronomical Observatories of Chinese Academy of Sciences, Beijing 100012, People's Republic of China*
✉ *E-mail: jjlei@tju.edu.cn*

Abstract: The exhaustive variable-block-size mode decision can efficiently remove the redundancies among the multiview videos, while it also leads to significant increase of computational complexity in the multiview video coding (MVC) encoder, and the high encoding complexity becomes a bottleneck for the MVC encoder to achieve real-time multimedia applications. To address this bottleneck, many fast mode decision methods have been proposed. However, most of them are only suitable for optimising the encoding complexity of the odd views of the MVC encoder. In this study, based on the property of the all-zero block and rate distortion (RD) cost of the DIRECT mode as well as the correlations between the current macroblock (MB) and its spatial–temporal nearby MBs, an early DIRECT mode decision method is proposed for reducing the encoding complexity of the MVC. Experimental results show that the proposed method achieves 48.25 and 55.64% on average encoding time saving for the even and odd views, respectively, whereas the RD performance degradation is quite acceptable. In summary, the proposed method efficiently reduces the encoding complexity for the MVC encoder.

1 Introduction

The multiview video coding (MVC) [1–3] is an extension of the H.264/ advanced video coding (AVC) standard [4], which was developed by the joint video team of the International Telecommunication Union-Telecommunication Standardisation Sector (ITU-T) video coding experts group and International Organisation for Standardisation/International Electrotechnical Commission (ISO/IEC) moving picture experts group. It aims to efficiently encode the multiview video which is generated by capturing the same scene simultaneously with multiple cameras from different viewpoints/angles. The multiview video is extremely useful for many multimedia applications, such as free viewpoint video, free viewpoint television, three-dimensional television etc. However, the volume of the raw multiview video data increases significantly as the increased number of cameras. Hence, an efficient multiview video encoding technique is crucial for the multiview video to be applied in real-time multimedia applications. In [5], a hierarchical B picture (HBP) prediction structure was designed for the MVC to remove the data redundancies among the multiview video sequences. An example of the MVC-HBP prediction structure is shown in Fig. 1, where the number of viewpoints equal to 8 (S_n, $n = 0, 1, 2, \ldots, 7$), and the group-of-picture (GOP) size is 12 (T_n, $n = 0, 1, 2, \ldots, 11$). On the basis of the techniques used in the views, all views can be classified into two groups: even views (S_0, S_2, S_4, and S_6) and odd views (S_1, S_3, S_5, and S_7). In the even views, the motion estimation (ME) is adopted to remove the temporal redundancies. In the odd views, besides the ME, the disparity estimation (DE) is used to exploit the inter-view redundancies. In the MVC variable-block-size mode decision process, the DIRECT/SKIP, B16 × 16, B16 × 8, B8 × 16, B8 × 8, B8 × 8Fret, Intra16 × 16, Intra8 × 8, Intra4 × 4, and PCM are checked sequentially. At last, the best mode, m^*, is determined according to the minimisation of the Lagrangian cost function

$$m^* = \arg\min_{m \in \mathbf{M}} D(m) + \lambda \cdot R(m), \tag{1}$$

where $\boldsymbol{M}$ is the candidate modes set, $\boldsymbol{M} = \{$DIRECT/SKIP, B16 × 16, B16 × 8, B8 × 16, B8 × 8, B8 × 8Fret, Intra16 × 16, Intra8 × 8, Intra4 × 4, PCM$\}$; $D(m)$ represents the total distortion by encoding the macroblock (MB) with the mode m, and is computed by the sum of squared difference; λ is the Lagrange multiplier; and $R(m)$ indicates the number of bits for encoding the MB with the mode m. Nevertheless, this 'try all and select the best' mode decision process leading to the computational complexity of the MVC encoder increased dramatically.

To reduce the computational complexity of the exhaustive mode decision process, many fast mode decision methods have been proposed for H.264. In [6], an adaptively fast mode decision was proposed for the H.264/AVC, which projects all candidate modes into a 2D map, and then the mode decision is performed according to a priority-based mode candidate list. In [7], a direct ME mode prediction was proposed, which uses the phase correlation to obtain the motion information between the current block and its reference blocks. In [8], based on the MB motion activity which is computed according to the motion vector of a set of spatial–temporal nearby MBs, a fast mode decision method was proposed. In [9], a fast mode decision method was proposed for the scalable video coding by using an all-zero block (AZB) detection technique [10]. These methods can efficiently reduce the computational complexity in the H.264 encoder; however, the characteristics of the MVC-HBP prediction structure are not considered. To consider the relationships among the MVC-HBP

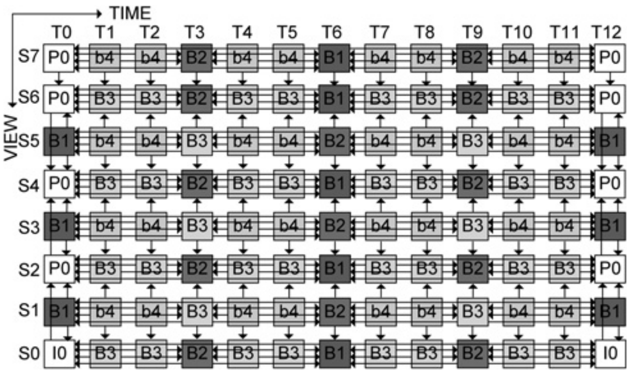

Fig. 1 *Example of the MVC-HBP prediction structure* ($S_n = 8$, *GOP size = 12)*

prediction structure, a lot of works have been proposed for reducing the computational complexity of the MVC encoder. In [11], a mode correlation-based mode decision algorithm was proposed for the fast MVC mode decision. First, based on the motion activity which is measured by the predicted motion vectors of the current MB and a group of spatial and inter-view neighbouring MBs, all candidate modes are classified into five classes. Then, the SKIP mode is checked first, if it satisfies to the early termination threshold, the other mode classes are skipped. Otherwise, the following mode classes will be further checked. In [12], based on the textural segmentation and correlations, a fast inter mode decision was proposed for the MVC. By converting the mode decision problem to a decision tree problem, an early SKIP mode decision method was proposed in [13]. In [14], based on the coding mode complexity which is computed by the mode selection of the spatial neighbouring MBs and inter-view corresponding MBs, an early SKIP mode decision was proposed. In [15], Wang *et al.* proposed an early DIRECT mode decision for the MVC by using the rate distortion (RD) cost correlation between the current MB and the spatial, temporal, and inter-view neighbouring MBs. On the basis of the coded block pattern, RD cost threshold, and inter-view correlation, Zhang *et al.* proposed a DIRECT mode early termination method for MVC [16]. However, the collocated MBs in the inter-views are obtained by the global disparity vector (GDV), whereas the GDV is computed according to the global displacement of the MB, which is not accurate enough for the toed-in camera arrangement and large depth-of-fields videos. In addition, the even views are used for obtaining the reference information, which results in no complexity optimisation for the even views of the MVC encoder.

In this paper, based on the characteristics of the AZB and RD cost, an early DIRECT mode decision method is proposed. The experimental results demonstrate that the proposed method can efficiently reduce the encoding complexity of the MVC encoder. The rest of this paper is organised as follows. Section 2 presents the motivations and statistical results. The details of the proposed early DIRECT mode decision method are illustrated in Section 3. Experimental results are shown in Section 4. Finally, Section 5 concludes this paper.

2 Motivations and statistical results

In the H.264-based MVC coding standard, the SKIP mode in the P slices uses the median prediction of motion vector (MV)s of spatial neighbouring MBs as its MV, and no residual information is transmitted. The DIRECT mode in the B slices is similar to the SKIP mode, while the residual information needs to be transmitted, and its MV is derived from spatial, temporal, or inter-view corresponding MBs. The DIRECT and SKIP modes have an excellent performance in coding efficiency and coding complexity. In this paper, the DIRECT mode denotes both the DIRECT and SKIP modes. It is well known that the natural video sequences contain a huge number of regions with simple content, slow motion, or background, and these regions are suitable for encoding in the DIRECT mode. Four multiview video sequences with various motion activities are encoded by the MVC reference software Joint Multiview Video Coding (JMVC) 8.0 to analyse the percentage of the DIRECT mode selected as the best mode in real video coding. The Ballroom [17] and Breakdancer [18] move fast. The Ballet [18] and Dog [19] are with medium and slow motion, respectively. Four basis quantisation parameters (QPs) (24, 28, 32, and 36) are tested, and the statistical results are tabulated in Table 1.

From Table 1, we can see that there are 56.73–88.94%, 71.35% on average, MBs selecting the DIRECT mode as their best mode. Another observation from Table 1 is that the DIRECT mode has a large probability to be selected as the best mode for video sequences with medium and slow motion, the Ballet and Dog are with 78.04 and 82.55%, respectively. While the probability has a little decrease for the video sequences with fast motion. For the Ballroom and Breakdancer, there are 67.82 and 64.69% MBs selecting the DIRECT as their best mode, respectively. On the other hand, there are different probabilities between the even views and odd views. The odd views have a larger probability than the even views. From Table 1, the probabilities of the even and odd views are 69.38 and 73.27%, respectively. In [20], it has been proved that the variable-block-size ME and DE consume about 70% (one reference frame) to 90% (five reference frames) of the total encoding time. However, the ME and DE are not required for the DIRECT mode, and the DIRECT mode holds a large proportion (71.35%) to be the best mode. Therefore, if the DIRECT can be determined early, and the mode decision process will be terminated early so that much more encoding complexity will be saved.

3 Proposed AZB and RD cost-based early DIRECT mode decision

The natural video sequences contain a substantial number of regions with simple content, background, or slow motion, and these regions are quite suitable for encoding in the DIRECT mode. In addition, when one block is encoded in the DIRECT mode, the prediction residual has a large probability to be transformed and quantised to zeros, and the block with all-zero quantised Discrete Cosine Transform (DCT) coefficients is called AZB. Therefore, after encoding the current MB with the DIRECT mode, if the prediction

Table 1 Percentages of the DIRECT mode to be the best mode (%)

View	QP	Ballroom	Ballet	Breakdancer	Dog	Average
even views	24	56.73	69.99	49.45	73.70	59.96
	28	63.12	74.48	59.79	78.89	67.27
	32	67.84	78.22	67.02	83.26	72.71
	36	72.21	81.86	73.76	86.77	77.58
	average	64.98	76.14	62.51	80.66	69.38
odd views	24	62.91	73.26	54.24	79.24	65.46
	28	68.68	78.09	63.99	83.37	72.01
	32	73.17	82.31	71.34	86.20	76.90
	36	77.92	86.11	77.91	88.94	81.59
	average	70.67	79.94	66.87	84.44	73.99
average		67.82	78.04	64.69	82.55	73.27

IET Image Process., 2016, Vol. 10, Iss. 1, pp. 9–15

Table 2 Statistical results of probability $P(A|B)$

QP	Even views				Odd views			
	Ballroom	Ballet	Vassar	Average	Ballroom	Ballet	Vassar	Average
24	0.947	0.969	0.939	0.951	0.902	0.954	0.936	0.931
28	0.973	0.986	0.993	0.984	0.921	0.973	0.971	0.955
32	0.983	0.991	0.998	0.990	0.935	0.980	0.980	0.965
36	0.985	0.993	0.999	0.992	0.952	0.986	0.986	0.975
average	0.972	0.985	0.982	0.979	0.928	0.973	0.968	0.956

residuals are all transformed and quantised to zeros, the current MB has a large probability to select the DIRECT mode as its best mode. To analyse the relationship between the AZB and the DIRECT mode which is selected as the best mode, let the event $\boldsymbol{A}$ represent the current MB is an AZB, the event $\boldsymbol{B}$ indicate the best mode of the current MB is the DIRECT mode, three multiview video sequences (Ballroom, Ballet [17] and Vassar [17]) are used to analyse the probability $P(\boldsymbol{A}|\boldsymbol{B})$, and the statistical results are listed in Table 2. It can be observed that when the DIRECT mode is selected as the best mode, the MB has a quite large probability to be an AZB, about 0.979 and 0.956 in the even and odd views, respectively. On the basis of this characteristic, $Q_B = \text{AZB}$ is set as the early DIRECT mode decision condition, which means after encoding one MB with the DIRECT mode, if it is an AZB, the best mode of this MB is the DIRECT mode, and the following INTER and INTRA predictions will be skipped.

To evaluate the coding efficiency of the early DIRECT mode decision condition, the RD performance comparison between the early DIRECT mode decision condition with $Q_B = \text{AZB}$ and the original JMVC is shown in Table 3. In Table 3, the Bjontegaard delta peak signal-to-noise ratio (BDPSNR) and BD bit rate (BDBR) are computed according to [21], and stand for the average PSNR differences in decibels for the same BRs, and the average BR differences in per cent for the same PSNR, respectively [22].

From Table 3, it can be seen that when setting the $Q_B = \text{AZB}$ as the early termination condition, it achieves a quite worst RD performance as compared with the original JMVC 8.0. For the Ballroom sequence, in the even views, the BDPSNR and BDBR between the AZB-based early DIRECT mode decision and the original JMVC 8.0 are −2.209 dB and 79.36%, respectively; in the odd views, the BDPSNR and BDBR between the AZB-based early DIRECT mode decision and the original JMVC 8.0 are −1.223 dB and 35.33%, respectively. For the Ballet sequence, in the even views, the BDPSNR and BDBR between the AZB-based early DIRECT mode decision and the original JMVC 8.0 are −2.409 dB and 45.72%, respectively; in the odd views, the BDPSNR and BDBR between the AZB-based early DIRECT mode decision and the original JMVC 8.0 are −0.904 dB, and 33.52%, respectively. From these values, it can be observed that when the single AZB is used as the early DIRECT mode decision condition, the RD performance degrades significantly. This is because when the AZB is used as the early DIRECT mode decision condition, there are about 2–4% MBs falsely encoded as the DIRECT mode; these 2–4% MBs skip all other modes, and will result in larger RD performance degradation. Hence, we can conclude that no matter how much encoding time saving is achieved by using the AZB-based early termination strategy, it is not suitable for the condition of early DIRECT mode decision. Hence, we should combine AZB with other conditions to design a stricter early DIRECT mode decision condition for achieving a trade-off between the RD performance and encoding complexity saving.

Table 3 BDPSNR and BDBR between the early DIRECT mode decision condition with $Q_B = \text{AZB}$ and the original JMVC 8.0

Sequence	Even views		Odd views	
	BDPSNR, dB	BDBR, %	BDPSNR, dB	BDBR, %
Ballroom	−2.209	79.36	−1.223	35.33
Ballet	−2.409	45.72	−0.904	33.52

In the RD-based MVC mode encoding process, the best mode is the mode which is with the minimum RD cost. Thus, if the RD cost of the DIRECT mode is different from the reminder inter modes (B16×16, B16×8, B8×16, B8×8, and B8×8Fret), the DIRECT mode can be classified into one separate group. Reference [23] has analysed the linear correlation of the RD cost between the DIRECT mode and the reminder inter modes. The linear correlation has been drawn that the RD cost of the DIRECT mode is quite different from these reminder inter modes, whereas the RD cost of the reminder inter modes are similar to each other. Hence, based on the RD cost property, all inter modes can be classified into two kinds of modes, DIRECT mode, and non-DIRECT mode, respectively. In addition, in MVC encoding process, the best mode selection and its RD cost of the current MB has a large correlation with its spatial–temporal neighbouring MBs [15, 24]. An illustration on the spatial–temporal neighbouring MBs of the current MB is shown in Fig. 2, where the 'CB' denotes the current MB; S_1–S_4 are the four spatial neighbouring MBs; T_5–T_{13} are the nine temporal neighbouring MBs. Thus, based on above analyses, the DIRECT mode can be determined early if

$$J_D \leq \alpha \cdot J_{ND}, \tag{2}$$

where J_D indicates the RD cost value of the DIRECT mode of the current MB; α is a regulating parameter; J_{ND} denotes the average RD cost value of the DIRECT mode of the spatial and temporal neighbouring MBs, and is computed as

$$J_{ND} = \frac{\sum_{i=1}^{13} K_i \cdot J_i}{\sum_{i=1}^{13} K_i}, \tag{3}$$

where i is the index of the neighbouring MB of the current MB, $i = 1, 2, \ldots, 13$, as shown in Fig. 2; J_i is the RD cost value of the DIRECT mode; K_i is used to decide whether the best mode of the ith neighbouring MB is the DIRECT mode, if the best mode of the neighbouring MB is the DIRECT mode, K_i equals to 1, otherwise, K_i equals to 0.

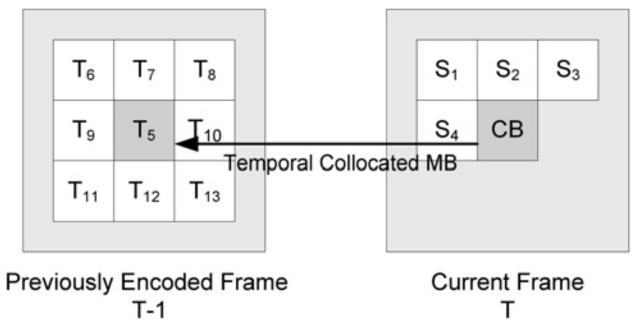

Fig. 2 *Illustration on the spatial and temporal neighbouring MBs of the current MB*

Table 4 Summary of encoding performance of different α values

α	Even views			Odd views		
	BDPSNR, dB	BDBR, %	TS, %	BDPSNR, dB	BDBR, %	TS, %
0.9	−0.002	0.09	−31.27	−0.021	0.89	−40.66
1.0	−0.003	0.10	−37.46	−0.032	1.07	−47.83
1.1	−0.005	0.21	−42.98	**−0.039**	**1.38**	**−56.09**
1.2	**−0.010**	**0.35**	**−46.71**	−0.041	1.58	−59.32
1.3	−0.015	0.53	−48.71	−0.052	1.72	−62.44
1.4	−0.016	0.63	−51.79	−0.056	2.05	−64.65
1.5	−0.018	0.76	−53.74	−0.064	2.26	−66.48

To achieve the best RD performance and encoding time saving, the AZB and RD cost are jointly applied to terminate early the mode decision process. Hence, the best mode of the current MB, M_B, is determined by

$$M_B = \begin{cases} \text{DIRECT}, & \text{if } Q_B = \text{AZB} \;\&\&\; J_D \le \alpha \cdot J_{ND}, \\ \text{non}-\text{DIRECT}, & \text{otherwise}, \end{cases} \quad (4)$$

where && is the logical AND, which means after encoding the MB with the DIRECT mode, if it is an AZB and the RD cost is less than or equal to the threshold J_{ND}, the best mode of the current MB is the DIRECT mode, then the following mode decision process is terminated. α is a regulating parameter, which is used to obtain the best RD performance and encoding time saving. If the α is larger, the more encoding time saving comes at the cost of larger RD degradation. Instead, if the α is smaller, the better RD performance is at the cost of lower encoding time saving.

For exploring the best value of α, which obtains the best trade-off between the RD performance and encoding time saving, a group of α values are encoded sequentially, $\alpha \in \{0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5\}$. The multiview video test sequence, Exit, which is with medium motion activity, is used for analysing the encoding performance, and the original JMVC 8.0 is used as the benchmark. The summary of encoding results is shown in Table 4, where the TS denotes the total encoding time saving, and it is defined as

$$\text{TS} = ((\text{Time}_\alpha - \text{Time}_o)/\text{Time}_o) \times 100\%, \quad (5)$$

where Time_α indicates the encoding time of the JMVC 8.0 with the fast mode decision of different α values; Time_o means the encoding time of the original JMVC 8.0 with the full mode decision. Since the even views will be referred by the odd views, the BDBR variation of the even views should in a small range; otherwise, the RD performance of the odd views will degrade dramatically. Hence, we set the upper bounds for the BDBR increase to 0.5% for the even views. To control the final encoding performance of the odd views, we set the upper bounds for the BDBR increase to 1.5% for the odd views. From Table 4, we can see that when the α is equal to 1.2, the encoding results of BDPSNR, BDBR, and TS are −0.010 dB, 0.35%, and −46.71%, respectively. These three values are quite acceptable. Thus, the α value is set to 1.2 for encoding the even views. In the odd views, when the α value is equal to 1.1, the BDBR is no larger than 1.5%, and the BDPSNR, BDBR, and TS are −0.039 dB, 1.38%, and −56.09%, respectively. Therefore, the α value is set to 1.1 for encoding the odd views.

To demonstrate the efficiency of the proposed AZB and RD cost-based early DIRECT mode decisions, four multiview video sequences (Ballroom, Doorflowers [25], Exit, and Vassar) are used to analyse the probabilities of the DIRECT mode falsely accepted as the best mode. The statistical results are tabulated in Table 5. From Table 5, it can be observed that for the even views, the error rate of the proposed early DIRECT mode decision is from 0 to 6.42%, 1.69% on average. For the odd views, the error rate of the proposed method is from 0.32 to 4.56%, 1.58% on average. In other words, the hit rate of the proposed early DIRECT mode decision method is 98.31 and 98.42% for the even and odd views, respectively. These values demonstrate that the proposed method can effectively determine the DIRECT mode, and terminate the mode decision process.

Table 5 Probability of the DIRECT mode false accepted as the best mode (%)

QP	Ballroom Even/odd	Doorflowers Even/odd	Exit Even/odd	Vassar Even/odd	Average Even/odd
24	6.23/4.56	3.14/3.31	3.87/3.26	6.42/3.55	4.92/3.67
28	2.07/1.29	1.08/1.59	0.69/1.21	1.08/0.59	1.23/1.17
32	0.85/1.16	0.42/0.87	0.20/0.58	0.22/0.92	0.42/0.88
36	0.52/0.87	0.15/0.42	0.10/0.32	0.00/0.76	0.19/0.59
average	2.42/1.97	1.20/1.55	1.22/1.34	1.93/1.46	1.69/1.58

Finally, based on above analyses, the overall proposed AZB and RD cost-based early DIRECT mode decision method is summarised as follows:

Step 1: Encode the current MB with the DIRECT mode, get its quantised DCT coefficients and RD cost, denote as Q_B and J_D, respectively;
Step 2: Compute the average RD cost of the DIRECT mode according to (3), and denote the average RD cost as J_{ND};
Step 3: The best mode of the current MB is selected according to (4). If Q_B = AZB and $J_D \le \alpha \cdot J_{ND}$, the α values are equal to 1.2 and 1.1 for the even and odd views, respectively, the best mode of the current MB is set as the DIRECT mode, go to step 5; otherwise, go to step 4;
Step 4: Encode the current MB with the non-DIRECT modes one by one, and determine the best mode according to (1). Go to step 5; and
Step 5: Go to step 1 and encode the next MB.

4 Experimental results

To evaluate the efficiency of the proposed early DIRECT mode decision method, the MVC reference software JMVC 8.0 is used as the software platform. The JMVC 8.0 adopts an HBP prediction structure to remove the spatial, temporal, and inter-view redundancies of the multiview video, which achieves high coding efficiency. The test conditions used in JMVC 8.0 are listed as follows: four basis QPs (24, 28, 32, and 36) are used; the GOP size equals to 12; the number of reference frames is 2; the ME/DE search range is set to 64; the maximum number of iterations for bi-prediction search is 4; the ME/DE search method is TZSearch; the RD optimisation is enabled; the entropy coding method is Context-based Adaptive Binary Arithmetic Coding (CABAC); and 49 frames to be encoded. Seven multiview video test sequences (Ballroom, Exit, Race1 [26], Breakdancers, Ballet, Doorflowers, and Dog) with various motion activities are used in the experiments. The hardware platform is Intel Core 2 Duo central processing unit E5800 at 3.16 and 3.17 GHz, 4.00 GB random access memory with Microsoft Windows 7 64 bit operating system.

We compare the proposed early DIRECT mode decision method with the Shen's method [14] and Wang's method [15] in terms of PSNR, BR and total encoding time saving. The JMVC 8.0 is used as the benchmark. The comparison results are summarised and tabulated in Table 6. In this table, the ΔPSNR, ΔBR, and ΔT are defined as

$$\begin{cases} \Delta\text{PSNR} & = \text{PSNR}_\Omega - \text{PSNR}_o \text{ (dB)}, \\ \Delta\text{BR} & = \dfrac{\text{BR}_\Omega - \text{BR}_o}{\text{BR}_o} \times 100\% \text{ (\%)}, \\ \Delta T & = \dfrac{\text{Time}_\Omega - \text{Time}_o}{\text{Time}_o} \times 100\% \text{ (\%)}, \end{cases} \quad (6)$$

where the PSNR_Ω, BR_Ω, and Time_Ω are the PSNR, BR, and encoding time of the method Ω, Ω ∈ {Shen [14], Wang [15], and proposed}; the PSNR_o, BR_o, and Time_o denote the PSNR, BR, and encoding time of the original JMVC 8.0, respectively.

From Table 6, it can be seen that for the even views, the proposed method achieves the total encoding time saving from 31.33 to

Table 6 Summary of encoding results

Sequences	QP	Even views	Odd views		
		Proposed vs. JMVC 8.0	Shen *et al.* [14] vs. JMVC 8.0	Wang *et al.* [15] vs. JMVC 8.0	Proposed vs. JMVC 8.0
		ΔPSNR/ΔBR/Δ*T*	ΔPSNR/ΔBR/Δ*T*	ΔPSNR/ΔBR/Δ*T*	ΔPSNR/ΔBR/Δ*T*
Ballet	24	−0.023/−0.61/−46.71	−0.033/−0.32/−33.59	−0.038/−0.65/−52.38	−0.032/−0.75/−54.40
	28	−0.013/−0.25/−48.13	−0.036/0.73/−45.36	−0.076/−0.54/−53.29	−0.066/−0.56/−53.78
	32	−0.007/−0.08/−50.40	−0.052/1.78/−50.28	−0.084/−0.72/−53.97	−0.094/−0.92/−54.75
	36	−0.002/−0.04/−54.87	−0.090/2.11/−52.90	−0.096/−0.41/−54.87	−0.087/−0.37/−57.26
	average	−0.011/−0.24/−49.62	−0.053/1.08/−45.53	−0.074/−0.58/−53.63	−0.070/−0.65/−55.05
	BDPSNR/BDBR	−0.007/0.23	−0.081/3.08	−0.058/1.81	−0.052/1.64
Ballroom	24	−0.058/−0.76/−43.51	−0.022/−0.07/−17.10	−0.047/−0.59/−51.28	−0.054/−0.69/−50.72
	28	−0.026/−0.33/−42.34	−0.014/0.11/−17.69	−0.027/−0.48/−53.37	−0.024/−0.37/−53.16
	32	−0.015/−0.16/−43.04	−0.020/0.32/−22.92	−0.038/−0.54/−55.41	−0.034/−0.42/−54.35
	36	−0.012/−0.10/−45.69	−0.029/0.74/−38.17	−0.045/−0.61/−56.23	−0.049/−0.45/−54.40
	average	−0.028/−0.34/−43.64	−0.021/0.28/−23.97	−0.039/−0.56/−54.07	−0.040/−0.48/−53.16
	BDPSNR/BDBR	−0.014/0.36	−0.029/0.77	−0.015/0.42	−0.018/0.50
Breakdancers	24	−0.060/−1.03/−31.33	−0.013/−0.26/−1.48	−0.062/−0.98/−41.27	−0.060/−1.12/−39.65
	28	−0.054/−0.62/−34.49	−0.027/−0.23/−5.99	−0.083/−1.12/−43.49	−0.080/−1.10/−42.96
	32	−0.030/−0.38/−38.11	−0.045/−0.65/−8.14	−0.091/−1.23/−46.73	−0.100/−1.31/−45.27
	36	−0.011/−0.15/−45.52	−0.076/−0.91/−17.26	−0.082/−1.27/−51.58	−0.078/−1.19/−52.38
	average	−0.039/−0.54/−37.36	−0.040/−0.51/−8.22	−0.080/−1.15/−45.77	−0.079/−1.18/−45.07
	BDPSNR/BDBR	−0.034/1.55	−0.024/1.02	−0.055/2.46	−0.055/2.44
Dog	24	−0.059/−1.01/−56.97	−0.035/−0.91/−43.71	−0.065/−0.98/−64.73	−0.061/−1.09/−65.72
	28	−0.035/−0.51/−58.64	−0.026/−0.73/−52.87	−0.072/−0.85/−66.73	−0.067/−0.73/−67.24
	32	−0.019/−0.29/−60.40	−0.033/−0.64/−58.33	−0.092/−1.01/−69.12	−0.096/−1.14/−68.81
	36	−0.008/−0.12/−62.33	−0.044/−0.74/−64.32	−0.095/−1.29/−69.86	−0.091/−1.33/−70.36
	average	−0.030/−0.48/−59.58	−0.035/−0.76/−54.81	−0.081/−1.03/−67.52	−0.079/−1.07/−68.03
	BDPSNR/BDBR	−0.014/0.37	−0.088/0.27	−0.046/1.45	−0.044/1.35
Doorflowers	24	−0.081/−1.37/−55.08	−0.062/0.15/−45.20	−0.098/−1.81/−57.28	−0.107/−2.36/−61.49
	28	−0.059/−0.73/−55.99	−0.054/−0.19/−56.68	−0.091/−1.42/−59.76	−0.102/−1.45/−60.99
	32	−0.036/−0.41/−59.75	−0.065/1.43/−63.62	−0.087/−1.27/−61.39	−0.099/−1.08/−61.95
	36	−0.020/−0.14/−62.80	−0.044/0.19/−67.87	−0.080/−1.19/−62.87	−0.082/−1.63/−63.65
	average	−0.049/−0.67/−58.41	−0.056/0.40/−58.34	−0.089/−1.42/−60.33	−0.098/−1.63/−62.02
	BDPSNR/BDBR	−0.034/1.15	−0.064/2.95	−0.051/1.88	−0.058/2.20
Exit	24	−0.036/−0.71/−41.72	−0.038/1.04/−30.39	−0.038/−0.62/−52.43	−0.035/−0.52/−51.42
	28	−0.014/−0.23/−45.45	−0.070/2.22/−39.41	−0.056/−0.68/−55.17	−0.056/−0.47/−56.43
	32	−0.005/−0.12/−48.22	−0.112/2.79/−45.55	−0.059/−0.67/−55.75	−0.058/−0.61/−56.54
	36	−0.001/−0.03/−51.44	−0.156/3.17/−49.62	−0.063/−0.75/−56.76	−0.062/−0.60/−59.95
	average	−0.014/−0.27/−46.71	−0.094/2.31/−41.24	−0.054/−0.68/−55.03	−0.053/−0.55/−56.09
	BDPSNR/BDBR	−0.010/0.35	−0.165/5.55	−0.036/1.32	−0.039/1.39
Race1	24	−0.092/−0.82/−39.76	−0.037/0.04/−8.80	−0.108/−1.11/−45.27	−0.119/−1.26/46.14
	28	−0.101/−0.95/−42.49	−0.084/0.33/−12.32	−0.133/−1.32/−46.88	−0.156/−1.58/−48.82
	32	−0.084/−0.62/−42.95	−0.144/1.30/−16.80	−0.137/−1.17/−49.96	−0.153/−1.23/−51.26
	36	−0.068/−0.48/−44.48	−0.209/3.05/−16.74	−0.098/−1.13/51.39	−0.114/−1.05/−53.97
	average	−0.086/−0.72/−42.42	−0.119/1.18/−13.67	−0.119/−1.18/−48.38	−0.135/−1.28/−50.04
	BDPSNR/BDBR	−0.058/1.36	−0.157/3.67	−0.044/1.08	−0.088/2.12
average PSNR/BR/TS		−0.037/−0.47/−48.25	−0.060/0.57/−35.11	−0.077/−0.94/−54.96	−0.079/−0.98/—55.64
average BDPSNR/BDBR		−0.024/0.77	−0.087/2.47	−0.044/1.49	−0.051/1.66

62.8%, 48.25% on average; meanwhile, the PSNR degrades from 0.001 to 0.101 dB, 0.037 dB on average; and the BR increase is from −1.37 to −0.03%, −0.47% on average. The average BDPSNR and BDBR between the proposed method and the original JMVC 8.0 are −0.024 dB and 0.77%, respectively. From these values, we can figure out that the proposed early DIRECT mode decision method reduces the computational complexity efficiently for the even views of the MVC encoder. However, the Shen's and Wang's methods were designed for the computational complexity reduction for the odd views, and are not suitable for optimising the encoding complexity of the even views. Thus, the even views are encoded by the original JMVC 8.0 so that no encoding time saving is gained.

For the odd views, the Shen's method saves the total encoding time from 1.48 to 67.87%, and 35.11% on average; meanwhile, the PSNR degrades from −0.013 to −0.209 dB, −0.060 dB on average; and the BR increases from −0.91 to 3.17%, 0.57% on average. The average BDPSNR and BDBR between the Shen's method and the original JMVC 8.0 are −0.087 dB and 2.47%, respectively. For the multiview video sequences with violent motion activities, the Shen's method cannot effectively reduce the complexity, such as Breakdancers and race1, only 8.22%, 13.67% encoding time saving, respectively. Since in these multiview video sequences with violent motion activities, the Shen's method cannot efficiently early terminate the mode decision process. The Wang's method reduces the total encoding time from 41.27 to 69.86%, 54.96% on average. Meanwhile, the PSNR degrades from 0.027 to 0.137 dB, 0.077 dB on average; and the BR increases from −0.41 to −1.81%, −0.94% on average. The average BDPSNR and BDBR between the Wang's method and the original JMVC 8.0 are −0.044 dB and 1.49%, respectively. The proposed method reduces 39.65–63.65%, 55.64% on average total encoding time; while the PSNR degradation is within 0.024–0.156 dB, 0.079 dB on average; the BR increase is within −1.58 to −0.37%, −0.98% on average; the BDPSNR and BDBR between the proposed method and the original JMVC 8.0 are −0.051 dB and 1.66%, respectively. From these values, we can see that (i) the proposed method obtains a better performance than the Shen's method in not only the RD performance but also the encoding time saving; (ii) the RD performance of the proposed method has a little decrease as compared with the Wang's method, this is because that the proposed method can optimise the encoding complexity of the even and odd views of the MVC encoder, while the Wang's only work for the odd views of the MVC encoder, the large computational complexity saving of the even views of the proposed method causes the RD degradation of the odd views. Finally, we can draw that the proposed method reduces the encoding complexity efficiently for the even and odd views of the MVC encoder.

To illustrate the RD performance and encoding time saving intuitively, the RD curves comparison between the proposed method and the original JMVC 8.0 are shown in Fig. 3, where it can be observed that the proposed method achieves the almost same RD performance with the original JMVC 8.0. Fig. 4 shows

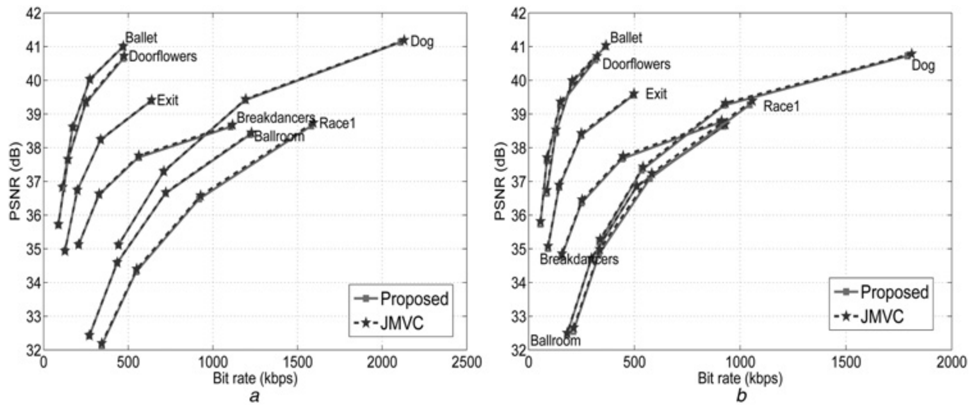

Fig. 3 *Comparison of the RD curves*

a Even views
b Odd views

the summary of encoding time saving. Fig. 4*a* presents the encoding time saving of the proposed method for the even and odd views, it can be observed that the proposed method efficiently reduces the computational complexity for the MVC encoder. For the even and odd views, at least 30 and 40% computational complexities have been removed, respectively. It can also be seen that the encoding time saving ratio of the odd views is larger than that of the even views, this is because that the odd views need the ME and DE, whereas the even views only need the ME, which results in the encoding complexity of the odd views is higher than that of the even views. Fig. 4*b* is the comparison of the encoding time saving among the Shen's method, Wang's method, and the proposed method. In Fig. 4*b*, the total encoder time saving of the MVC encoder, TS_{MVC}, is computed as $TS_{MVC} = (TS_{Even} + TS_{odd})/2$, where TS_{Even} and TS_{Odd} indicate the total encoding time saving of the even and odd views of the MVC encoder, respectively. We can see that the Shen's method can achieve from 4 to 29% encoding time saving for the MVC encoder, the Wang's method can reduce the encoding complexity of the MVC encoder from 22 to 33%, and the proposed method saves the encoding time from 41 to 63%. It can be observed that the proposed method obtains the best encoding time saving for the MVC encoder, this is because the proposed method reduces the encoding complexity efficiently for both the odd and even views of the MVC encoder, whereas the Shen's and Wang's methods only optimise the encoding complexity of the odd views of the MVC encoder.

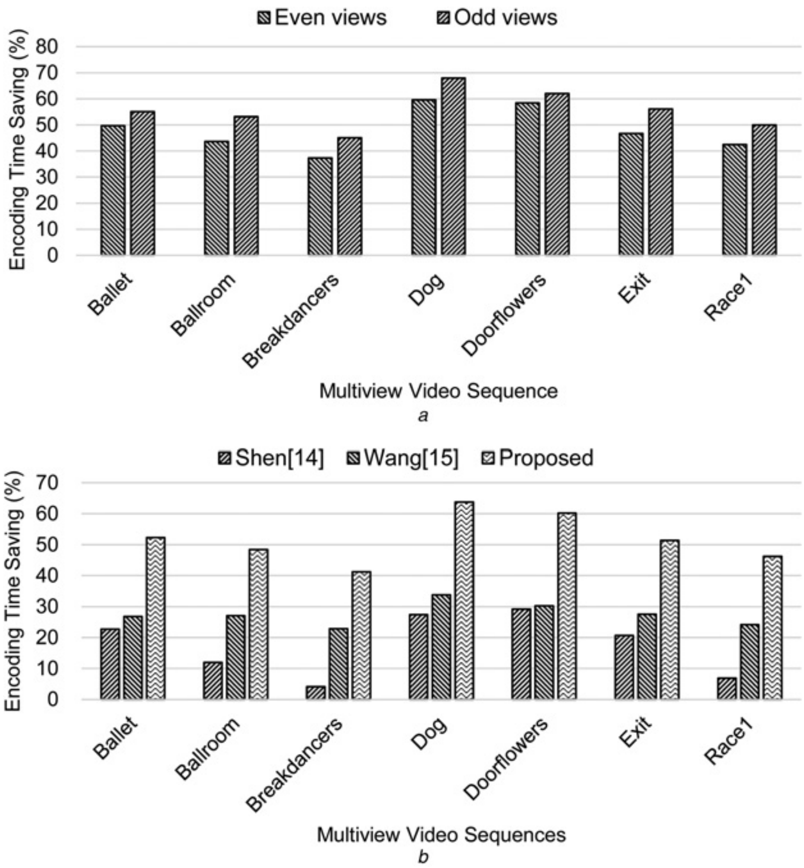

Fig. 4 *Summary of encoding time saving*

a Encoding time saving of the even and odd views of the proposed method
b Encoding time saving comparison among Shen *et al.* [14], Wang *et al.* [15], and the proposed method

Eventually, we can conclude that the proposed method reduce the encoding complexity of the MVC encoder efficiently.

5 Conclusion

In this paper, an early DIRECT mode decision algorithm is proposed for the MVC. First, an experiment is performed to analyse the coding performance of the AZB-based early DIRECT mode decision condition. Then, the AZB and RD cost are jointly used for the early termination of the MVC mode decision process. Experimental results show that the proposed method obtains an excellent coding performance in terms of the RD performance and encoding time saving. In summary, the proposed method efficiently reduces the encoding complexity for both the odd views and even views.

6 Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under grants 61232016, 61471348, and 61102088, in part by the Project through the Priority Academic Program Development of Jiangsu Higher Education Institutions, in part by the Startup Foundation for Introducing Talent of Nanjing University of Information Science and Technology (NUIST) under grant 2243141501012, and in part by the Shenzhen Overseas High-Caliber Personnel Innovation and Entrepreneurship Project under grant KQCX20140520154115027. Also in part by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant 15KJB510019.

7 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*, 2011, **99**, (4), pp. 626–642
2 Shao, F., Yu, M., Jiang, G., *et al*.: 'Colour correction pre-processing and chrominance reconstruction post-processing for multi-view video coding', *IET Image Process.*, 2012, **6**, (2), pp. 129–138
3 Lei, J., Feng, K., Wu, M., *et al*.: 'Rate control of hierarchical B prediction structure for multi-view video coding', *Multimedia Tools Appl.*, 2014, **72**, (1), pp. 825–842
4 ITU-T and ISO/IEC JTC 1: 'Advanced video coding for generic audiovisual services'. ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC), 2010
5 Merkle, P., Smolic, A., Müller, K., *et al*.: 'Efficient prediction structures for multi-view video coding', *IEEE Trans. Circuits Syst. Video Technol.*, 2007, **17**, (11), pp. 1461–1473
6 Zhao, T., Wang, H., Kwong, S., *et al*.: 'Fast mode decision based on mode adaptation', *IEEE Trans. Circuits Syst. Video Technol.*, 2010, **20**, (5), pp. 695–705
7 Paul, M., Lin, W., Lau, C.T., *et al*.: 'Direct intermode selection for H.264 video coding using phase correlation', *IEEE Trans. Image Process.*, 2011, **20**, (2), pp. 461–473
8 Zeng, H., Cai, C., Ma, K.-K.: 'Fast mode decision for H.264/AVC based on macroblock motion activity', *IEEE Trans. Circuits Syst. Video Technol.*, 2009, **19**, (4), pp. 491–499
9 Jung, S.-W., Baek, S.-J., Park, C.-S., *et al*.: 'Fast mode decision using all-zero block detection for fidelity and spatial scalable video coding', *IEEE Trans. Circuits Syst. Video Technol.*, 2010, **20**, (2), pp. 201–206
10 Wang, H., Kwong, S., Kok, C.W.: 'An efficient mode decision algorithm for H.264/AVC encoding optimization', *IEEE Trans. Multimedia*, 2007, **9**, (4), pp. 882–888
11 Zeng, H., Ma, K.-K., Cai, C.: 'Fast mode decision for multiview video coding using mode correlation', *IEEE Trans. Circuits Syst. Video Technol.*, 2011, **21**, (11), pp. 1659–1666
12 Zhu, W., Tian, X., Zhou, F., *et al*.: 'Fast inter mode decision based on textural segmentation and correlations for multiview video coding', *IEEE Trans. Consum. Electron.*, 2010, **56**, (3), pp. 1696–1704
13 Yu, T., Zhang, Y., Cosman, P.C.: 'Classification based fast mode decision for stereo video coding'. Proc. Int. Conf. Image Processing, Melbourne, VIC, Australia, September 2013, pp. 1724–1728
14 Shen, L., Liu, Z., Yan, T., *et al*.: 'Early SKIP mode decision for MVC using inter-view correlation', *Signal Process., Image Commun.*, 2010, **25**, (2), pp. 88–93
15 Wang, F., Zeng, H., Shen, Q., *et al*.: 'Efficient early direct mode decision for multi-view video coding', *Signal Process., Image Commun.*, 2013, **28**, (7), pp. 736–744
16 Zhang, Y., Kwong, S., Jiang, G., *et al*.: 'Statistical early termination model for fast mode decision and reference frame selection in multiview video coding', *IEEE Trans. Broadcast.*, 2012, **58**, (1), pp. 10–23
17 Vetro, A., McGuire, M., Matusik, W., *et al*.: 'Multiview video test sequences from MERL for the MPEG multiview working group'. ISO/IEC JTC1/SC29/WG11, Doc. M12077, 2005
18 Zitnick, C.L., Kang, S.B., Uyttendaele, M., *et al*.: 'High-quality video view interpolation using a layered representation'. Proc. ACM SIGGRAPH Transactions on Graphics, New York, NY, USA, August 2004, pp. 600–608 (2004)
19 Tanimoto, M., Fujii, T., Fukushima, N.: '1D parallel test sequences for MPEG-FTV'. ISO/IEC JTC1/SC29/WG11, Doc.M15378, 2008
20 Huang, Y.W., Hsieh, B.Y., Chien, S.Y., *et al*.: 'Analysis and complexity reduction of multiple reference frames motion estimation in H.264/AVC', *IEEE Trans. Circuits Syst. Video Technol.*, 2006, **16**, (4), pp. 507–522
21 Bjontegaard, G.: 'Calculation of average PSNR differences between RD-curves'. VCEG 13th meeting, Doc. VCEG-M33, 2001
22 Hanhart, P., Ebrahimi, T.: 'Calculation of average coding efficiency based on subjective quality scores', *J. Vis. Commun. Image Represent.*, 2014, **25**, (3), pp. 555–564
23 Hu, S., Zhao, T., Wang, H., *et al*.: 'Fast inter-mode decision based on rate-distortion cost characteristics'. Proc. 11th Pacific Rim Conf. Multimedia (PCM), Shanghai, China, September 2010, vol. 2, pp. 145–155
24 Pan, Z., Zhang, Y., Kwong, S.: 'Efficient motion and disparity estimation optimization for low complexity multiview video coding', *IEEE Trans. Broadcast.*, 2015, **61**, (2), pp. 166–176
25 Feldmann, I., Mueller, M., Zilly, F., *et al*.: 'HHI test material for 3D video'. ISO/IEC JTC1/SC29/WG11, Doc. M15413, 2008
26 Kawada, R.: 'KDDI multiview video sequences for MPEG 3DAV use'. ISO/IEC JTC1/SC29/WG11, Doc. M10533, 2004