

Journal of Electronic Imaging

[SPIDigitalLibrary.org/jei](https://spiedigitallibrary.org/jei)

Global-local correlation-based early large-size mode decision for multiview video coding

Wei Zhu
Yayu Zheng
Peng Chen
Jiani Xie



Global-local correlation-based early large-size mode decision for multiview video coding

Wei Zhu,^{a,*} Yayu Zheng,^a Peng Chen,^a and Jiani Xie^b

^aZhejiang University of Technology, College of Information Engineering, No. 288 Liuhe Road, Hang Zhou 310023, China

^bState Intellectual Property Office, No. 6 Xitucheng Road, Beijing 100088, China

Abstract. Multiview video coding (MVC) is a recent extension of H.264/AVC, and it consumes huge encoding time to select the optimal macroblock (MB) mode, among different size candidate modes. As compared with the small-size mode (Inter16 × 8, Inter8 × 16, Inter8 × 8, Intra8 × 8, and Intra4 × 4), the large-size mode (Skip/Direct, Inter16 × 16, and Intra16 × 16) occupies most of the MB mode proportion with much less computational complexity. Thus, if the large-size mode could be early decided as the optimal MB mode, the complexity of mode decision could be effectively reduced. In this work, an early large-size mode decision algorithm is proposed based on the global correlation of rate-distortion (RD) costs between neighbor views and the local correlation of RD costs among candidate modes. Average RD costs of large-size and small-size MB modes in the neighbor view are employed as a global reference for the threshold of early decision. And RD costs of estimated modes are used to calculate the local adjustment for the threshold. Experimental results demonstrate that the proposed algorithm can significantly reduce the whole encoding time while maintaining an RD performance similar to that of the original MVC encoder. © The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: 10.1117/1.JEI.23.1.013027]

Keywords: multiview video coding; mode decision; early termination; global–local correlation.

Paper 13351 received Jun. 27, 2013; revised manuscript received Dec. 26, 2013; accepted for publication Jan. 28, 2014; published online Feb. 19, 2014.

1 Introduction

Multiview video is captured from a set of viewpoints, and it is useful in many multimedia applications, such as three-dimensional (3-D) television, free viewpoint television, and glass-free portable 3-D display. Multiview video coding (MVC) was developed for the storage and transmission of very large multiview video data, and it has been standardized as the extension of H.264/AVC.¹ Figure 1 shows an illustration of the basic prediction structure in the MVC reference software JMVC, where the group of picture (GOP) length is eight for each view. According to the backward compatibility, all views can be classified into a base view and some nonbase views. Only the base view (S0 in Fig. 1) is backward compatible with H.264/AVC; nonbase views (S1 and S2 in Fig. 1) are encoded with new coding tools for providing complete multiview video bitstreams.² Every GOP in each view includes one anchor frame and some nonanchor frames. Frames with the time instants T0 and T8 in Fig. 1 are anchor frames, and other frames are nonanchor frames. For the random access of video bitstreams, anchor frames are not allowed to adopt temporal prediction. To improve the compression performance, nonanchor frames adopt temporal prediction and interview prediction.³ In Fig. 1, solid arrows represent temporal prediction and dotted arrows represent interview prediction. According to prediction directions for nonanchor frames, all views can also be classified into temporal views and interview views. Nonanchor frames employ both temporal prediction and interview prediction for interview views, while they only employ temporal

prediction for temporal views. In Fig. 1, views S0 and S2 belong to temporal views, and view S1 belongs to interview views.

Like H.264/AVC, MVC also needs to select the optimal macroblock (MB) mode among multiple candidate modes for each frame, and it has five intermodes (Skip/Direct, Inter16 × 16, Inter16 × 8, Inter8 × 16, and Inter8 × 8) and three intramodes (Intra16 × 16, Intra8 × 8, and Intra4 × 4). Except for Skip/Direct, other intermodes are needed to consume a lot of time for motion/disparity estimation in reference frames. In MVC reference software JMVC, the exhaustive mode decision algorithm is used to select the optimal MB mode, and it checks all candidate modes in sequence. Besides, due to the use of interview prediction, the computational complexity of mode decision for single view is greater than that of H.264/AVC. Thus, the computational complexity of MVC is very high, and it has hindered the practical use for real-time and mobile applications.

Many state-of-the-art fast mode decision algorithms have been developed for H.264/AVC. Wu et al.⁴ presented a fast intermode decision by using of the spatial homogeneity and the temporal stationarity characteristic of video objects. Hu et al.⁵ proposed a fast intermode decision algorithm based on rate-distortion (RD) cost characteristics, which includes an early skip mode decision and a three-stage mode prediction. Sung and Wang⁶ introduced a multiphase classification scheme that builds a mode decision tree according to the clustering of RD costs. These algorithms can be employed to speed up the mode decision of MVC, and their ideas could also enlighten the design of fast mode decision algorithms for MVC. However, the complexity of MVC is still very high, and it could be further reduced by using characteristics of MVC. To address this issue, various mode decision

*Address all correspondence to: Wei Zhu, E-mail: weizhu@zjut.edu.cn

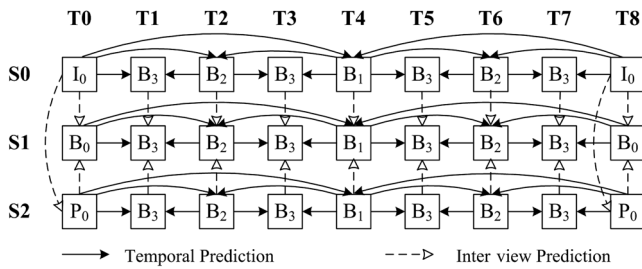


Fig. 1 Illustration of the basic prediction structure in JMVC. S0, S1, and S2 represent three views, T0 to T8 represent nine time instants. In each view, I, B, and P represent frame types, and their subscript values represent corresponding temporal levels.

algorithms have been studied, including Refs. 7–19. These algorithms presented some effective optimization techniques, such as adaptive termination strategy,^{7–10} candidate modes selection,^{10–13} prediction direction selection,^{12–14} and early Skip/Direct mode decision.^{14–18} For the reduction of the whole complexity, Shen et al.¹³ combined the candidate modes selection with the fast motion estimation and the prediction direction selection; meanwhile, Khattak et al.¹⁹ provided a complete framework that includes not only mode decision but also reference frame selection and fast motion/disparity estimation. In these algorithms, the correlation of coding information between neighbor views and the RD costs of MB mode in MVC are usually employed to arrive at a faster mode decision. As Skip/Direct mode occupies the largest proportion of MB modes with negligible computational complexity, several algorithms only focused on early Skip/Direct mode decision. Zeng et al.¹⁵ introduced an early decision algorithm by using RD costs of nearby MBs. Zatt et al.¹⁶ proposed an early decision algorithm based on mode correlation in the 3-D-neighborhood. Shen et al.¹⁷ presented an early decision algorithm based on the analysis of prediction mode distribution regarding the corresponding MBs in the neighbor view. Zhang et al.¹⁸ proposed an efficient statistical Skip/Direct mode termination model named SDMET to adjust the RD cost threshold adaptively by using statistical information of coded MBs. As mentioned above, early Skip/Direct mode decision algorithms can reduce the complexity effectively with high RD performance, and they also can be combined with fast algorithms of motion/disparity estimation and multireference frames selection^{19–23} to further reduce the complexity. However, since they mainly utilize the local correlation of coding information between neighbor views, the global correlation of coding information has not been exploited. Moreover, they cannot perform very well for video scenes with fast motions and large disparities, and they also have not considered the complexity reduction for anchor frames.

In this article, an early large-size mode decision algorithm based on the global-local correlation of RD costs is proposed to reduce the computational complexity of MVC. According to mode sizes and RD properties, all candidate modes in mode decision are classified into two types: the large-size mode and the small-size mode. The large-size mode includes Skip/Direct, Inter 16×16 , and Intra 16×16 , and the small-size mode includes Inter 16×8 , Inter 8×16 , Inter 8×8 , Intra 8×8 , and Intra 4×4 , where Inter 8×8 further contains four submodes (sub 8×8 , sub 8×4 , sub 4×8 , and sub 4×4). Compared with the small-size mode, the large-size mode

occupies much more MB mode proportion with much less computational complexity. Because of including Inter 16×16 and Intra 16×16 modes, the large-size mode also occupies more proportion than Skip/Direct mode, especially for frames with fast motions and large disparities. Therefore, the proposed algorithm focuses on the early decision of large-size mode instead of the early decision of Skip/Direct mode. The global correlation of RD costs between views is adopted to calculate the basis portion of the early decision threshold. For each MB, the local correlation of RD costs among different size modes is employed to calculate the adjustable portion of the threshold, and the minimal RD cost of the large-size mode is compared with the threshold to early terminate mode decision. In addition, the proposed algorithm also considers the optimization for frames in the base view by using the nearest forward-coded frame on temporal direction. Therefore, the proposed algorithm can be applied to all interframes of all views for effectively reducing the whole computational complexity of MVC.

The rest of this article is organized as follows: In Sec. 2, the characteristics of large-size and small-size modes are analyzed. Then, an early large-size mode decision based on the global-local correlation of RD costs is proposed in Sec. 3. Experimental results and conclusions are given in Secs. 4 and 5, respectively.

2 Motivation and Analysis

As an extension of H.264/AVC, MVC also employs the RD optimization technique²⁴ to select the optimal MB mode with the minimum RD cost, among candidate inter- or intramodes, and RD cost of each mode includes rate portion and distortion portion. Both inter and intramodes include different size modes, which have distinct characteristics of rate portion and distortion portion. By providing more precise motion/disparity estimation, small-size intermodes which include Inter 16×8 , Inter 8×16 , and Inter 8×8 can obtain less distortion portions of RD costs than large-size intermodes (Skip/Direct, Inter 16×16). However, because small-size intermodes need more bits for encoding motion vector of each partition block, they have larger rate portions of RD costs than that of large-size intermodes. For MBs with complex motion/disparity, small-size intermodes can obtain a much less distortion portion of RD cost than large-size intermodes, and they are more likely to be selected as the optimal MB mode. For MBs with smooth motion/disparity, large-size intermodes usually can provide the same level performance of distortion portions as small-size modes, and they have less rate portions than small-size intermodes due to encoding only 16×16 partition information. Similar to intermodes, small-size intramodes (Intra 8×8 and Intra 4×4) can also obtain less distortion portions and more rate portions than that of large-size intramode (Intra 16×16). For MBs with complex textures, small-size intramodes will often have less RD costs than large-size intramode. For MBs with smooth texture, large-size intramode usually obtains less RD cost than small-size intramode. In this article, according to the RD properties of different size modes, all intermodes and intramodes are classified into the large-size mode (includes Skip/Direct, Inter 16×16 , and Intra 16×16 modes) and the small-size mode (includes

Inter16 × 8, Inter8 × 16, Inter8 × 8, Intra8 × 8, and Intra4 × 4 modes).

In order to verify the above theoretical analysis, JMVC 8.0 was used to investigate statistic characteristics of large-size and small-size modes, and three typical test sequences (“Exit,” “Ballroom,” and “Race1”) with eight views (S0 to S7) were chosen. The GOP length was set to 12, and five GOPs were selected for each view. For each frame, the default coding conditions were used: maximum two reference frames are available for forward reference list and backward reference list, respectively, only one interview reference frame is allowed for each reference list, search method “TZ search” is enabled with search range 96. Table 1 gives the MB proportion of large-size and small-size modes, and the experimental basis quantization parameter (QP) is 32. It can be seen that the large-size mode occupies a proportion of 85% to 97%, which is much larger than the small-size mode. The mode proportion inside large-size MB mode under different basis QPs for “Race1” sequence is further shown in Fig. 2. It can be observed that Skip/Direct occupies the largest proportion, while Inter16 × 16 occupies the second largest proportion and Intra16 × 16 occupies a considerable proportion. Furthermore, the encoding time proportions of large-size and small-size modes are given in Table 2. The large-size mode consumes 7% to 15% of the encoding time for three sequences over eight views, whereas the small-size mode consumes 83% to 86% of the encoding time, which is much greater than that of the large-size mode. From Tables 1 and 2, it can be found that the large-size mode occupies a majority of the proportion, but only consumes a small part of the encoding time. Thus, if the large-size mode can be identified early as the optimal MB mode, the estimation of small-size mode can be skipped. In addition, it also indicated that the statistical proportions of MB modes are similar, among various views for each sequence. This is because those different

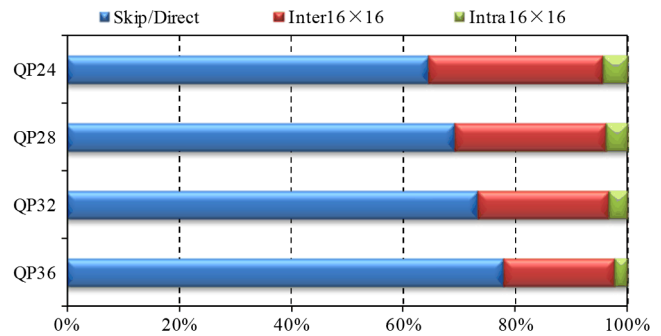


Fig. 2 Mode proportion inside large-size macroblock (MB) mode under different basis quantization parameters (QPs) for view S1 of “Race1” sequence.

views’ video are originated from the same scene, and the effect of occlusions between views is small in most cases. So, the statistical information of neighboring views can be adopted for the early decision of large-size mode.

Further studies on RD cost characteristics of large-size mode and small-size mode are performed, and experimental conditions are the same as Table 1. RD costs distribution of large-size and small-size MB modes in a single frame is shown in Fig. 3. It can be seen that most of large-size MB mode RD costs are less than the average RD cost of small-size MB mode in Fig. 3(a), and most of small-size MB mode RD costs are larger than the average RD cost of large-size MB mode in Fig. 3(b). These indicate that most of large-size MB modes have relatively low RD costs and most of small-size MB modes have relatively high RD costs, which are consistent with our previous theoretical analysis, and it can draw the conclusion that the average RD cost of large-size MB modes and the average RD cost of small-size MB modes can be used to calculate the thresholds for the early decision of large-size mode.

Table 1 Macroblock (MB) mode proportion for “Exit,” “Ballroom,” and “Race1” sequences.

View IDs	Proportion of large-size MB mode (%)			Proportion of small-size MB mode (%)		
	Exit	Ballroom	Race1	Exit	Ballroom	Race1
S0	97.3	89.8	89.7	2.7	10.2	10.3
S1	95.9	89.5	92.0	4.1	10.5	8.0
S2	94.0	85.6	87.8	6.0	14.4	12.2
S3	95.3	89.8	91.5	4.7	10.2	8.5
S4	92.8	85.4	87.2	7.2	14.6	12.8
S5	94.0	89.6	91.4	6.0	10.4	8.6
S6	89.7	85.0	86.6	10.3	15.0	13.4
S7	90.6	86.8	90.0	9.4	13.2	10.0
Avg.	93.7	87.7	89.5	6.3	12.3	10.5

Table 2 Encoding time proportion for “Exit,” “Ballroom,” and “Race1” sequences.

View IDs	Encoding time proportion of large-size mode (%)			Encoding time proportion of small-size mode (%)		
	Exit	Ballroom	Race1	Exit	Ballroom	Race1
S0	14.4	13.3	7.7	83.3	84.9	84.7
S1	14.6	13.8	9.0	85.0	85.5	85.1
S2	14.8	13.7	10.1	83.0	85.2	84.5
S3	14.4	13.7	12.7	84.8	85.2	85.7
S4	13.9	13.7	10.1	83.4	85.3	84.6
S5	14.0	14.2	11.8	85.0	85.4	85.5
S6	13.8	13.5	8.9	83.8	85.3	84.6
S7	13.2	14.8	11.5	85.3	85.3	84.9
Avg.	14.1	13.8	10.2	84.2	85.3	85.0

Note: Experimental conditions are the same as Table 1.

Moreover, the RD cost gap between large-size mode and small-size mode is further analyzed. Figure 4 shows the sorting of large-size mode RD costs with the red curve, and the corresponding small-size mode RD costs with same sorting indexes are also given. It can be seen

that most large-size mode RD costs are smaller than their corresponding small-size mode RD costs. This is consistent with the statistical results in Table 1 and Fig. 3. If the large-size mode RD cost is relatively low, the large-size mode is likely to be selected as the optimal MB mode. The RD cost gap between the large-size mode and the small-size mode is increasing with the large-size mode RD cost. Therefore, if the large-size mode RD cost is a small value, the misjudgment cost of early termination would also be small.

Characteristics of large-size and small-size modes in MVC are summarized as follows:

1. The large-size mode occupies most of the proportion of MB modes, while it consumes only a small part of the encoding time. The small-size mode occupies a smaller proportion of MB modes, but consumes most of the encoding time.

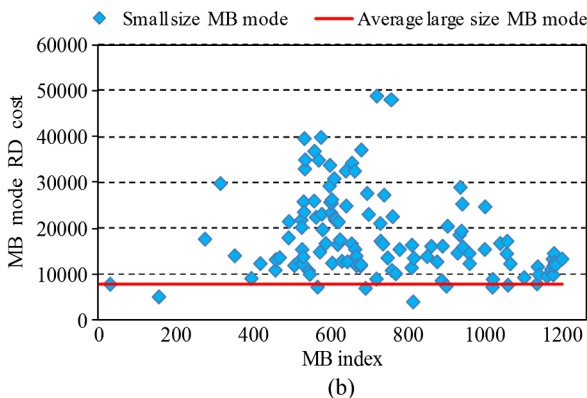
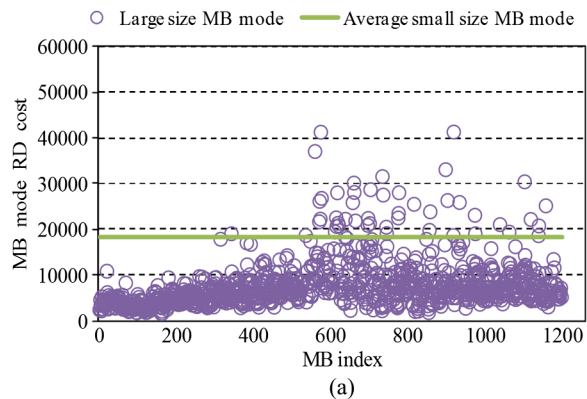


Fig. 3 Distribution of MB mode rate-distortion (RD) costs on view S1 frame 6 for “Ballroom” sequence. (a) RD costs of large-size MB mode and the average RD cost of small-size MB mode. (b) RD costs of small-size MB mode and the average RD cost of large-size MB mode.

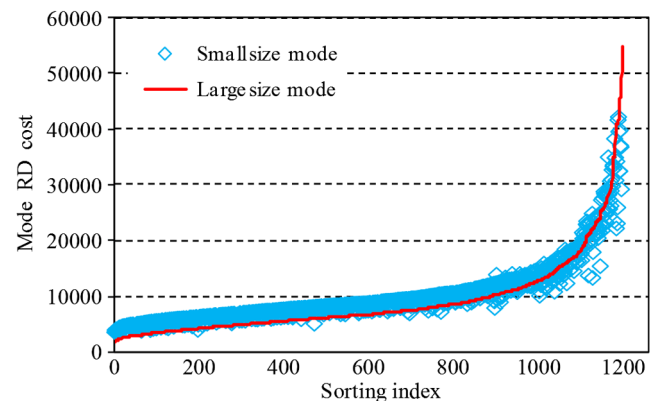


Fig. 4 Sorting of large-size mode RD costs and their corresponding small-size mode RD costs on view S1 frame 6 for the “Ballroom” sequence.

2. Although Skip/Direct occupies the largest proportion inside large-size MB mode, the proportions of Inter 16×16 and Intra 16×16 are also considerable.
3. Compared with small-size mode RD costs, most of large-size mode RD costs are relatively low, and their gaps are small when RD costs are low.
4. Different test sequences have various MB mode proportions, while different views in the same test sequence have similar MB mode proportions as they originate from the same scene. Thus, there is a global correlation of coding information between views, and the average RD cost of large-size MB modes and the average RD cost of small-size MB modes in neighboring views can be used for the mode decision of current view.

3 Proposed Early Large-Size Mode Decision Algorithm

According to the analysis in the previous section, the proposed algorithm focuses on the early decision of large-size mode, which is estimated first, then, its RD cost is compared with a global-local adaptive threshold to early terminate mode decision. The detail processes are introduced as follows.

Based on the motivation in Fig. 3, the average RD cost of large-size MB modes ($AvgJ_{Large}$) and the average RD cost of small-size MB modes ($AvgJ_{Small}$) of the coded frame are employed. For nonbase views, the coded frame in the forward neighboring view with the same time instant as the current frame is selected to calculate $AvgJ_{Large}$ as follows:

$$AvgJ_{Large} = \frac{\sum_{i=1}^{N_{Large}} J_{Large}(i)}{N_{Large}}, \quad (1)$$

where J_{Large} is the large-size mode RD cost of the MB i with large-size mode as its optimum MB mode, N_{Large} is the number of J_{Large} . $AvgJ_{Small}$ can be calculated in the same way. Due to the global correlation between views, $AvgJ_{Large}$ and $AvgJ_{Small}$ are adaptive to the video scene and coding feature. Therefore, $AvgJ_{Large}$ and $AvgJ_{Small}$ of neighboring views can be employed as a global measure for the early decision algorithm of current view. The curves of $AvgJ_{Large}$ and $AvgJ_{Small}$ under different basis QPs for “Ballroom” sequence are shown in Fig. 5. It can be seen that both $AvgJ_{Large}$ and $AvgJ_{Small}$ increase with basis QPs. Because the calculation of RD cost is relative to QP, $AvgJ_{Large}$, and $AvgJ_{Small}$ are adaptive to the change of QP, and they are suitable to be employed as the global reference. For each MB, the estimation of Skip/Direct, Inter 16×16 , and Intra 16×16 are performed, and the minimum RD cost of these estimated modes is selected as the large-size mode RD cost J_{Large} . Then, the early decision of large-size mode is determined in Eq. (2):

$$EarlyDecision(n) = \begin{cases} 1, & \text{if } J_{Large}(n) < EarlyTH \\ 0, & \text{otherwise} \end{cases}, \quad (2)$$

where n is the index of the current MB, and $EarlyTH$ is the early decision threshold. If J_{Large} is smaller than the termination threshold, $EarlyTH$, the optimal large-size mode with the minimum RD cost is selected early as the final MB mode and mode decision process is terminated. The selection of

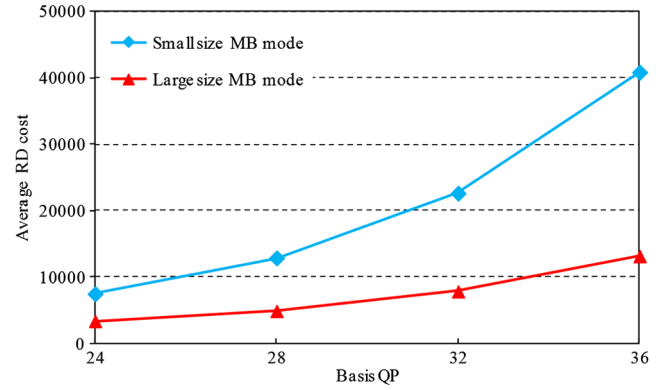


Fig. 5 Average MB mode RD costs on view S1 under different basis QPs for “Ballroom” sequence.

$EarlyTH$ directly affects the performance of the proposed algorithm, and its calculation is analyzed as follows:

First, $AvgJ_{Large}$ of the coded frame can be adopted as a measurement for J_{Large} in the current frame, and it is multiplied with a parameter α as the $EarlyTH$. To study the relation between parameter α and the variation of RD performance which is caused by the misjudgment of large-size mode, the increments of total RD costs for MBs, which select small-size mode as the optimal mode are shown in Fig. 6, where the experimental basis QP is 32. It can be seen that the increments of RD costs are very small when α is 1, while the increments obviously grow when α is larger than 1. Thus, $AvgJ_{Large}$ (α is equal to 1) can be employed as the basis portion of the threshold. Although $AvgJ_{Large}$ is adaptive to video content of the current frame, the gap between $AvgJ_{Large}$ and $AvgJ_{Small}$ also changes with basis QPs in Fig. 5. If the threshold is directly calculated by multiplying J_{Large} with a fixed parameter, the performance of early termination will not be stable. The threshold may be larger than $AvgJ_{Small}$ under small basis QPs, then the RD performance of the early decision will drop dramatically due to the low-decision accuracy. And the threshold also may be much less than $AvgJ_{Small}$ under large basis QPs, which will lead to low time savings because of the small early termination ratio. So, a good threshold calculation method that solves the above problem is expected in the proposed algorithm.

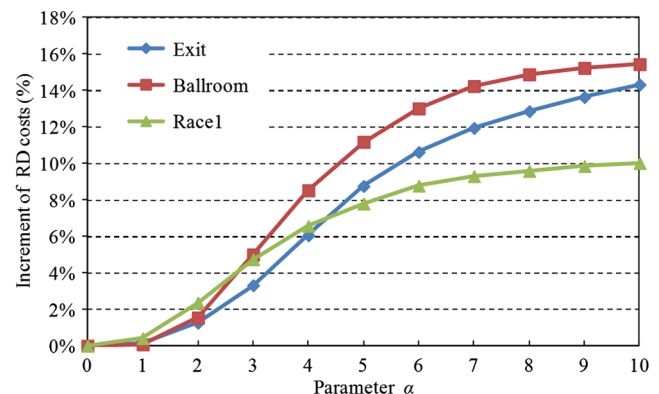


Fig. 6 The relationship between the parameter α and the increment of total RD costs for MBs with small-size MB mode.

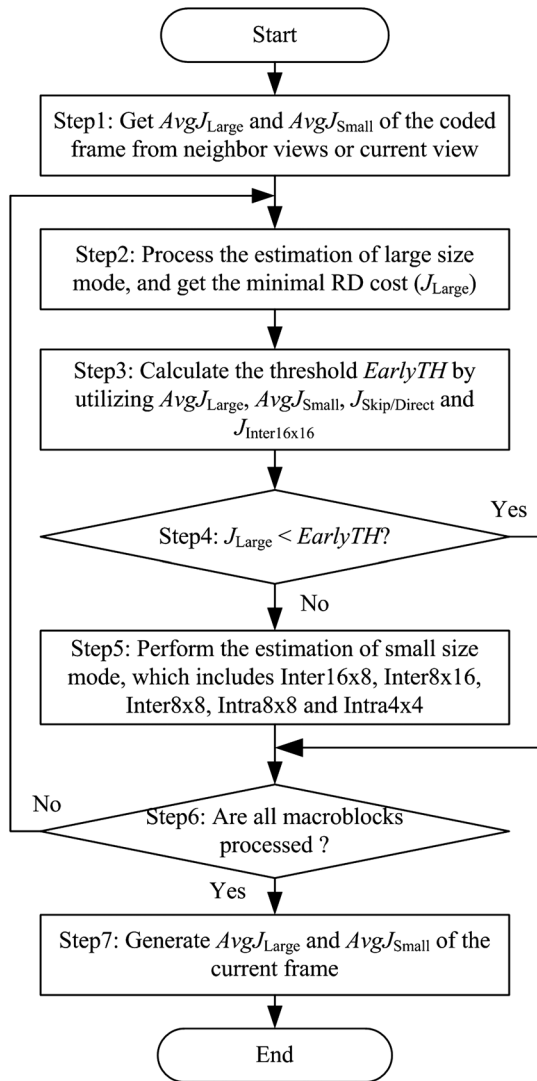


Fig. 7 Flow diagram of the proposed early large-size mode decision algorithm for multiview video coding.

Second, the threshold is further adjusted by utilizing the local feature of MB to improve the computational performance while maintaining a high RD performance. After the estimation of large-size mode, its RD costs reflect the local feature of current MB, and it can be used to predict the probability of a small-size MB mode. To reduce the misjudgment of large-size MB mode, MBs which selected the small-size mode as the final MB mode are employed to study the relation between large-size mode and small-size mode. For these MBs, Skip/Direct mode RD cost ($J_{\text{Skip/Direct}}$) and Inter16×16 mode RD cost ($J_{\text{Inter16} \times 16}$) are compared, and ratios of the smaller RD cost are given in Table 3, where the experimental conditions are same with Sec. 2. It can be seen that $J_{\text{Inter16} \times 16}$ occupies a majority of the proportion. Therefore, if $J_{\text{Inter16} \times 16}$ is less than $J_{\text{Skip/Direct}}$, there is more probability for current MB to select the small-size mode, and the threshold should be decreased for maintaining RD performance. Conversely, if $J_{\text{Skip/Direct}}$ is less than $J_{\text{Inter16} \times 16}$, the threshold could be increased for achieving more time saving.

Based on the above analysis, $EarlyTH$ in Eq. (2) is finally calculated as follows for each MB:

Table 3 Proportions of the smaller rate-distortion (RD) cost between $J_{\text{Skip/Direct}}$ and $J_{\text{Inter16} \times 16}$.

Sequences	Basis quantization parameters (QPs)	Proportions of the smaller RD cost (%)	
		$J_{\text{Skip/Direct}}$	$J_{\text{Inter16} \times 16}$
Exit	24	15.5	84.5
	28	14.3	85.7
	32	14.8	85.2
	36	15.6	84.4
Ballroom	24	17.6	82.4
	28	17.0	83.0
	32	17.5	82.5
	36	18.0	82.0
Race1	24	18.2	81.8
	28	17.6	82.4
	32	16.9	83.1
	36	17.9	82.1
Average		16.7	83.3

$$\begin{aligned}
 EarlyTH = & AvgJ_{\text{Large}} + \frac{J_{\text{Inter16} \times 16}}{J_{\text{Inter16} \times 16} + J_{\text{Skip/Direct}}} \\
 & \times (AvgJ_{\text{Small}} - AvgJ_{\text{Large}}), \quad (3)
 \end{aligned}$$

where $AvgJ_{\text{Large}}$ is the basis portion of $EarlyTH$, and the proportional divisor calculated by $J_{\text{Inter16} \times 16}$ and $J_{\text{Skip/Direct}}$ is used to adjust the gap between $AvgJ_{\text{Small}}$ and $AvgJ_{\text{Large}}$ with the local feature of current MB. In Eq. (3), if $J_{\text{Inter16} \times 16}$ is less than $J_{\text{Skip/Direct}}$, $EarlyTH$ is closer to $AvgJ_{\text{Large}}$, and if $J_{\text{Skip/Direct}}$ is less than $J_{\text{Inter16} \times 16}$, $EarlyTH$ is closer to $AvgJ_{\text{Small}}$. Thus, the range of $EarlyTH$ is from $AvgJ_{\text{Large}}$ to $AvgJ_{\text{Small}}$, and its value is dependent on the local MB coding feature. For different video sequences, a few frames may have no small-size MB mode under high basis QPs. In this particular case, $AvgJ_{\text{Small}}$ is replaced with $5 \times AvgJ_{\text{Large}}$ according to extensive experiments for getting a preferable performance.

In addition, the proposed algorithm is also extended to the base view by using $AvgJ_{\text{Small}}$ and $AvgJ_{\text{Large}}$ of the nearest forward-coded frame in the current view. Thus, it can reduce the computational complexity of all views. Besides, the proposed algorithm need not to store all MB RD costs of the coded frame because the use of global correlation, and only two classified average RD costs of current frame are stored for the optimization of following frames.

To verify the effectiveness of the proposed algorithm, simulations have been conducted under the same test conditions as Sec. 2. Table 4 gives decision accuracies and

Table 4 Decision accuracies and termination ratios of large-size MB mode under different thresholds.

Sequences	Basis QPs	Decision accuracies of large-size MB mode under different thresholds (%)			Termination ratios of large-size MB mode under different thresholds (%)			Proportion of large-size MB mode
		<i>EarlyTH</i>	<i>AvgJ_{Large}</i>	<i>AvgJ_{Small}</i>	<i>EarlyTH</i>	<i>AvgJ_{Large}</i>	<i>AvgJ_{Small}</i>	
Exit	24	96.8	98.6	92.7	79.3	50.2	90.1	86.7
	28	98.4	99.6	95.2	83.8	55.0	93.4	90.9
	32	99.0	99.7	96.7	86.7	58.7	95.4	93.7
	36	99.3	99.8	97.9	89.0	61.2	96.6	95.8
Ballroom	24	93.5	96.8	87.9	76.1	51.4	86.9	79.2
	28	95.8	98.4	90.8	78.9	54.2	89.5	83.7
	32	97.3	99.3	93.3	81.6	56.3	91.6	87.7
	36	98.3	99.6	95.5	84.1	58.9	93.4	91.3
Race1	24	90.8	93.6	86.8	74.7	55.1	85.4	79.4
	28	94.4	96.8	90.9	78.6	56.8	88.9	85.3
	32	96.6	98.5	93.6	81.8	57.7	91.9	89.5
	36	97.7	99.2	95.4	84.4	58.5	93.9	92.4
Average		96.5	98.3	93.1	81.6	56.2	91.4	88.0

termination ratios of the proposed algorithm under different thresholds (*EarlyTH*, *AvgJ_{Large}*, and *AvgJ_{Small}*). The decision accuracy and termination ratio are defined as follows:

$$\begin{cases} \text{Accuracy}(\%) = \frac{N_{\text{Hit}}}{N_{\text{Early}}} \times 100 \\ \text{Ratio}(\%) = \frac{N_{\text{Early}}}{N_{\text{MB}}} \times 100 \end{cases}, \quad (4)$$

where N_{MB} represents the number of MBs, N_{Early} represents the number of early termination in Eq. (2), and N_{Hit} is the number of MBs which early select large-size MB mode in Eq. (2), and their optimal MB modes are also large-size mode. For the threshold *EarlyTH*, the average decision accuracy is 96.5%, and the average termination ratio is 81.6% that is close to the average proportion of large-size MB mode. This indicate that *EarlyTH* can achieve large termination ratios with high decision accuracies. For threshold *AvgJ_{Large}*, the average decision accuracy is 98.3%, while the average termination ratio is only 56.2%, which is 31.8% less than the average proportion of large-size MB mode. For the threshold *AvgJ_{Small}*, all termination ratios are larger than the corresponding proportion of large-size MB mode, which leads to only 93.1% average decision accuracy. These results demonstrate that *EarlyTH*, which is calculated using global-local RD costs, is more suitable for the proposed algorithm than thresholds only using global RD costs.

Table 5 Experimental configuration of JMVC.

Basis QP	24, 28, 32, 36
Delta QP values	0, 3, 4, 5, 6, 7
View order	0-2-1-4-3-6-5-7
Group of picture (GOP) length	12
Frames coded	5 GOPs × 8 views
Entropy coding	CABAC
RDO	On
Search mode	4 (TZ search)
Search range	96
Bi prediction iteration	4
Iteration search range	8
Reference number	2

A flow diagram of the proposed algorithm is illustrated in Fig. 7, and the detailed steps of proposed algorithm are given as follows:

1. If the current frame is an intraframe, perform the estimation of all intramodes, then go to step 7. Otherwise, get $AvgJ_{Large}$ and $AvgJ_{Small}$ from the coded frame for the current frame. If the current frame belongs to non-base views, then the coded frame with the same time instant in the forward neighboring view is selected. If the current frame belongs to the base view, then the nearest forward-coded frame in the current view is employed.
2. Process the estimation of large-size mode (Skip/Direct, Inter16 × 16, and Intra16 × 16) for current MB, and obtain Skip/Direct mode RD cost ($J_{Skip/Direct}$), Inter16 × 16 mode RD cost ($J_{Inter16 \times 16}$), and Intra16 × 16 mode RD cost. Then select the optimal large-size mode with the minimal RD cost (J_{Large}).
3. Calculate the threshold $EarlyTH$ in Eq. (3) by utilizing $AvgJ_{Large}$, $AvgJ_{Small}$, $J_{Skip/Direct}$, and $J_{Inter16 \times 16}$.
4. Perform the early large-size mode decision in Eq. (2). If the J_{Large} is less than $EarlyTH$, then the optimal large-size mode is decided as the final MB mode, go to step 6.

Table 6 Performance of proposed algorithm under different basis QPs.

Sequences	Basis QP	$\Delta Time (\%) / \Delta PSNR (dB) / \Delta Bit (\%)$		Sequences	Basis QP	$\Delta Time (\%) / \Delta PSNR (dB) / \Delta Bit (\%)$	
		Temporal views	Interview views			Temporal views	Interview views
Exit	24	48.5/ -0.018/ -0.19	67.9/ -0.020/ 0.24	Break-dancers	24	54.4/ -0.033/ -0.05	64.6/ -0.044/ 0.07
	28	53.6/ -0.019/ -0.26	69.7/ -0.024/ -0.09		28	57.1/ -0.028/ -0.20	65.6/ -0.031/ -0.28
	32	58.0/ -0.019/ -0.09	70.7/ -0.031/ -0.08		32	58.6/ -0.012/ -0.03	66.1/ -0.021/ -0.42
	36	60.2/ -0.012/ -0.10	71.3/ -0.038/ -0.53		36	58.5/ -0.009/ -0.04	65.3/ -0.017/ -0.02
	Avg.	55.1/ -0.017/ -0.16	69.9/ -0.028/ -0.12		Avg.	57.2/ -0.020/ -0.08	65.4/ -0.028/ -0.16
Ballroom	24	45.2/ -0.035/ 0.00	66.3/ -0.044/ 0.36	Door-flowers	24	69.0/ -0.017/ 0.10	75.6/ -0.027/ 0.53
	28	47.7/ -0.040/ -0.06	66.7/ -0.047/ 0.05		28	69.6/ -0.026/ -0.21	74.5/ -0.028/ 0.39
	32	50.5/ -0.043/ -0.18	67.0/ -0.045/ -0.06		32	69.1/ -0.016/ 0.16	73.4/ -0.016/ 0.42
	36	53.1/ -0.044/ -0.14	66.9/ -0.045/ -0.09		36	68.9/ -0.012/ 0.14	72.3/ -0.015/ 0.33
	Avg.	49.1/ -0.040/ -0.10	66.7/ -0.045/ 0.06		Avg.	69.2/ -0.018/ 0.05	74.0/ -0.022/ 0.42
Race1	24	63.4/ -0.050/ 0.08	73.2/ -0.067/ 0.03	Lovebird1	24	69.6/ -0.020/ -0.01	79.0/ -0.020/ 0.25
	28	64.7/ -0.051/ -0.02	74.0/ -0.066/ -0.40		28	71.4/ -0.030/ -0.32	77.7/ -0.024/ 0.04
	32	65.6/ -0.033/ -0.31	73.9/ -0.045/ -0.63		32	72.0/ -0.045/ -0.55	74.3/ -0.038/ -0.07
	36	66.1/ -0.032/ -0.33	73.1/ -0.057/ -0.95		36	72.5/ -0.034/ -0.48	72.0/ -0.030/ -0.21
	Avg.	65.0/ -0.042/ -0.14	73.5/ -0.059/ -0.49		Avg.	71.3/ -0.032/ -0.34	75.7/ -0.028/ 0.00
Ballet	24	60.7/ -0.019/ 0.15	74.6/ -0.023/ 0.19	Dog	24	60.6/ -0.026/ -0.22	73.8/ -0.031/ -0.10
	28	62.6/ -0.014/ -0.10	73.8/ -0.025/ 0.12		28	62.4/ -0.036/ -0.29	73.6/ -0.036/ -0.32
	32	63.6/ -0.014/ -0.12	72.6/ -0.023/ -0.09		32	63.3/ -0.047/ -0.44	72.5/ -0.050/ -0.38
	36	63.9/ -0.005/ -0.18	70.9/ -0.023/ -0.24		36	63.9/ -0.044/ -0.46	70.8/ -0.056/ -0.52
	Avg.	62.7/ -0.013/ -0.06	73.0/ -0.024/ 0.00		Avg.	62.6/ -0.038/ -0.35	72.6/ -0.043/ -0.33
Average performance of temporal views				Average performance of interview views			
61.5/ -0.028/ -0.15				71.4/ -0.035/ -0.08			

5. Perform the estimation of small-size mode (Inter16 × 8, Inter8 × 16, Inter8 × 8, Intra8 × 8, and Intra4 × 4). Then select the final optimal mode with the minimal RD cost, among all estimated modes.
6. If all MBs of current frame have been processed, go to step 7, else go to step 2 for the mode decision of the next MB.
7. Generate $AvgJ_{Large}$ and $AvgJ_{Small}$ of the current frame for the early large-size mode decision of subsequent encoding frames.

4 Experimental Results

The proposed algorithm was implemented on the MVC reference software JMVC 8.0, and its detailed configuration is given in Table 5. JMVC adopts the I-B-P prediction structure for view coding order and hierarchical B picture prediction structure for temporal coding order.^{3,25} Eight typical test sequences were chosen for the simulation, which include three 640 × 480 sequences (“Exit,” “Ballroom,” and “Race1”), four 1024 × 768 sequences (“Breakdancers,” “Ballet,” “Doorflowers,” and “Lovebird1”), and one 1280 × 960 sequence (“Dog”). These sequences are representative in camera setups, video scenes, and frame rates, and eight views (S0 to S7) were chosen for each sequence. Compared with the exhaustive mode decision in JMVC, the encoding time saving (Δ Time) was calculated to evaluate the computational performance, and the change of peak signal-to-noise ratio (Δ PSNR) and the change of bit rate (Δ Bits) were calculated to evaluate the RD performance under each QP. To evaluate the overall RD performance under four basis QPs, Bjontegaard delta peak signal-to-noise ratio (BDPSNR) and Bjontegaard delta bit rate (BDBR)²⁶ were also employed. A negative BDPSNR or a positive BDBR indicates a coding loss and is not preferred.

Because interview views employ not only temporal prediction but also interview prediction for nonanchor frames,

they consume more encoding time than temporal views. Table 6 gives the performances of the proposed algorithm for temporal views and interview views separately. For temporal views of the eight sequences, the proposed algorithm reduces encoding time from 49.1% to 71.3%, and the PSNR loss is from 0.013 to 0.042 dB with the change of bit rate from -0.35% to 0.05%. The average time saving of the eight sequences is 61.5%, while the average loss of PSNR is only 0.028 dB with a 0.15% decrement of average bit rate. For interview views of the eight sequences, the proposed algorithm reduces encoding time from 65.4% to 75.7%, and the loss of PSNR is from 0.022 to 0.059 dB with the change of bit rate from -0.49% to 0.42%. The average time saving of the eight sequences is 71.4%, while the average PSNR loss is only 0.035 dB with a 0.08% decrement of average bit rate. These demonstrate that the proposed algorithm achieves significant time saving while maintaining an RD performance similar to that of the original encoder. In Table 6, it can be seen that the proposed algorithm achieves about 10% more average time savings for interview views when compared with that for temporal views. This is because interview views consume more encoding time originally than temporal views. From Table 6, it can also be found that the proposed algorithm achieves more encoding time savings for the sequences only with a few motions, such as “Ballet,” “Doorflowers,” “Lovebird1,” and “Dog” sequences. This is because the original proportions of large-size MB mode in these sequences are larger than in other sequences, and the proposed algorithm saves more encoding time adaptively. For “Ballroom,” “Race1,” “Exit,” and “Breakdancer” sequences which have a lot of motions, the proposed algorithm also reduces the encoding time significantly from 65.4% to 73.5% for interview views.

BD-metric²⁶ was adopted to assess RD performances of the proposed algorithm and state-of-the-art mode decision algorithms. Table 7 gives the encoding time saving,

Table 7 Performance of the proposed algorithm and state-of-the-art algorithms under four basis QPs.

Sequences	Δ Time (%) / BDPSNR (dB) / BDBR (%)				
	Temporal views		Interview views		
	Proposed	SDMET in Ref. 18	Proposed	SDMET in Ref. 18	FIMD in Ref. 14
Exit	55.1 / -0.01 / 0.51	44.1 / -0.02 / 0.82	69.9 / -0.02 / 0.87	57.2 / -0.02 / 0.97	62.4 / 0.00 / 0.19
Ballroom	49.1 / -0.04 / 0.91	34.1 / -0.03 / 0.68	66.7 / -0.05 / 1.20	51.6 / -0.03 / 0.78	52.6 / -0.01 / 0.37
Race1	65.0 / -0.04 / 0.88	39.5 / -0.06 / 1.34	73.5 / -0.04 / 0.92	42.4 / -0.08 / 1.82	52.6 / -0.01 / 0.15
Ballet	62.7 / -0.01 / 0.33	54.5 / -0.02 / 0.95	73.0 / -0.02 / 0.67	65.6 / -0.02 / 0.97	65.0 / 0.00 / 0.03
Breakdancers	57.2 / -0.02 / 0.83	34.4 / -0.06 / 2.93	65.4 / -0.02 / 0.93	42.4 / -0.05 / 2.14	43.3 / 0.00 / 0.05
Doorflowers	69.2 / -0.02 / 0.71	64.3 / -0.06 / 2.20	74.0 / -0.04 / 1.19	65.6 / -0.07 / 2.55	67.2 / 0.00 / 0.09
Lovebird1	71.3 / -0.02 / 0.59	53.1 / -0.01 / 0.27	75.7 / -0.03 / 0.84	55.1 / -0.01 / 0.35	68.2 / 0.00 / 0.04
Dog	62.6 / -0.02 / 0.66	50.2 / -0.03 / 0.74	72.6 / -0.03 / 0.91	57.6 / -0.03 / 0.86	56.9 / -0.01 / 0.16
Average	61.5 / -0.02 / 0.68	46.8 / -0.04 / 1.24	71.4 / -0.03 / 0.94	54.7 / -0.04 / 1.31	58.5 / 0.00 / 0.13

BDPSNR and BDBR under four basis QPs for the proposed algorithm, and the state-of-the-art early Skip/Direct mode decision algorithm SDMET.¹⁸ For temporal views of the eight sequences, SDMET achieves a 46.8% time saving on average with a 0.04-dB BDPSNR loss and a 1.24% BDBR increment, and the proposed algorithm achieves a 14.7% more time saving with slightly better RD performance than SDMET. For interview views of the eight sequences, SDMET achieves a 54.7% time savings on average with a 0.04-dB BDPSNR loss and a 1.31% BDBR increment, and the proposed algorithm achieves a 16.7% more time savings with also a slightly better RD performance than SDMET. Table 7 also gives results of the fast intermode decision algorithm based on textural segmentation and correlations (indicated as FIMD), which has been presented in our previous work.¹⁴ FIMD includes an early Skip/Direct mode decision method and two assistant methods (selection of disparity estimation and the reduction of Inter8 × 8 mode estimation), and it was implemented on interview views. In Table 7, it can be seen that FIMD achieves a 58.5% time saving on average with a 0.00-dB BDPSNR loss and a 0.13% BDBR increment. The RD performance of FIMD almost maintains the same as the original JMVC, while the proposed algorithm has a 12.9% more time saving with a similar RD performance.

For a better observation, Fig. 8 shows the histogram comparison of the time savings among the proposed algorithm, SDMET and FIMD for interview views of different sequences. It can be observed that the proposed algorithm achieves larger time savings than both SDMET and FIMD over all sequences. For sequences which have lots of fast motions and large disparities, such as “Race1” and “Breakdancers,” the time savings of SDMET and FIMD are less significant than the proposed algorithm. For “Race1” sequence, the proposed algorithm achieves up to 30% more time saving than SDMET, and it also achieves up to 20% more time saving than FIMD. For sequences which have a few motions and simple textures, such as “Ballet” and “Doorflowers,” the time savings of SDMET and FIMD are close to the proposed algorithm. The above comparisons indicate that the proposed algorithm also has more stable time savings than these two

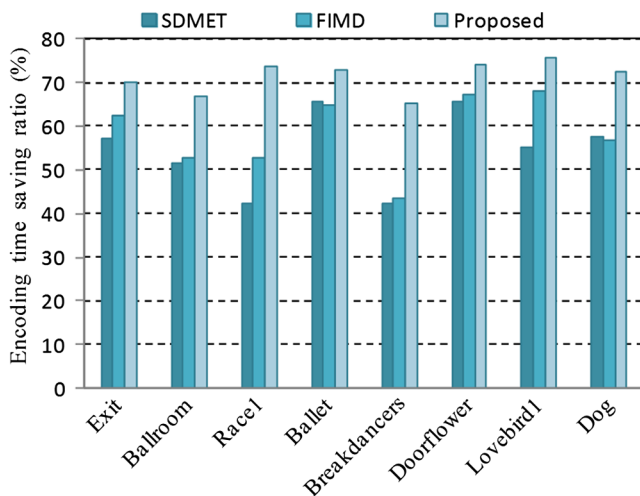


Fig. 8 Encoding time saving ratios of SDMET, FIMD, and the proposed algorithm for interview views of the eight sequences.

state-of-the-art algorithms. This is because that the original proportion of large-size MB mode is larger than the proportion of Skip/Direct MB mode for all sequences, especially for sequences with lots of fast motions. Thus, the performances of SDMET and FIMD are more sensitive to motions than the proposed algorithm. Additionally, SDMET and FIMD have not considered the optimization for anchor frames, which also affects their computational performances.

Moreover, Fig. 9 illustrates the time savings curves of the proposed algorithm, SDMET, and FIMD for interview views under different QPs. It can be seen that the proposed algorithm achieves a more stable and larger time savings than SDMET over different QPs. For “Race1” sequence, which has fast global motions, the proposed algorithm obtains about 30% more time savings than SDMET under all QPs. For “Dog” sequence, which has slow motions and small disparities, the time savings of SDMET is close to the proposed algorithm under higher QPs, while it has less time savings than the proposed algorithm under lower QPs. Due to the complex textures and picture noises in the static background of “Dog” sequence, the distribution of Skip/Direct MB mode is dispersive under lower QPs, and SDMET has lesser decision ratios for ensuring small RD degradation. With the increase of QPs, more MBs select Skip/Direct as the optimal MB mode, and SDMET also achieves much more time savings than that under lower QPs. Owing to the use of global RD costs obtained from neighbor views, the proposed algorithm has taken both QP and video contents

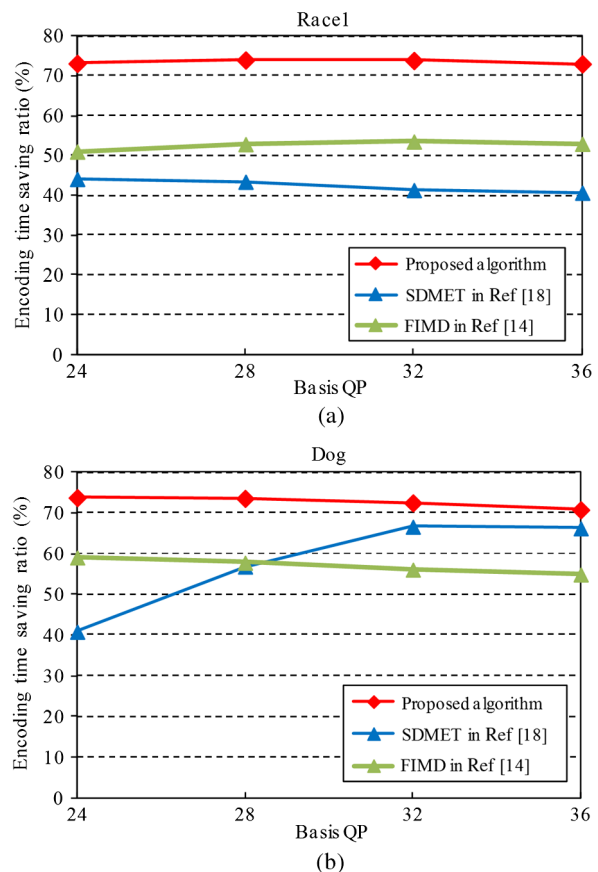


Fig. 9 Time saving curves of proposed algorithm, SDMET, and FIMD under different basis QPs for interview views. (a) Time saving curves for “Race1” sequence. (b) Time saving curves for “Dog” sequence.

Table 8 Performance of the proposed integrated scheme and VAMEDE under four basis QPs for interview views.

Sequences	Proposed algorithm + FDE in Ref. 21					VAMEDE in Ref. 13				
	Δ Time (%)	Δ PSNR (dB)	Δ Bit (%)	BDPSNR (dB)	BDBR (%)	Δ Time (%)	Δ PSNR (dB)	Δ Bit (%)	BDPSNR (dB)	BDBR (%)
Exit	87.5	-0.03	-0.04	-0.03	1.06	82.0	-0.06	1.95	-0.11	4.04
Ballroom	83.2	-0.05	0.31	-0.06	1.48	72.7	-0.05	3.19	-0.18	4.67
Race1	85.7	-0.06	-0.92	-0.01	0.34	75.4	-0.10	4.05	-0.27	6.87
Ballet	89.3	-0.05	-0.22	-0.04	1.12	83.5	-0.03	0.24	-0.04	1.11
Breakdancers	81.3	-0.03	-0.28	-0.03	1.06	71.1	-0.04	2.17	-0.10	3.98
Doorflowers	89.5	-0.03	0.22	-0.03	1.13	86.7	-0.02	1.51	-0.06	2.16
Lovebird1	91.1	-0.03	-0.10	-0.03	0.80	87.9	-0.01	0.13	-0.02	0.46
Dog	82.7	-0.05	-0.78	-0.02	0.55	77.3	-0.03	1.08	-0.07	1.97
Average	86.3	-0.04	-0.23	-0.03	0.94	79.6	-0.04	1.79	-0.10	3.16

into consideration, and its performance is not sensitive to QP and video contents. Compared with FIMD in our previous work,¹⁴ the proposed algorithm achieves about 20% more time saving under all QPs for “Race1” sequence, and it also achieves about 15% more time saving under all QPs for “Dog” sequence. It is because FIMD employs more strict decision conditions for maintaining the same RD performance as the original encoder, which leads to less time savings. Besides, the computational performance of FIMD depends largely on the proportion of Skip/Direct MB mode, which is originally less than the proportion of large-size mode. Therefore, the proposed algorithm can obtain more time savings.

Although the proposed algorithm can reduce about 71% of encoding time for interview views, the computational complexity of interview views is still very large due to the disparity estimation for interview prediction. To further reduce the complexity, the proposed algorithm can be combined with state-of-the-art fast search algorithms, and it was integrated with the fast disparity estimation algorithm (indicated as FDE) in our previous work²¹ for interview views. Table 8 gives the performance of the integrated scheme. It can be seen that the integrated scheme achieves 86.3% average time saving, which is about 15% more time savings than the proposed algorithm. And the average coding efficiency loss of all sequences is 0.04-dB PSNR loss and 0.23% bit rate decrement (0.03-dB BDPSNR loss and 0.94% BDBR increment), which is same with the average performance of the proposed algorithm in Table 7. The view-adaptive motion estimation and disparity estimation (VAMEDE),¹³ which includes mode size decision, fast motion estimation, and selective disparity estimation, was also implemented on JMVC for interview views, and it achieves 79.6% average time saving with 0.04-dB PSNR loss and 1.79% bit rate increment (0.10-dB BDPSNR loss and 3.16% BDBR increment). The average performance of VAMEDE in Table 8 is consistent with that mentioned in Ref. 13, and the reason of

the slightly coding efficiency degradation is as follows. VAMEDE is developed based on JMVC platform, which is the MVC reference software before JMVC. JMVC adopts the motion skip mode to highly improve the overall coding performance. However, the motion skip mode is excluded in JMVC, and VAMEDE would cause relatively larger coding efficiency loss on JMVC than that on JMVC. Compared with VAMEDE, our integrated scheme achieves about 6% more time saving on average with a same level PSNR loss and less bit rate increments. For “Lovebird1” and “Doorflowers” sequences which have a few smooth motions, our integrated scheme achieves about 3% more time saving than VAMEDE with similar coding efficiency loss. For “Race1” and “Ballroom” sequences which have a lot of complex motions, our integrated scheme achieves about 10% more time saving than VAMEDE with less coding efficiency loss.

5 Conclusion

To reduce the computational complexity of MVC, this work presents a fast mode decision algorithm which focuses on the early decision of large-size mode. Based on the global correlation of RD costs between views and the local correlation of RD costs among candidate modes, the average RD costs of large-size and small-size MB modes in the neighboring view are combined with large-size mode RD costs of the current MB for the early selection of large-size mode as the optimal MB mode. Compared with the exhaustive mode decision, experimental results show that the proposed algorithm saves the encoding time significantly with negligible loss of RD performance, and it also achieves a better performance than the state-of-the-art algorithms, especially for test sequences with fast motions and large disparities. Moreover, the proposed algorithm was integrated with the FDE of our previous work, and the integrated scheme also achieves better performance than the state-of-the-art algorithm VAMEDE.

Acknowledgments

This work was supported in part by the Zhejiang Provincial Natural Science Foundation of China under Grants LQ12F01008 and Y1110532 and the Natural Science Foundation of China under Grant 61303139. The authors would like to thank journal reviewers for their insightful comments and valuable suggestions.

References

1. ITU-T, and ISO/JTC1/EC 1, "Advanced video coding for generic audio-visual services," *ITU-T Recommendation H.264 and ISO/IEC 14496 (MPEG-4 AVC)* (2010).
2. A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/AVC standard," *Proc. IEEE* **99**(4), 626–642 (2011).
3. P. Merkle et al., "Efficient prediction structures for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.* **17**(11), 1461–1473 (2007).
4. D. Wu et al., "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.* **15**(7), 953–958 (2005).
5. S. Hu et al., "Fast inter-mode decision based on rate-distortion cost characteristics," in *Pacific-Rim Conf. Multimedia*, pp. 145–155, Springer, Shanghai (2010).
6. Y. Sung and J. Wang, "Fast mode decision for H.264/AVC based on rate-distortion clustering," *IEEE Trans. Multimedia* **14**(3), 693–702 (2012).
7. L. Ding et al., "Content-aware prediction algorithm with inter-view mode decision for multiview video coding," *IEEE Trans. Multimedia* **10**(8), 1553–1564 (2008).
8. B. Zatt et al., "A multi-level dynamic complexity reduction scheme for multiview video coding," in *IEEE Int. Conf. Image Process.*, pp. 749–752, IEEE, Brussels (2011).
9. T. Zhao et al., "Multiview coding mode decision with hybrid optimal stopping model," *IEEE Trans. Image Process.* **22**(4), 1598–1609 (2013).
10. L. Shen et al., "Low-complexity mode decision for MVC," *IEEE Trans. Circuits Syst. Video Technol.* **21**(6), 837–843 (2011).
11. H. Zeng, K. Ma, and C. Cai, "Fast mode decision for multiview video coding using mode correlation," *IEEE Trans. Circuits Syst. Video Technol.* **21**(11), 1659–1666 (2011).
12. L. Shen et al., "Selective disparity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," *IEEE Trans. Broadcast.* **55**(4), 761–766 (2009).
13. L. Shen et al., "View-adaptive motion estimation and disparity estimation for low complexity multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.* **20**(6), 925–930 (2010).
14. W. Zhu et al., "Fast inter mode decision based on textural segmentation and correlations for multiview video coding," *IEEE Trans. Consumer Electron.* **56**(3), 1696–1704 (2010).
15. H. Zeng, K. Ma, and C. Cai, "Mode-correlation-based early termination mode decision for multi-view video coding," in *IEEE Int. Conf. on Image Process.*, pp. 3405–3408, IEEE, Hongkong (2010).
16. B. Zatt et al., "An adaptive early skip mode decision scheme for multiview video coding," in *Picture Coding Symp.*, pp. 42–45, IEEE, Nagoya (2010).
17. L. Shen et al., "Early SKIP mode decision for MVC using inter-view correlation," *Signal Process.: Image Commun.* **25**(2), 88–93 (2010).
18. Y. Zhang et al., "Statistical early termination model for fast mode decision and reference frame selection in multiview video coding," *IEEE Trans. Broadcast.* **58**(1), 10–23 (2012).
19. S. Khattak et al., "Fast encoding techniques for multiview video coding," *Signal Process.: Image Commun.* **28**(6), 569–580 (2013).
20. L. Shen et al., "Macroblock-level adaptive search range algorithm for motion estimation in multiview video coding," *J. Electron. Imaging* **18**(3), 033003 (2009).
21. W. Zhu et al., "Fast disparity estimation using spatio-temporal correlation of disparity field for multiview video coding," *IEEE Trans. Consumer Electron.* **56**(2), 957–964 (2010).
22. Y. Zhang et al., "Efficient multi-reference frame selection algorithm for hierarchical B pictures in multiview video coding," *IEEE Trans. Broadcast.* **57**(1), 15–23 (2011).
23. Z. Deng et al., "Iterative search strategy with selective bi-directional prediction for low complexity multiview video coding," *J. Vis. Commun. Image R.* **23**, 522–534 (2012).
24. T. Wiegand et al., "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.* **13**(7), 688–703 (2003).
25. A. Vetro et al., "Joint Multiview Video Model (JMVM) 8.0," *ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16*, Doc. JVT-AA207, Geneva (2008).
26. G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *ITU-T Q6/SG16*, Doc. VCEG-M33, Austin (2001).

Wei Zhu received his BSc and PhD degrees from Zhejiang University, Hangzhou, China, in 2004 and 2010, respectively. He is currently a lecturer of Zhejiang University of Technology. His major research fields are video coding, video analysis, and parallel processing.

Yayu Zheng received his BSc and PhD degrees from Zhejiang University, Hangzhou, China, in 2002 and 2008, respectively. He is currently an associate professor of Zhejiang University of Technology. His major research fields are networking multimedia systems and video coding.

Peng Chen received his BSc and PhD degrees from Zhejiang University, Hangzhou, China, in 2003 and 2009, respectively. He is currently an associate professor of Zhejiang University of Technology. His major research fields are signal processing and system design for multimedia.

Jiani Xie received her BSc and MS degrees from Zhejiang University, Hangzhou, China, in 2004 and 2008, respectively. She is currently in State Intellectual Property Office. Her major research field is 3-D video processing.