Fast Mode Decision for Multiview Video Coding Using Mode Correlation

Huanqiang Zeng, Kai-Kuang Ma, *Senior Member, IEEE,* and Canhui Cai, *Senior Member, IEEE*

***Abstract*—Exhaustive mode decision has been exploited in multiview video coding for effectively improving the coding efficiency, but at the expense of yielding much higher computational complexity. In this paper, a fast mode decision algorithm, called the *mode correlation-based mode decision* (MCMD), is proposed to speed up the encoding process by reducing the number of the modes required to be checked. In our approach, all the prediction modes are first categorized into five motion-activity classes, and only one of them will be chosen to identify the optimal mode in a hierarchical manner, as follows. For each *macroblock* (MB), the proposed MCMD algorithm always begins with checking whether the rate-distortion cost computed at the SKIP mode (i.e., Class 1) is below an adaptive threshold for providing a possible early termination chance. If this early termination condition is not met, one of the remaining four motion-activity classes will be chosen for further mode checking according to the analysis of the *predicted motion vector* (PMV) of the current MB. The above-mentioned adaptive threshold and PMV are derived by exploiting the mode correlation between the current MB and a set of adjacent MBs (i.e., *region of support*) in the current view and its neighboring view. Experimental results have shown that compared with exhaustive mode decision, which is a default approach set in the *joint multiview video model* (JMVM) reference software, the proposed MCMD algorithm achieves a reduction of the computational complexity by 73.39% on average, while incurring only 0.07 dB loss in *peak signal-to-noise ratio* (PSNR) and 2.22% increment on the total bit rate.**

***Index Terms*—Early termination, JMVM, mode correlation, mode decision, motion activity, multiview video coding.**

I. INTRODUCTION

WITH the rapid development of camera and display techniques, more interactive and realistic multimedia applications, such as 3DTV and free viewpoint television (FTV), have been emerging [1], [2]. These multiview videos are acquired by using multiple cameras, synchronously capturing the same scene from different viewpoints. In order to store and transmit these huge multiview video data for any practical application, an efficient *multiview video coding* (MVC) technique is indispensable.

A straightforward solution for MVC is to independently encode each video acquired at a specific view by using a state-of-the-art video codec (e.g., H.264/AVC) [3]. However, this approach would yield a low coding efficiency because it does not exploit the *inter*-view correlation existing among the multiview videos. For improving coding efficiency, the Joint Video Team (JVT) standardization body, which is composed of ITU-T VCEG and ISO/IEC MPEG standard committees, has been developing a *joint multiview video model* (JMVM) based on the core of H.264/AVC video coding standard [4], [5]. In the JMVM, not only the sophisticated intra-prediction and the variable block-size *motion estimation* (ME) as adopted in H.264/AVC are used to exploit the spatial and temporal correlation within a single view (i.e., *intra*-view) but also the new technique, the variable block-size *disparity estimation* (DE), is incorporated to employ the *inter*-view correlation between neighboring views for further improving the coding efficiency. However, the higher coding efficiency is achieved at the expense of heavy computational complexity [6]. Therefore, how to reduce the computational complexity while maintaining almost the same video coding quality and the total bit rate through a *fast* algorithm has become the main objective of encoding optimization for any practical multiview video codec realization.

In recent years, fast mode decision methods have been widely studied for *single*-view video coding [7]–[10], but they are not applicable to MVC because the prediction structure of MVC, which involves DE, is different from that of single-view video coding. Therefore, a novel algorithm of fast mode decision for the JMVM is needed and proposed in this paper.

Multiple fast mode decision methods for the JMVM can be found in the literature [11]–[17]. Ding *et al.* [11] presented a content-aware prediction algorithm to efficiently reduce the computational complexity of ME process by using the information of the corresponding *macroblock* (MB) in the neighboring view. Huo *et al.* [12] suggested a scalable prediction structure to skip the DE process adaptively based on the observation that the contribution of DE to the coding efficiency varies from picture to picture. Shen *et al.* [13] presented a fast mode decision method by utilizing the fact that the optimal mode distributions of neighboring views are inherently similar. Therefore, only those more likely modes are chosen and then checked. Lin *et al.* [14] identified those MBs with slow motion content by simply comparing their *motion vector* (MV) lengths with a pre-determined threshold. For those MBs with slow motion, only the more likely prediction direction is selected for

---

Manuscript received July 14, 2010; revised January 5, 2011; accepted February 15, 2011. Date of publication April 15, 2011; date of current version November 2, 2011. This paper was recommended by Associate Editor A. Kaup.

H. Zeng and K.-K. Ma are with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (e-mail: zeng0043@ntu.edu.sg; ekkma@ntu.edu.sg).

C. Cai is with the Institute of Information Science and Technology, Huaqiao University, Quanzhou 362021, China (e-mail: chcai@hqu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2133350

checking in order to speed up the mode decision process. Shen *et al.* [15] presented a fast mode decision method by checking whether the current MB is with motion homogeneity, which is measured according to the MVs of the spatially adjacent MBs in the current view and the corresponding MB in the neighboring view. If the checking result is positive, only the ME process with a block size of 16 × 16 is required to be performed so that the computational complexity can be effectively decreased. Peng *et al.* [16] presented a three-stage early termination method according to three thresholds, which are calculated via the derived formulas that involve some statistical measurements obtained from multiple multi-view video sequences with various kinds of motion contents. Shen *et al.* [17] introduced a model to evaluate the weighting factor of the SKIP mode for the current MB. If this weighting factor is larger than a fixed threshold, the SKIP mode will be chosen as the optimal mode and the mode decision process will be early terminated.

In this paper, a more efficient mode decision algorithm for the JMVM, called the *mode correlation-based mode decision* (MCMD), is proposed. The key points are described as follows. All the prediction modes used in the JMVM are grouped into five motion-activity classes and organized in a hierarchical manner. Then, for the current MB, a comparison of the *rate-distortion* (RD) cost of the SKIP mode (i.e., Class 1) against the adaptive threshold provides an early chance to possibly skip the checking process of the remaining modes. If such early termination is not granted, only one of the remaining four motion-activity classes will be selected to identify the optimal mode according to further analysis of the *predicted motion vector* (PMV) of the current MB. Both the aforementioned adaptive threshold and PMV are derived by utilizing mode correlation. Experimental results have shown that the proposed MCMD algorithm significantly reduces the computational complexity while maintaining almost the same video coding quality and the targeted total bit rate as that of the exhaustive mode decision, which is the default approach implemented in the JMVM reference software [4], [5].

The rest of this paper is organized as follows. An overview of the mode decision conducted in the JMVM is provided in Section II. The proposed fast mode decision algorithm, MCMD, is presented in Section III in detail. Extensive simulation results are documented and discussed in Section IV. Finally, conclusions are drawn in Section V.

II. Overview of Mode Decision in the JMVM

Beside the *intra*-view correlation as contained in single-view video, multiview video contains a substantial amount of *inter*-view correlation between two neighboring views. Hence, unlike the existing single-view video coding standards, the JMVM employs *hierarchical B picture* (HBP) prediction structure [18] to efficiently exploit the intra-view and inter-view correlation. Fig. 1 illustrates the HBP prediction structure with eight views (denoted by V_i, for $i = 0, 1, \ldots, 7$), and the length of group of picture (GOP) is 8 (denoted by T_i, for $i = 0, 1, \ldots, 7$). In this HBP prediction structure, all the views can be categorized into two types: the *main* views (namely, V_0, V_2, V_4, and V_6) and the *auxiliary* views (namely, V_1, V_3, V_5, and V_7) [19]. In the main views, the ME is performed within the same view to exploit the temporal correlation. In the auxiliary views, apart from the ME, the DE is conducted by referring to the pictures at the same time instant but from two neighboring views to employ the inter-view correlation. Moreover, all the pictures can be classified into two classes: the *anchor* pictures (i.e., the pictures at T_0 and T_8) and the *non-anchor* pictures (i.e., the pictures at T_i, for $i = 1, 2, 3, 4, 5, 6, 7$), which are presented between two adjacent anchor pictures.

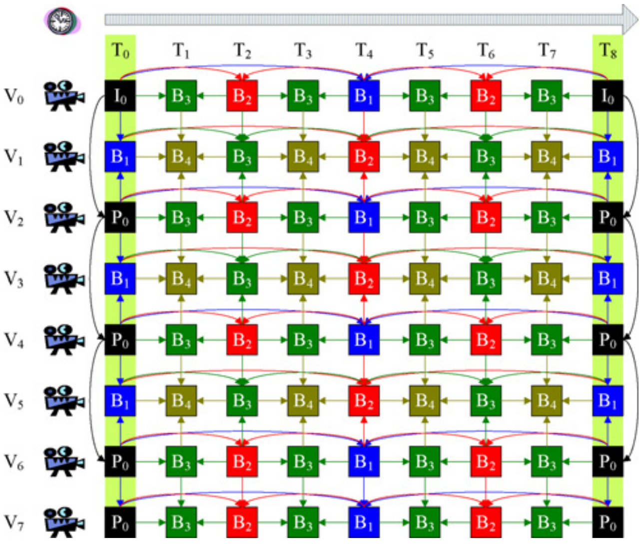


Fig. 1. HBP prediction structure used in the JMVM.

In order to achieve higher coding efficiency, the JMVM provides seven block sizes (or the so-called *modes*) as shown in Fig. 2 to conduct both ME and DE. The above-mentioned seven block sizes are 16 × 16, 16 × 8, 8 × 16, 8 × 8, 8 × 4, 4 × 8, and 4 × 4, among which the last four block sizes are jointly denoted as P8 × 8 in the JMVM, as illustrated in Fig. 2(b). For the *inter*-frame MB coding, there are 11 candidate modes: SKIP, inter_16 × 16, inter_16 × 8, inter_8 × 16, inter_8 × 8, inter_8 × 4, inter_4 × 8, inter_4 × 4, intra_4 × 4, intra_8 × 8, and intra_16 × 16. For the *intra*-frame MB coding, only intra_4 × 4, intra_8 × 8, and intra_16 × 16 are applicable. Furthermore, the Lagrangian *rate-distortion optimization* (RDO) function [20], [21] is adopted as its mode decision criterion as follows:

$$J_{RD} = SSD(s, c|QP) + \lambda_{MODE} \cdot R(s, c|QP) \qquad (1)$$

where λ_{MODE} is the Lagrangian multiplier, QP is the quantization parameter, SSD means the sum of squared differences incurred between the original luma block (denoted by s) and its prediction reconstructed block (denoted by c), and R represents the total number of bits required for representing the headers, MVs, *disparity vectors* (DVs), and coefficients.

It should be pointed out that the aforementioned modes require to be individually checked based on this mode decision criterion, and the results are compared to identify the optimal mode that corresponds to the minimum RD cost; this approach is the so-called *exhaustive mode decision*. Consequently, the computational complexity resulted from such exhaustive mode

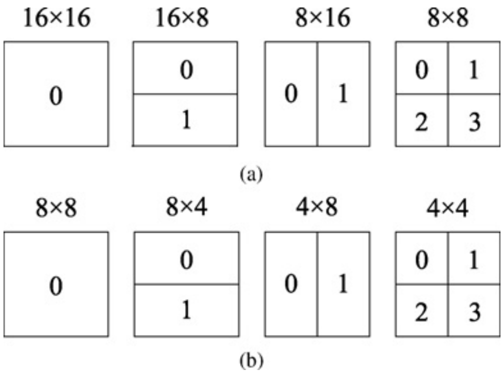


Fig. 2. Seven block sizes used in the JMVM. (a) 16 × 16 MB and its sub-blocks. (b) 8 × 8 block and its sub-blocks.

decision is extremely high, and thus a *fast* mode decision algorithm is highly desirable, especially for real-time applications.

III. PROPOSED MODE CORRELATION-BASED MODE DECISION (MCMD) ALGORITHM

A. Motivation

First, it has been well-recognized that the SKIP mode is the most frequently encountered optimal mode in H.264/AVC for single-view video sequences [7]–[9]. It is expected that this might be also true for multiview video sequences. To verify this intuition, extensive simulation experiments have been conducted based on a set of multiview video sequences as listed in Table I [22]. The test conditions are as follows: each test sequence is encoded using the HBP prediction structure under a GOP length = 16, QP = 20, 24, 28, 32, and 36, respectively, RDO and context-adaptive binary arithmetic coding (CABAC) entropy coding are enabled, and the search range of ME and DE is ±64. By exploiting the exhaustive mode decision in the JMVM under the above-mentioned test conditions, we study the distribution of optimal mode in the non-anchor pictures of the auxiliary views, and the results are documented in Table II. This study clearly indicates that the SKIP mode also dominates among all the available modes for multiview video sequences, especially for those sequences containing motionless or slow-motion content. This observation implies that the SKIP mode checking should be done at the beginning of the mode decision process for each MB to provide a chance for early termination.

Second, ME is used to conduct temporal prediction, while DE is employed to perform inter-view prediction. It can be further observed from Table II that the average percentage of ME turning out to be the optimal choice is 25.11%, while that of DE is 3.07%. It means that temporal prediction via ME is more likely to be a better choice, as compared to inter-view prediction via DE. This is because most real-world video sequences are full of stationary background with slow-motion objects. In such cases, the correlation between two consecutive pictures from the same view will be much stronger than that of the two pictures acquired at the same time instant but from two different views. On the contrary, for those highly textured regions with fast motion or scene changes, the temporal correlation from the intra-view becomes much weaker than the inter-view correlation; hence, exploiting the inter-view prediction is more effective in this case. In addition, various block sizes are exploited to more accurately capture the true motion of the real-world video sequences. One can easily perceive that a large block size is more suitable for a homogeneous region under slow motion, while a small block size is more proper to use for a region containing a fast moving object.

TABLE I

MULTIVIEW VIDEO SEQUENCES

Sequences		Resolution	Frames
KDDI	*Flamenco1*	320 × 240	250
KDDI	*Race1*	640 × 480	250
MERL	*Ballroom*	640 × 480	250
MERL	*Exit*	640 × 480	250
MERL	*Vassar*	640 × 480	250
Tanimoto Lab	*Akko&Kayo*	640 × 480	250
Tanimoto Lab	*Rena*	640 × 480	250
Microsoft	*Breakdancers*	1024 × 768	100
HHI	*Jungle*	1024 × 768	250
HHI	*Uli*	1024 × 768	250

TABLE II

DISTRIBUTION OF THE OPTIMAL MODE RESULTED FROM EXPLOITING THE EXHAUSTIVE MODE DECISION IN THE JMVM USING THE MULTIVIEW VIDEO SEQUENCES LISTED IN TABLE I UNDER DIFFERENT QP VALUES

QP	SKIP	ME (%)				DE	Intra
	(%)	16 × 16	16 × 8	8 × 16	P8 × 8	(%)	(%)
20	53.77	13.93	6.08	6.59	12.95	5.52	1.16
24	66.27	12.41	4.66	4.90	7.51	3.57	0.68
28	73.71	10.21	3.35	3.28	6.32	2.67	0.46
32	79.06	8.98	2.87	2.76	3.91	2.11	0.31
36	83.50	6.75	2.03	2.19	3.84	1.49	0.20
Average	71.26	10.46	3.80	3.94	6.91	3.07	0.56

Third, there exists a mass amount of spatial, temporal, and inter-view correlations in multiview video as follows. From the spatial point of view, large homogeneous areas are presented in the spatial domain quite frequently. From the temporal point of view, motionless or slow-motion scenes are oftentimes encountered in the real-world video sequences. From the inter-view point of view, a pair of pictures at the same time instant from two neighboring views tends to represent the same content. This leads to the fact that the resultant coding information (namely, the optimal mode, the MVs, and the corresponding RD cost) of the current MB are intimately related to that of its adjacent MBs. Thus, we shall utilize these information as the *mode correlation* on the design of fast mode decision algorithm, as follows.

B. Proposed MCMD Algorithm

Based on the above analysis, all the prediction modes available in the JMVM are classified into five motion-activity classes as summarized in Table III, and then by using the mode correlation, only one of them will be chosen for further mode checking in a hierarchical manner. For the current MB, the proposed MCMD algorithm always starts with checking the SKIP mode in Class 1, which may arrive at the early termination for the entire mode decision process. That is, if the RD

TABLE III

MOTION-ACTIVITY CLASSES AND THEIR INVOLVED MODES TO BE CHECKED

Class	Motion Activity	Involved Modes
1	Motionless	SKIP
2	Slow motion (homogeneous region)	SKIP ME16 × 16
3	Moderate motion	ME16 × 8, ME8 × 16
4	Fast motion	ME_P8 × 8 (i.e., 8 × 8, 8 × 4, 4 × 8, 4 × 4)
5	Highly-textured region in fast motion or with scene changes	DE (16 × 16, 16 × 8, 8 × 16, P8 × 8) intra_4 × 4, intra_8 × 8, intra_16 × 16

TABLE IV

EMPIRICALLY-DETERMINED WEIGHT W_i OF THE CORRESPONDING MB_i BASED ON THE ROS AS ILLUSTRATED IN FIG. 3

MB Index i	1, 2, 4, 5	3, 7, 9, 10, 12	6, 8, 11, 13
W_i	1.30	0.96	0.75

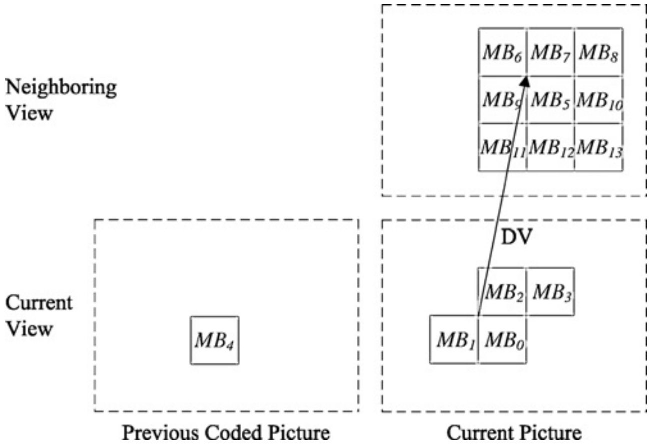


Fig. 3. ROS of the current MB, MB_0.

cost computed at the SKIP mode is small enough and below an adaptive threshold T_{SKIP}, the SKIP mode will be selected as the optimal mode, and the mode decision process proceeds to the next MB. Otherwise, only one of the remaining four motion-activity classes will be selected based on the city-block distance of the PMV of the current MB. The above-mentioned adaptive threshold T_{SKIP} and the PMV are derived by exploiting the mode correlation between the current MB and a set of nearby MBs (defined as *region of support*, ROS). Fig. 3 shows the ROS, which consists of the spatially and temporally nearby MBs in the current view and the corresponding MB and its eight adjacent MBs in the neighboring view. Note that MB_4 is located at the same position as the current MB (i.e., MB_0) in the previous coded picture, and the corresponding MB in the neighboring view (i.e., MB_5) is located based on the global DV yielded between the current view and the neighboring view as introduced in [23]. Here, we consistently only consider the forward view of the current view as its neighboring view in the ROS due to the following two justifications.

1) Our experimental results have shown that the performance of using both forward and backward views as neighboring views is almost the same as that of using either forward view or backward view as its neighboring view.
2) The last auxiliary view (i.e., V_7 in Fig. 1) has only one neighboring view (i.e., V_6 in Fig. 1).

In this paper, the adaptive threshold T_{SKIP} is proposed as follows:

$$T_{SKIP} = \frac{\sum_{i=1}^{13} W_i \cdot K_i \cdot RDcost(SKIP)_i}{\sum_{i=1}^{13} W_i \cdot K_i} \tag{2}$$

where $RDcost(SKIP)_i$ and W_i are the RD costs of MB_i computed at the SKIP mode and the weights of the corresponding MB_i in the ROS, for $i = 1, 2, \ldots, 13$, respectively. Based on the intuition that the closer the neighboring MB to the current MB, the larger the weight should be assigned, the weights W_i are empirically determined from our extensive simulation experiments and the results are documented in Table IV. Since adjacent MB choosing non-SKIP mode as its optimal mode has little contribution to decide whether the SKIP mode is the optimal mode of the current MB, only the RD cost of those adjacent MBs in the ROS, which selected the SKIP mode as the optimal mode, will be used. Hence, K_i is defined as

$$K_i = \begin{cases} 1, & \text{if the SKIP mode was the optimal mode of } MB_i \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Note that if all the K_i are equal to 0, the checking of early termination will be skipped for this special case.

If the aforementioned early termination is not granted, the optimal prediction mode will be identified from one of the remaining four classes (namely, Classes 2 to 5), by evaluating the city-block distance of the PMV of the current MB as follows.

1) For the current MB (i.e., MB_0), establish the MV set $\{mv_1, mv_2, mv_3, \ldots, mv_{13}\}$ according to the defined ROS as shown in Fig. 3, where $mv_i = (x_i, y_i)$ is the MV of the corresponding MB_i, for $i = 1, 2, \ldots, 13$, respectively. If the MB involved in the ROS has multiple MVs for different block sizes, a single MV of this MB can be obtained through a bottom-up merging procedure as reported in [24].
2) Derive the PMV of MB_0, $PMV = (x_p, y_p)$, from the MV set as follows:

$$\begin{cases} x_p = \dfrac{\sum_{i=1}^{13} W_i \cdot Z_i \cdot x_i}{\sum_{i=1}^{13} W_i} \\[2ex] y_p = \dfrac{\sum_{i=1}^{13} W_i \cdot Z_i \cdot y_i}{\sum_{i=1}^{13} W_i} \end{cases} \tag{4}$$

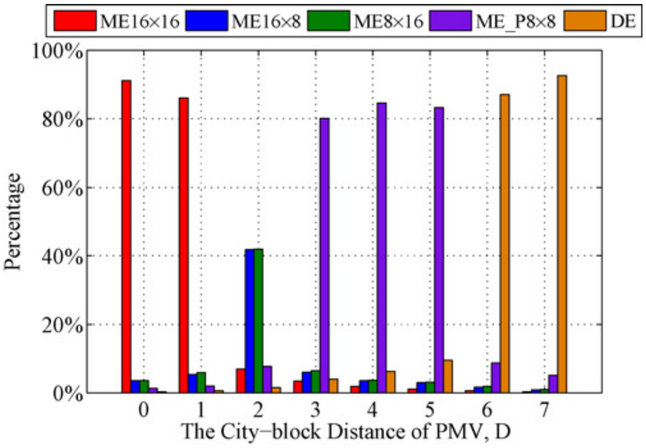


Fig. 4. Percentage of the optimal block sizes resulted from exploiting the exhaustive mode decision in the JMVM under different values of city-block distance of PMV D at QP=28.

where W_i as shown in Table IV are the weights of the corresponding MB_i, for i = 1, 2, ..., 13, respectively; Z_i is the normalization factor of the corresponding mv_i in the established MV set, which is to further normalize all the mv_i to the forward prediction direction. Here, Z_i is defined as follows:

$$Z_i = \begin{cases} +1, & \text{if } mv_i \text{ was the forward MV} \\ -1, & \text{if } mv_i \text{ was the backward MV.} \end{cases} \tag{5}$$

3) Calculate the city-block distance of the PMV of MB_0, D as follows:

$$D = |x_p| + |y_p|. \tag{6}$$

4) The motion activity of MB_0 is determined according to

$$\text{Motion Activity} = \begin{cases} \text{Slow}, & \text{if } D \le D_1 \\ \text{Moderate}, & \text{if } D_1 < D \le D_2 \\ \text{Fast}, & \text{if } D_2 < D \le D_3 \\ \text{Highly textured}, & \text{if } D_3 < D \end{cases} \tag{7}$$

where "Slow," "Moderate," "Fast," and "Highly textured" correspond to Classes 2, 3, 4, and 5 in Table III, respectively.

To determine the thresholds D_1, D_2, and D_3, extensive simulations have been conducted on a set of multiview video sequences as listed in Table I. The test conditions are as follows: each test sequence is encoded using the HBP prediction structure under a GOP length = 16, various QP values are used, RDO and CABAC entropy coding are enabled, and the search range of ME and DE is ±64. By exploiting the exhaustive mode decision in the JMVM under the aforementioned test conditions, we investigate the relationship between the optimal block sizes and the city-block distance of PMV D in the non-anchor pictures of the auxiliary views. Here, let us take the case QP = 28 as an example and the corresponding results are presented in Fig. 4. For example, one can see that, when D is equal to 0 or 1, the ME16 × 16 (i.e., Class 2) has the highest percentage (i.e., almost 90%) to be the optimal block size. Thus, the threshold D_1 is set to 1. Likewise, the thresholds D_2 and D_3 are determined as 2 and 5, respectively. It should be pointed out that similar arguments are also applied for other QP values.

5) Only the modes that belong to the identified motion-activity class are individually checked, and the one with the minimum cost yielded is selected as the optimal mode.

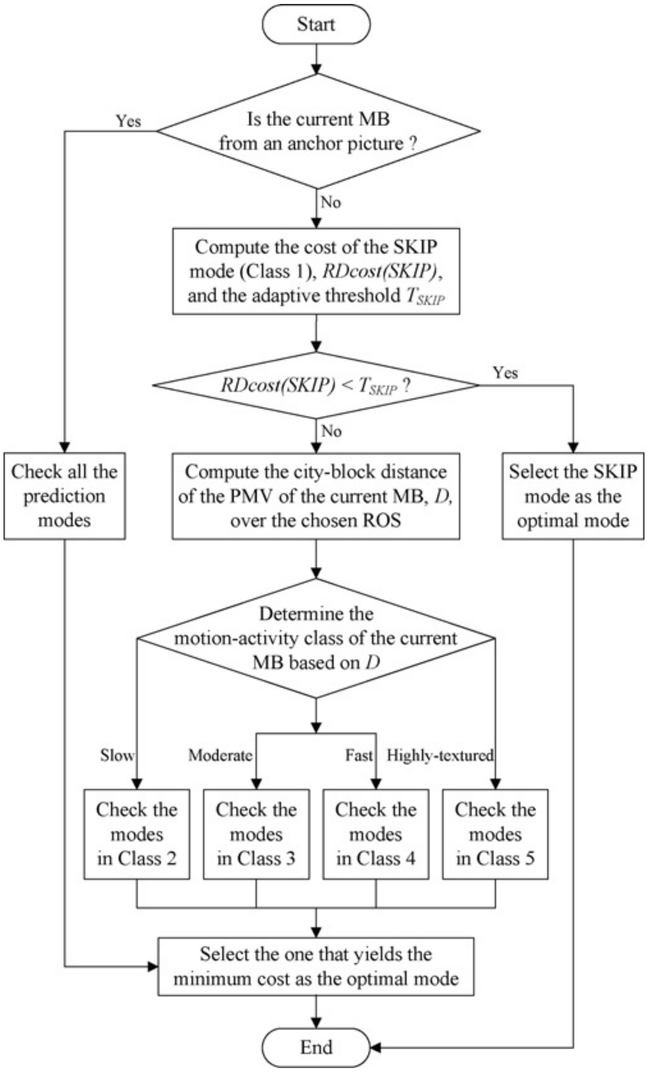


Fig. 5. Flowchart of the proposed fast mode decision method, MCMD.

The entire algorithm is summarized in a flowchart as shown in Fig. 5. It is worth noting that the ROS of some special MBs are not completely available. For example, the boundary MBs located in the first row, the first column, and the last column. In such cases, the exhaustive mode decision will be performed for these special MBs instead of the proposed MCMD algorithm for ensuring the coding efficiency.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. *Test Conditions*

To evaluate the performance, the proposed MCMD algorithm has been incorporated into the JMVM 8.0 [4], [5]

TABLE V

EXPERIMENTAL RESULTS OF THE PROPOSED MCMD ALGORITHM COMPARED WITH THE EXHAUSTIVE MODE DECISION APPROACH ON MULTIPLE MULTIVIEW VIDEO SEQUENCES

Sequences	BDPSNR (dB)	BDBR (%)	$\triangle PSNR$ (dB)	$\triangle B$ (%)	$\triangle T$ (%)
Akko&Kayo	−0.09	+2.18	−0.08	+0.09	−75.39
Alt_Moabit	−0.02	+0.77	−0.05	−0.83	−72.85
Ballroom	−0.08	+2.27	−0.06	+0.86	−74.05
Breakdancers	−0.07	+3.72	−0.09	−0.82	−66.85
Champagne_tower	−0.07	+2.06	−0.09	−0.71	−74.75
Dog	−0.03	+0.92	−0.05	−0.48	−72.85
Door_Flowers	−0.05	+2.76	−0.08	−1.04	−77.73
Exit	−0.07	+3.24	−0.09	−0.67	−83.04
Flamenco1	−0.11	+2.51	−0.10	+0.50	−75.94
Flamenco2	−0.10	+2.44	−0.08	+0.34	−73.09
Janine2	−0.12	+2.96	−0.10	+0.53	−71.81
Jungle	−0.11	+2.64	−0.09	+0.51	−71.60
Race1	−0.06	+1.89	−0.06	−0.02	−76.36
Rena	−0.05	+1.14	−0.10	−1.23	−68.44
Uli	−0.12	+2.86	−0.09	+0.82	−70.23
Vassar	−0.02	+1.24	−0.04	−0.74	−69.31
Average	−0.07	+2.22	−0.08	−0.18	−73.39

and experimented on multiple multiview video sequences, covering a wide variety of motion activities [22], [25], [26]. Three different views from each test sequence are chosen for simulation. The first and the third views are used as the reference views (i.e., V_0 and V_2 in Fig. 1, respectively); while the second view (i.e., V_1 in Fig. 1) is utilized for the implementation of the proposed MCMD algorithm. The personal computer used for conducting these experiments is made up of 2.66 GHz Intel Core2 processor and 4 GB memory. The test conditions are set as follows.

1) Each test sequence is encoded using the HBP prediction structure under a GOP length = 16.
2) The QP is set at 20, 24, 28, and 32, respectively.
3) The RDO is enabled.
4) The CABAC entropy coding is used.
5) The search range of ME and DE is ±64.

B. Performance Evaluations and Comparisons

Table V shows the experimental results based on a set of multiview video sequences by comparing the proposed fast mode decision algorithm, MCMD, with the exhaustive mode decision of JMVM. Table VII compares the performance resulted from the proposed MCMD algorithm, and the method recently proposed by Shen *et al.* [15]. The performance evaluation is made with respect to the results of the auxiliary view obtained by the exhaustive mode decision, and their differences are measured by the following indexes.

1) The time saving $\triangle T$ is computed according to

$$\triangle T = \frac{T_{Proposed} - T_{JMVM}}{T_{JMVM}} \times 100\% \tag{8}$$

where $T_{Proposed}$ and T_{JMVM} denote the total encoding time of the proposed method and the JMVM reference software, respectively.

TABLE VI

HIT RATIO OF THE SKIP MODE RESULTED FROM THE PROPOSED EARLY TERMINATION SCHEME ON MULTIPLE MULTIVIEW VIDEO SEQUENCES UNDER DIFFERENT QP VALUES

QP	20	24	28	32	Average
Hit ratio (%)	82.64	86.87	89.42	90.48	87.35

TABLE VII

EXPERIMENTAL RESULTS OF TWO METHODS: (A) SHEN *et al.* [15] AND (B) OUR PROPOSED MCMD ALGORITHM. ALL THE INCREMENTAL DIFFERENCES ARE AVERAGED OVER FOUR QPS AND RESULTED FROM COMPARING EACH METHOD WITH RESPECT TO THE EXHAUSTIVE MODE DECISION APPROACH, INDIVIDUALLY

Sequences	Method	$\triangle PSNR$ (dB)	$\triangle B$ (%)	$\triangle T$ (%)
Ballroom	(A)	−0.04	+2.02	−72.31
	(B)	−0.06	+0.86	−74.05
Exit	(A)	−0.02	+0.43	−80.21
	(B)	−0.09	−0.67	−83.04
Flamenco1	(A)	−0.09	+0.84	−66.70
	(B)	−0.10	+0.50	−75.94
Jungle	(A)	−0.02	+0.17	−42.38
	(B)	−0.09	+0.51	−71.60
Race1	(A)	+0.00	+0.40	−38.49
	(B)	−0.06	−0.02	−76.36
Uli	(A)	−0.04	−0.03	−51.12
	(B)	−0.09	+0.82	−70.23
Average	(A)	−0.04	+0.64	−58.54
	(B)	−0.08	+0.33	−75.20

2) To show the RD performance, two kinds of measurements are adopted in this paper.
    2.1) The Bjøntegaard delta PSNR (BDPSNR) (in dB) and Bjøntegaard delta bit rate (BDBR) (in percentage) as suggested in [27] are used to measure the averaged PSNR and the bit rate difference between the RD performance produced by the proposed method and the JMVM reference software, respectively.
    2.2) The $\triangle PSNR$ (in dB) and $\triangle B$ (in percentage) are utilized to compute the PSNR changes and the total bit rate changes averaged over four QPs, that is

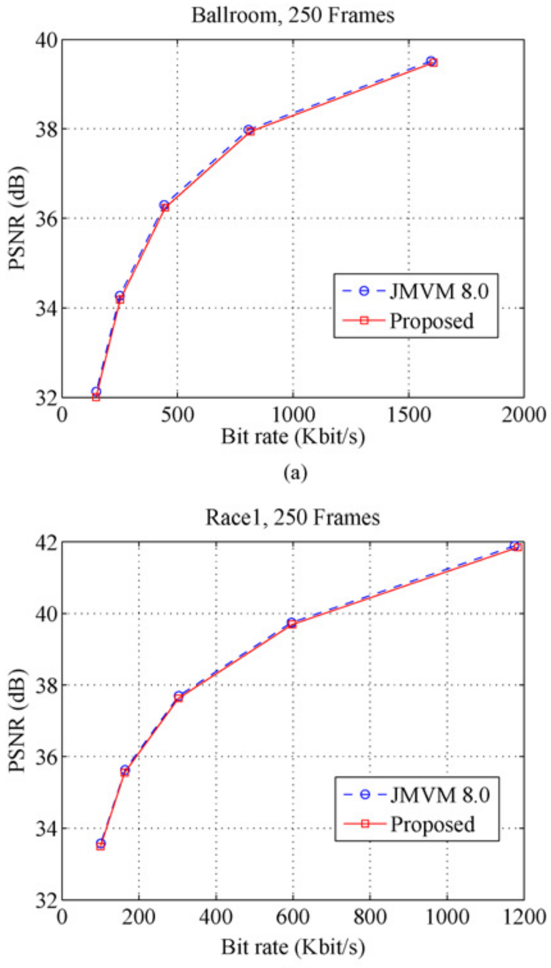
$$\triangle PSNR = PSNR_{Proposed} - PSNR_{JMVM} \tag{9}$$

where $PSNR_{Proposed}$ and $PSNR_{JMVM}$ denote the PSNR values resulted from the proposed method and the JMVM reference software, respectively, as follows:

$$\triangle B = \frac{B_{Proposed} - B_{JMVM}}{B_{JMVM}} \times 100\% \tag{10}$$

where $B_{Proposed}$ and B_{JMVM} denote the bit rate resulted from the proposed method and the JMVM reference software, respectively.

From the results as shown in Table V, it can be seen that the proposed MCMD algorithm constantly achieves, on average, 73.39% computational time saving with only 0.07 dB loss in PSNR and 2.22% increment in the total bit rate in terms of BDPSNR and BDBR, comparing with the outcomes



(a)

(b)

Fig. 6. RD curves of two multiview video sequences. (a) *Ballroom*. (b) *Race1*.

resulted by applying the exhaustive mode decision of JMVM. Fig. 6(a) and (b) presents the RD performance of two multiview video sequences *Ballroom* and *Race1* as examples, showing that the proposed fast mode decision algorithm has a similar RD performance as that of the exhaustive mode decision of the JMVM.

To further show the accuracy of the proposed early termination scheme by comparing the RD cost of the SKIP mode with an adaptive threshold T_{SKIP}, extensive simulations have been conducted to investigate the *hit ratio* of the SKIP mode resulted from the proposed early termination scheme on multiple multiview video sequences, and the results are documented in Table VI. It can be observed that if the SKIP mode is assigned to the current MB as its optimal mode by the proposed early termination scheme, the hit ratio that the SKIP mode is indeed the optimal mode determined by the exhaustive mode decision is 87.35% on average. All these verify that the proposed early termination scheme performs well in general.

In addition to achieve superior performance to that of the exhaustive mode decision, experimental results as demonstrated in Table VII further indicate that the proposed MCMD algorithm consistently outperforms the recent approach proposed by Shen *et al.* [15]—with about additional 16.66% encoding time saving, 0.04 dB PSNR loss, and 0.31% total bit rate reduction.

V. CONCLUSION

In this paper, a novel fast mode decision algorithm for the JMVM, called the mode correlation-based mode decision (MCMD), was proposed. The speed-up gain was achieved by properly grouping all the prediction modes into classes, followed by arranging them in a hierarchical manner such that only one of the checking paths will be chosen for identifying the optimal mode. For each MB, if the RD cost computed at the SKIP mode is below an adaptive threshold, the SKIP mode will be chosen as the optimal mode, and the mode decision process is terminated. Otherwise, further analysis of the PMV of the current MB is employed to skip those unlikely modes. Note that both the adaptive threshold and PMV are derived by using the mode correlation. Experimental results have verified that the proposed fast mode decision algorithm, MCMD, effectively reduces the computational load by 73.39% on average as compared with that of the exhaustive mode decision based on the experimental platform JMVM 8.0, while only incurring a negligible loss of PSNR (about 0.07 dB on average) and slightly increase of the total bit rate (about 2.22% on average).

REFERENCES

[1] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. Kopylow, "A survey of 3DTV displays: Techniques and technologies," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1647–1658, Nov. 2007.

[2] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3D video and free viewpoint video: Technologies, applications and MPEG standards," in *Proc. IEEE ICME*, Jul. 2006, pp. 2161–2164.

[3] M. Fierl and B. Girod, "Multiview video compression," *IEEE Signal Process. Mag.*, vol. 24, no. 6, pp. 66–76, Nov. 2007.

[4] T. Wiegand, G. J. Sullivan, G. Bjonte, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[5] *Joint Multiview Video Model (JMVM) 8.0*, document JVT-AA208, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Geneva, Switzerland, Apr. 2008.

[6] A. Smolic, K. Mueller, N. Stefanoski, J. Ostermann, A. Gotchev, G. B. Akar, G. Triantafyllidis, and A. Koz, "Coding algorithms for 3DTV—A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1606–1621, Nov. 2007.

[7] I. Choi, J. Lee, and B. Jeon, "Fast coding mode selection with rate-distortion optimization for MPEG-4 part-10 AVC/H.264," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 12, pp. 1557–1561, Dec. 2006.

[8] C. S. Kannangara, I. E. G. Richardson, M. Bystrom, J. Solera, Y. Zhao, A. Maclennan, and R. Cooney, "Low complexity skip prediction for H.264 through Lagrangian cost estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 2, pp. 202–208, Feb. 2006.

[9] H. Q. Zeng, C. H. Cai, and K.-K. Ma, "Fast mode decision for H.264/AVC based on macroblock motion activity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 491–499, Apr. 2009.

[10] H. Q. Zeng, K.-K. Ma, and C. H. Cai, "Hierarchical intra mode decision for H.264/AVC,"*IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 907–912, Jun. 2010.

[11] L.-F. Ding, P.-K. Tsung, S.-Y. Chien, W.-Y. Chen, and L.-G. Chen, "Content-aware prediction algorithm with inter-view mode decision for multiview video coding," *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 1553–1564, Dec. 2008.

[12] J. Y. Huo, Y. L. Chang, M. Li, and Y. Z. Ma, "Scalable prediction structure for multiview video coding," in *Proc. IEEE ISCAS*, May 2009, pp. 2593–2596.

[13] L. Q. Shen, T. Yan, Z. Liu, Z. Y. Zhang, P. An, and L. Yang, "Fast mode decision for multiview video coding," in *Proc. IEEE ICIP*, Nov. 2009, pp. 2953–2956.

[14] J.-P. Lin and A. C.-W. Tang, "A fast direction predictor of inter frame prediction for multi-view video coding," in *Proc. IEEE ISCAS*, May 2009, pp. 2589–2592.

[15] L. Q. Shen, Z. Liu, S. X. Liu, Z. Y. Zhang, and P. An, "Selective dispaprity estimation and variable size motion estimation based on motion homogeneity for multi-view coding," *IEEE Trans. Broadcast.*, vol. 55, no. 4, pp. 761–766, Dec. 2009.

[16] Z. J. Peng, G. Y. Jiang, and M. Yu, "A fast multiview video coding algorithm based on dynamic multi-threshold," in *Proc. IEEE ICME*, Jun. 2009, pp. 113–116.

[17] L. Q. Shen, Z. Liu, T. Yan, Z. Y. Zhang, and P. An, "Early SKIP mode decision for MVC using inter-view correlation," *Signal Process.: Image Commun.*, vol. 25, no. 2, pp. 88–93, Feb. 2010.

[18] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction sturcture for multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[19] X. M. Li, D. B. Zhao, S. W. Ma, and W. Gao, "Fast disparity and motion estimation based on correlations for multiview video coding," *IEEE Trans. Consumer Electron.*, vol. 54, no. 4, pp. 2037–2044, Nov. 2008.

[20] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Process. Mag.*, vol. 15, no. 6, pp. 74–90, Nov. 1998.

[21] T. Wiegand, H. Schwarz, A. Joch, F. Kossentini, and G. J. Sullivan, "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.

[22] Y. P. Su, A. Vetro, and A. Smolic, *Common Test Conditions for Multiview Video Coding*, document JVT-U211, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Hangzhou, China, Oct. 2006.

[23] H.-S. Koo, Y.-J. Jeon, and B.-M. Jeon, *MVC Motion Skip Mode*, document JVT-W081, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, San Jose, CA, Apr. 2007.

[24] K.-C. Hou, M.-J. Chen, and C.-T. Hsu, "Fast motion estimation by motion vector merging procedure for H.264," in *Proc. IEEE ICME*, Jul. 2005, pp. 1444–1447.

[25] I. Feldmann, M. Muller, F. Zilly, R. Tanger, K. Muller, A. Smolic, P. Kauff, and T. Wiegand, *HHI Test Material for 3D Video*, document MPEG08/M15413, ISO/IEC JTC1/SC29/WG11, Archamps, France, May 2008.

[26] *MPEG-FTV Project* [Online]. Tanimoto Lab., Dept. Inform. Electron., School Eng., Nagoya Univ., Nagoya, Japan. Available: http://www.tanimoto.nuee.nagoya-u.ac.jp

[27] G. Bjøntegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33, VCEG 13th Meeting, Austin, TX, Apr. 2001.

Huanqiang Zeng received the B.S. and M.S. degrees from Huaqiao University, Quanzhou, China, in 2005 and 2008, respectively, both in electronic engineering. He is currently pursuing the Ph.D. degree at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

From May 2008 to July 2008, he was a Research Associate with the Temasek Laboratories, Nanyang Technological University. His current research interests include areas of image processing, video communication, multiview video coding, and computer vision.

Kai-Kuang Ma (S'80–M'84–SM'95) received the B.E. degree in electronic engineering from Chung Yuan Christian University, Chung-Li, Taiwan, the M.S. degree in electrical engineering from Duke University, Durham, NC, and the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, NC.

He is currently a Full Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From 1992 to 1995, he was a Technical Staff Member with the Institute of Microelectronics, Singapore, working on digital video coding and the Moving Picture Experts Group (MPEG) standards. From 1984 to 1992, he was with IBM Corporation, Kingston, NY, and Research Triangle Park, Durham, engaged in various digital signal processing and very large scale integrated advanced product development. His current research interests include areas of digital image processing and computer vision. These include digital image/video coding and standards, image/video segmentation, denoising and enhancement, interpolation and super-resolution, and error resilience and concealment. His recent research on computer vision specifically focused on camera array image processing, wide-baseline matching, and registration for scene analysis and recognition; these are resulted from multiple defense application-related research grants, including DARPA. He has produced extensive publications in various international journals, conferences, and MPEG standardization meetings. He holds one U.S. patent on fast motion estimation algorithm.

Prof. Ma was a Distinguished Lecturer with the IEEE Circuits and Systems Society from 2008 to 2009. He has been actively contributing various technical services for numerous international conferences, especially, serving as Technical Program Co-Chair of the IEEE International Conference on Image Processing in 2004, the International Symposium on Intelligent Signal Processing and Communication Systems in 2007, and the Pacific-Rim Symposium on Image and Video Technology in 2010. He is currently an Associate Editor or Editorial Board Member of six international journals: the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2007 to present, the IEEE TRANSACTIONS ON COMMUNICATIONS from 1997 to present, the IEEE TRANSACTIONS ON MULTIMEDIA from 2002 to present, the *International Journal of Image and Graphics* from 2002 to present, the *Journal of Visual Communication and Image Representation* from 2005 to present, and the *Research Letters in Signal Processing* from 2008 to present. He is an elected member of three technical committees: the Image and Multidimensional Signal Processing Committee of the IEEE Signal Processing Society, the Multimedia Communications Committee of the IEEE Communications Society, and the Digital Signal Processing of the IEEE Circuits and Systems Society. He was the MPEG Chairman and the Head of Delegation, Singapore, from 1997 to 2001. He was the Chairman of IEEE Signal Processing Singapore Chapter from 2000 to 2002. He is a member of Eta Kappa Nu.

Canhui Cai (M'06–SM'08) received the B.S. degree from Xidian University, Xian, China, in 1982, the M.S. degree from Shanghai University, Shanghai, China, in 1985, and the Ph.D. degree from Tianjin University, Tianjin, China, in 2003, all in electronic engineering.

Since 1984, he has been a Faculty Member with Huaqiao University, Quanzhou, China, and is currently a Professor of electronic and communication engineering. He was a Visiting Professor with Delft University of Technology, Delft, The Netherlands, from 1992 to 1993, and a Visiting Professor with the University of California at Santa Barbara, Santa Barbara, from 1999 to 2000. His current research interests include video communications, image and video signal processing, and computer vision. He is the author or co-author of four books, and has published over 80 papers in journals and conference proceedings.

Dr. Cai was a General Co-Chair of Intelligent Signal Processing and Communication Systems in 2007.