Early DIRECT Mode Decision for MVC Using MB Mode Homogeneity and RD Cost Correlation

Yue Li, Gaobo Yang, Ning Chen, Yapei Zhu, and Xiangling Ding

***Abstract*—Multi-view video coding (MVC) adopts variable size mode decision to achieve high coding efficiency. However, its high computational complexity is a bottleneck of enabling MVC into practical real-time applications. In this paper, an early termination strategy is proposed for DIRECT mode decision of MVC by exploiting mode homogeneity and rate distortion (RD) cost correlation. By comparing the RD cost between DIRECT mode and Inter16×16 mode, an adaptive threshold is defined based on the MB's mode homogeneity and RD cost so as to early terminate the remaining inter and intra modes. Experimental results show that compared with the original JMVC model, the proposed approach can reduce the total encoding time from 65.08% to 91.45% (80.43% on average). Meanwhile, the Bjontegaard delta peak signal-to-noise ratio only decreases 0.031 dB and Bjontegaard delta bit rate increases 0.97% on average, which is a negligible loss of coding efficiency and superior to the performance of state-of-the-art methods.**

***Index Terms*—DIRECT mode decision, mode homogeneity, RD cost, multi-view video coding.**

I. INTRODUCTION

THE MULTI-VIEW video (MVV) and three-dimensional (3D) video have been emerging in recent years, which provide users with real depth perception, interactivity and novel visual enjoyment [1]. Multi-view video plus depth (MVD) is an advanced 3D video representation format, and it consists of multi-view texture video and corresponding depth video [2]. Since multi-view video is simultaneously captured from multiple cameras with different viewpoints or angles, it has a huge amount of data due to high spatial-view-temporal redundancy, and should be efficiently encoded for storage and transmission. To this end, Joint Video Team (JVT) has developed multi-view video coding (MVC) standard [3], in which various prediction techniques such as motion estimation (ME) and disparity estimation (DE) are adopted to remove the temporal and inter-view redundancies. Moreover, hierarchical B picture (HBP) prediction structure, multi-reference frame and exhaustive searching process of mode decision are jointly used to further improve coding efficiency [4]. However, these advanced techniques lead to very high computational complexity, which has been a bottleneck of applying MVC into real time applications such as 3D live broadcasting and interactive free viewpoint television (FTV) [5].

To address the high computational complexity issue, lots of fast ME and DE algorithms have been proposed. Deng *et al.* [6] proposed a fast motion and disparity estimation method by using a pair of motion and disparity vectors. Shen *et al.* [7], [8] proposed the concept of motion homogeneity, which is measured by motion vectors (MV) of spatial neighboring MBs and inter-view collocated MBs, to speed up ME and DE. Zhu *et al.* [9], [10] proposed hexagon-based search pattern for fast block ME. Later, Pan *et al.* [11] further presented a multiple hexagon search algorithm for efficient ME and DE. In addition, Pan *et al.* [12] proposed hybrid optimization strategies to early terminate the process of ME/DE, including initial search point, optimal theory and block matching strategy. These methods can efficiently reduce the search range of ME and DE. However, exhaustive mode decision is not early terminated to further reduce computational complexity of MVC.

In the literature, there are a few low-complexity mode decision methods, which exploit RD cost, mode correlation and/or prediction model to speed up the process of mode decision. Among them, references [13], [14], [18], and [20] exploit adaptive RD cost thresholds, and references [15] and [16] exploit mode correlation strategy for fast mode decision. In addition, Zhang *et al.* [17] proposed a statistical early termination model for efficient mode decision and reference frame selection. Zhao *et al.* [19] proposed a hybrid optimal stopping model for fast mode decision, which achieves a good compromise between RD degradation and complexity reduction. In our previous work [21], a MB position constraint model is also proposed for fast mode decision. However, these fast mode decision algorithms are still not efficient enough to early terminate the DIRECT mode before further checking other modes which still requires complex ME and DE. Since DIRECT mode does not need time-consuming ME and DE, if it can be early terminated, the process of ME and DE and the exhaustive mode decision of the remaining modes can be skipped. This will significantly decrease computational complexity.

Manuscript received January 6, 2016; revised March 28, 2016; accepted April 29, 2016. Date of publication May 30, 2016; date of current version August 31, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61572183 and Grant 61379143, in part by the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20120161110014, in part by the Priority Academic Program Development of Jiangsu Higer Education Institutions, and in part by the Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology.

Y. Li, G. Yang, N. Chen, and X. Ding are with the School of Information Science and Engineering, Hunan University, Changsha 410082, China (e-mail: yueli@hnu.edu.cn; yanggaobo@hnu.edu.cn; ningchen@163.com; xianglingding@163.com).

Y. Zhu is with the Faculty of Physics and Electronic Information Science, Hengyang Normal University, Hengyang 421002, China (e-mail: zyp1016@hynu.edu.cn).

Digital Object Identifier 10.1109/TBC.2016.2570018

TABLE I
TEST CONDITIONS

GOP length	12
Number of reference frames	2
Search range	64
Number of bi-prediction iteration	4
Search for iteration	8
Number of encoded frames	49
Basis quantization parameter (QP)	24,28,32,36

TABLE II
PERCENTAGE OF THE DIRECT MODE AS THE OPTIMAL MODE (UNIT:%)

QP	Flamenco1	Ballroom	Ballet	Champagner
24	64.66	61.72	77.84	83.10
28	70.80	68.32	82.20	87.74
32	77.22	72.21	85.69	92.86
36	82.72	76.25	88.80	95.21
average	73.85	69.63	83.63	89.73

TABLE III
STATISTICAL RESULTS OF PROBABILITY P(B|A) (UNIT:%)

QP	Flamenco1	Ballroom	Ballet	Champagner
24	68.54	70.08	92.89	94.55
28	75.69	82.43	94.45	96.37
32	84.18	84.39	95.53	97.64
36	90.53	86.55	96.52	98.58
average	79.74	80.86	94.84	96.79

TABLE IV
STATISTICAL RESULTS OF PROBABILITY P(D|C) (UNIT:%)

QP	Flamenco1	Ballroom	Ballet	Champagner
24	97.21	94.06	93.82	91.71
28	97.07	96.03	93.32	90.19
32	96.72	95.78	92.17	89.37
36	96.39	94.43	91.55	90.30
average	96.85	95.08	92.72	90.39

It is claimed in [22] that DIRECT mode is highly possible to be the optimal mode among all candidate modes of MVC. Thus, an early termination of DIRECT mode decision is an effective solution to reduce computational complexity without sacrificing coding efficiency. Zhang *et al.* [17] proposed an efficient statistical DIRECT mode early termination (SDMET) model for MVC, which estimates rate distortion degradation, false acceptance rate and false reject rate of early DIRECT mode decision. Shen *et al.* [22] also proposed an early SKIP mode decision algorithm by analyzing prediction mode distribution of those MBs in reference view. Moreover, an early DIRECT mode decision is proposed based on adaptive RD cost threshold [23], which is computed from spatial, temporal and inter-view RD costs. Zhang *et al.* [24] proposed an efficient DIRECT mode early decision algorithm by exploiting rate distortion cost and inter-view correlation. Pan *et al.* [25] proposed an early DIRECT mode decision algorithm based on all-zero block and RD cost. However, these early DIRECT mode decision approaches separately treat RD cost and mode correlation as the condition for early DIRECT mode termination.

In this paper, an efficient DIRECT mode decision approach is proposed for MVC by exploiting mode homogeneity and RD cost correlation to further reduce computational complexity and maintain good RD performance. The main contributions of the proposed approach are two-folds. First, mode homogeneity is firstly exploited by analyzing the mode distribution in spatial and inter-view. Secondly, an adaptive threshold is defined for early DIRECT mode termination, which is derived from spatial and inter-view correlation of MVV. Therefore, a significant speed-up gain is achieved by skipping the rest modes when it meets early termination conditions. The rest of this paper is organized as follows. The motivations and statistical analysis are presented in Section II. Section III presents the early DIRECT mode termination approach. Experimental results and analyses are provided in Section IV. Finally, conclusion is made in Section V.

II. MOTIVATIONS AND ANALYSES

Similar to H.264/AVC video coding standard, MVC adopts exhaustive mode decision to improve coding efficiency. For the homogeneous regions without motion or simply slow motion in natural video sequences, DIRECT mode is intuitively more likely to be the optimal mode among all candidate modes. In order to analyze the possibility of DIRECT mode selected as the optimal mode, four typical video sequences including Flamenco1 (320x240), Ballroom (640x480), Ballet (1024x768) and Champanger (1280x960) are encoded to analyze their best mode distributions with JMVC8.3.1. These video sequences are most representative since they have different motion activities. Among them, Flamenco1 and Ballroom have fast motion, Ballet has medium motion and Champanger has relatively slow motion. The test conditions for statistical analysis are summarized in Table I. Three views are used in the test experiments. Even views are treated as reference views and odd view is used to count the percentages of DIRECT mode when it is chosen as the optimal mode. Table II summarizes the statistical results.

From Table II, the percentages of DIRECT mode selected as the optimal mode vary from 69.63% to 89.73% for four video sequences. For videos with medium and slow motion, there are higher possibilities that DIRECT mode is selected as the optimal mode. The percentages are 83.63% and 89.73% for Ballet and Champagner, respectively. However, the percentage decreases significantly for video sequences with fast motions. It is well-known that DIRECT mode does not require computation-intensive ME and DE, which can achieve high encoding efficiency with little encoding time cost. Therefore, if DIRECT mode can be fully exploited and early terminated, encoding time will be significantly reduced.

Let A represent the event that DIRECT mode is selected as the optimal mode for current MB, and let B represent the fact that the mode of spatial neighboring MBs or inter-view collocated MBs whose optimal mode is also DIRECT mode, the MBs location correlation is shown in Fig. 1. In the following, the relationship between A and B is considered.

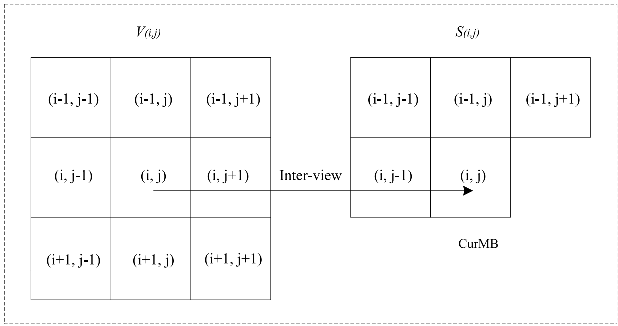


Fig. 1. Current MB and its spatial neighboring and inter-view collocated MBs.

Table III shows the statistical results of conditional probability P(B|A). From it, we can observe that more than 79.74% of those spatial neighboring or inter-view collocated MBs have the same DIRECT mode with current MB. Especially, Champagner sequence has the highest percentage up to 96.79%, simply because it has a relatively large region of static background. That is, the MBs in this regions have homogeneous modes. Even for Flamenco1 sequence with complex motion, it has a minimum probability of 79.74%. Further, we analyze the correlation of RD costs. Let C represent the RD cost of current MB whose mode is DIRECT mode, and D represent the event C is less than or equal to the maximum RD cost of spatial neighboring or inter-view collocated MBs whose modes are also DIRECT modes. Table IV shows the statistical results of probability P(D|C). From it, we know that the probability of P(D|C) is bigger than 90%. That is, when current MB is encoded with DIRECT mode, its RD cost has a strong correlation with those spatial neighboring or inter-view collocated MBs when they are also encoded with DIRECT mode. Thus, both mode homogeneity and RD cost correlation should be exploited for early termination of DIRECT mode decision. Actually, this is the main idea behind the proposed approach.

III. Proposed Early DIRECT Mode Decision Approach

A. Mode Homogeneity

Mode homogeneity (MH) refers to the fact those MBs within a region share the same mode. In this paper, a uniform mode field is defined to calculate mode homogeneity for each MB. Let $MB_{i,j}$ be a MB located at the ith row and jth column ($i \in [1, W/16], j \in [1, H/16]$), W and H be the width and the height of video frame, respectively. Mode homogeneity is firstly defined as

$$MH = \sum_{(i,j)\in Z} \left| M_{(i,j)} - argmin\left\{M_{(i,j)}\right\} \right| \tag{1}$$

where $M_{(i,j)}$ is the mode of $MB_{(i,j)}$, and Z represents spatial neighboring or inter-view collocated MBs of current MB, which is shown in Fig. 1. Therefore, spatial mode homogeneity is defined as

$$MH_s = \sum_{(i,j)\in Z_s} \left| M_{s(i,j)} - argmin\left\{M_{s(i,j)}\right\} \right| \tag{2}$$

where Z_s represents those spatial neighboring MBs, *i.e*, ($MB_{s(i-1,j-1)}$, $MB_{s(i-1,j)}$, $MB_{s(i-1,j+1)}$, $MB_{s(i,j-1)}$). Similarly, inter-view mode homogeneity is defined as

$$MH_v = \sum_{(i,j)\in Z_v} \left| M_{v(i,j)} - argmin\left\{M_{v(i,j)}\right\} \right| \tag{3}$$

where Z_v represents the inter-view collocated MBs, *i.e*, $MB_{v(i,j)}$, in previous view of current MB ($v(i) \in [i-1, i+1]$, $v(j) \in [j-1, j+1]$).

Current MB has a strong mode correlation with its spatial neighboring or inter-view collocated MBs. When MH equals zero and the minimum value of modes is DIRECT mode, current MB is more likely to select DIRECT mode as its optimal mode. Therefore, either MH_s or MH_v can be used to early determine whether current MB choose DIRECT mode as its optimal mode or not.

Following Equations (2) and (3), MH_s and MH_v are firstly computed for current MB, respectively. If MH_s or MH_v equals zero, the DIRECT mode is determined as the optimal mode, and mode decision is early terminated. To verify the early DIRECT mode decision strategy by exploiting mode homogeneity, the decrease of peak signal-to-noise ratio (ΔPSNR) and bit rate increase (ΔBR) are adopted for performance evaluation, which are computed as

$$\Delta PSNR = PSNR_{Pro} - PSNR_{Ori} \tag{4}$$

$$\Delta BR = (BR_{Pro} - BR_{Ori})/BR_{Ori} \times 100\% \tag{5}$$

where $PSNR_{Pro}$, BR_{Pro} are PSNR and bit rate of the proposed approach, respectively. And $PSNR_{Ori}$, BR_{Ori} are PSNR and bit rate of the original JMVC, respectively. The same test conditions in Table I are used here, and the experimental results are shown in Table V. From the second and third columns of Table V, we can observe that the average values of $\Delta PSNR$ and ΔBR are -1.26 dB and 77.69%, respectively. Apparently, both $\Delta PSNR$ and ΔBR are not acceptable if only mode homogeneity is exploited for early DIRECT mode decision. Actually, there are a large amount of MBs whose optimal modes are improperly selected as DIRECT mode by simply exploiting mode homogeneity. The reason is that some MBs are error classified into the DIRECT mode, and the error classified MBs are further used to predict other MBs' modes. In this way, the prediction mode errors might be accumulated. Therefore, mode homogeneity should be combined with other constrains to reduce prediction error as possible, so as to reduce the computational complexity while maintaining the RD performance.

B. RD Cost Correlation

Fig. 1 shows the location correlation between current MB and its spatial neighboring and inter-view collocated MBs. Current MB not only has mode correlation with these spatial neighboring and inter-view collocated MBs, but also has RD cost correlation with them. In fact, the original MVC selects the optimal mode from all candidate modes by checking RD costs for exhaustive mode decision, and the DIRECT mode is firstly checked. If the RD cost of DIRECT mode is smaller

TABLE V
THE PSNR AND BR ANALYSES FOR EARLY DIRECT MODE DECISION

	MH		RDCC		MH+RDCC	
Video sequences	ΔPSNR(dB)	ΔBR(%)	ΔPSNR(dB)	ΔBR(%)	ΔPSNR(dB)	ΔBR(%)
Flamenco1	-2.33	98.72	-0.34	-1.08	-0.04	-0.52
Ballroom	-0.91	97.78	-0.11	1.70	-0.02	-0.13
Ballet	-1.01	74.45	-0.09	-0.23	-0.02	-0.54
Champanger	-0.79	39.81	-0.04	-0.59	-0.04	-0.69
Average	-1.26	77.69	-0.15	-0.05	-0.03	-0.47

TABLE VI
PROBABILITY OF THE DIRECT MODE EARLY TERMINATE AS THE OPTIMAL MODE (UNIT: %)

	MH+RDCC					MH+RDCC+U(MH/RDCC)				
Video sequences	QP=24	QP=28	QP=32	QP=36	average	QP=24	QP=28	QP=32	QP=36	average
Flamenco1	60.14	62.38	65.97	73.25	65.43	95.29	96.51	97.79	98.65	97.06
Ballroom	60.98	71.38	72.38	74.32	69.76	94.16	95.65	94.43	96.78	95.25
Ballet	76.90	80.27	81.57	83.14	80.47	98.30	98.62	98.90	99.19	98.75
Champanger	78.54	80.81	82.17	84.97	81.62	98.78	99.16	99.42	99.64	99.25
Average	69.14	73.71	75.52	78.92	**74.32**	96.63	97.49	97.64	98.57	**97.58**

than an adaptive threshold, it can be early selected as the optimal mode and thus the remaining modes might be skipped. This will save the computational complexity of computing and checking the RD costs for the rest modes. Please note that selection of an adaptive threshold is the key issue for fast DIRECT mode decision because it compromises between computational complexity and RD performance. In this paper, two adaptive thresholds are defined based on the RD cost correlation for those spatial neighboring and inter-view collocated MBs as follows.

$$T_s = \underset{(i,j)\in Z_s}{argmax}\left\{w_{(i,j)} \cdot RD\cos t_{(i,j)}\right\} \tag{6}$$

$$T_v = \underset{(i,j)\in Z_v}{argmax}\left\{w_{(i,j)} \cdot RD\cos t_{(i,j)}\right\} \tag{7}$$

where Z_s and Z_v represent the locations of those spatial neighboring and inter-view collocated MBs, respectively. And $RD\cos t_{(i,j)}$ is the RD cost of MB, $w_{(i,j)}$ is used to check whether the optimal mode of MB is DIRECT mode or not. If $M_{(i,j)}$ is DIRECT mode, $w_{(i,j)}$ equals 1, otherwise it equals 0. That is, $w_{(i,j)}$ is defined as

$$w_{(i,j)} = \begin{cases} 1 & if\, M_{(i,j)} = DIRECT \\ 0 & others \end{cases} \tag{8}$$

The fourth and fifth columns of Table V, which are labeled as RDCC (RD cost correlation), show the RD performance of early DIRECT mode decision based on RDCC. It can be observed that the values of $\Delta PSNR$ and ΔBR are negligible for those video sequences with medium and slow motions. That is, a desirable encoding efficiency is achieved for them. However, for fast motion sequences such as Flamenco1 and Ballroom, their $\Delta PSNR$ are -0.34 dB and -0.11 dB, respectively. To achieve negligible RD degradation, mode homogeneity and RD cost correlation are jointly exploited to early terminate DIRECT mode decision. Thus, the condition of early DIRECT mode decision is defined as follow.

$$M_{(i,j)} = \begin{cases} DIRECT, & ((T_{DIRECT} \le T_s)\&\&(MH_s = 0)) \\ & \|((T_{DIRECT} \le T_v)\&\&(MH_v = 0)) \\ others\,, & otherwise \end{cases} \tag{9}$$

where T_{DIRECT} is RD cost of current MB when it is encoded with DIRECT mode, and the operators && and || are logical AND and OR, respectively. If MH_s or MH_v is equal to zero and T_{DIRECT} is less than thresholds T_s or T_v, DIRECT mode will be selected as the optimal mode for current MB, and the rest modes will be skipped.

The sixth and seventh columns of Table V, which are labeled as MH+RDCC, show the performance of early DIRECT mode decision by jointly exploiting mode homogeneity and RD cost correlation. From it, we know that $\Delta PSNR$ and ΔBR are -0.03 dB and 0.47% on average, respectively. It means that by jointly exploiting mode homogeneity and RD cost correlation for early DIRECT mode decision, a good tradeoff is achieved between RD performance and encoding efficiency. From Table VI, we further know that for video sequences with various motions, the probability of early determining DIRECT mode as the optimal mode varies from 65.43% to 81.62% under the conditions of MH+RDCC and the average probability is 74.32%. However, there are still a lots of MBs, whose modes should be early terminated as the DIRECT mode, are still required to check the remaining modes. To further early terminate DIRECT mode decision and reduce encoding time, the thresholds of mode homogeneity and RD cost are updated respectively, which is discussed in next subsection.

C. Updating Adaptive Thresholds of Mode Homogeneity and RD Cost

To further terminate DIRECT mode earlier and reduce computational complexity, the mode correlation between DIRECT mode and subsequent Inter16x16 mode is further analyzed, which is shown in Table VII. The test conditions are the same

TABLE VII
STATISTICAL RESULTS OF PROBABILITIES P(A|B), P(A|C) AND P(A|BC) (UNIT: %)

	QP	Flamenco1	Ballroom	Ballet	Champanger
P(A|B)	24	93.98	95.97	98.38	98.96
	28	95.58	96.43	98.68	99.29
	32	97.26	96.90	98.94	99.53
	36	98.39	97.23	99.27	99.73
	average	96.30	96.63	98.82	99.38
P(A|C)	24	97.62	97.30	98.85	99.56
	28	97.91	97.53	99.20	99.69
	32	98.72	97.90	99.42	99.77
	36	99.26	98.20	99.58	99.87
	average	98.38	97.73	99.26	99.72
P(A|BC)	24	98.44	98.94	99.42	99.68
	28	98.62	99.04	99.54	99.77
	32	99.04	99.07	99.64	99.82
	36	99.34	99.10	99.73	99.89
	average	98.86	99.04	99.58	99.79

with Table I. Let A represent the fact that the optimal mode of current MB is either DIRECT or Inter16x16 mode, B represent the fact that the optimal mode of spatial adjacent MBs is also either DIRECT or Inter16x16 mode, and event C denote that the optimal mode of inter-view collocated MBs is either DIRECT mode or Inter16x16 mode as well. Consequently, event BC represents the fact that both B and C are satisfied simultaneously. Table VII reports the statistical results. Apparently, the probabilities P(A|B), P(A|C) and P(A|BC) are about 96.30%-99.38%, 97.73%-99.72% and 98.86%-99.79% for four sequences with different motions, respectively. Thus, they always meet

$$P(A|BC) > P(A|C) > P(A|B) \tag{10}$$

This implies the fact that jointly exploiting the spatial and inter-view mode correlation can improve the accuracy of early DIRECT mode decision. Thus, we are motivated to update the threshold of mode homogeneity by jointly exploiting spatial and inter-view mode correlation after checking Inter16x16, which is defined as

$$MH_{s\_v} = \sum_{(i,j)\in Z_{s\_v}} \left| M_{s\_v(i,j)} - argmin\left\{M_{s\_v(i,j)}\right\}\right| \tag{11}$$

where $M_{s\_v(i,j)}$ is the mode of those spatial neighboring and inter-view collocated MBs, $argmin\{M_{s\_v(i,j)}\}$ is either the DIRECT mode or the Inter16x16 mode, and Z_{s_v} indicates the positions of those spatial neighboring and inter-view collocated MBs, as shown in Fig. 1. Meanwhile, by jointly exploiting the spatial and inter-view RD cost correlations after checking Inter16x16 mode, the threshold T_{s_v} is adaptively refined as

$$T_{s\_v} = \underset{(i,j)\in Z_{s\_v}}{argmax}\left\{w_{(i,j)} \cdot RD\cos t_{(i,j)}\right\} \tag{12}$$

where $w_{(i,j)}$ is defined in equation (8). After checking Inter16x16 mode, the condition of early DIRECT mode decision for current MB is defined as

$$M_{(i,j)} = \begin{cases} 16x16, & MH_{v\_s} = 0 \,||T_{D\_16} \le T_{v\_s} \\ others, & otherwise \end{cases} \tag{13}$$

where 16x16 represents DIRECT mode or Inter16x16 mode, T_{D_16} is the minimum RD cost of current MB when it is encoded with either DIRECT or Inter16x16 mode.

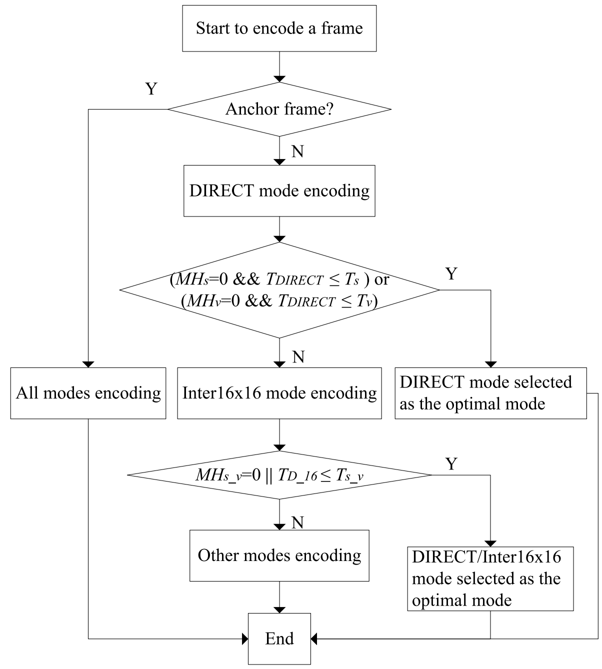


Fig. 2. Flowchart of the proposed algorithm.

In Table VI, the right column MH+RDCC+U(MH/RDCC) reports the experimental results by combining MH, RDCC and U(MH/RDCC) which are presented in subsections *A*, *B*, and *C*, respectively. From it, it is apparent that the correct probability of early terminating DIRECT mode as the optimal mode increases from 95.25% to 99.25%, and its average is 97.58% for four video sequences with different motion activities. That is, DIRECT mode decision is early terminated in an efficient and accurate way.

D. The Proposed Early DIRECT Mode Decision Approach

The flowchart of the proposed early DIRECT mode decision approach is shown in Fig. 2. The main steps are summarized as follows.

1) Start to encode a frame.
2) If current frame is an anchor frame, encode all the MBs in this frame with all candidate modes, and then jump to step 5. Otherwise, encode current MB with DIRECT mode and obtain its RD cost, which is denoted as T_{DIRECT}, and go to step 3.
3) If it meets the condition $((T_{DIRECT} \le T_s)\&\&(MH_s = 0))\,||((T_{DIRECT} \le T_v)\&\&(MH_v = 0))$, simply select DIRECT mode as the optimal mode and thus early terminate the mode decision process. Otherwise, go to step 4.
4) Encode current MB with Inter16x16 mode, compare its RD cost with the RD cost T_{DIRECT} of DIRECT

TABLE VIII
THE RESULTS OF PROPOSED METHOD WITH PANIET [25] AND MH+RDCC UNDER DIFFERENT VIDEO SEQUENCES

		PanIET[25] vs JMVC			MH+RDCC vs JMVC		
Sequences	Resolution	BDPSNR(dB)	BDBR(%)	ΔTS(%)	BDPSNR(dB)	BDBR(%)	ΔTS(%)
Flamenco1	320x240	-0.007	0.14	-46.85	-0.005	0.12	-43.37
Race1	320x240	0.000	-0.01	-60.05	-0.012	0.22	-78.57
Golf1	320x240	-0.028	0.56	-60.58	-0.014	0.26	-75.39
Ballroom	640x480	-0.010	0.25	-42.84	-0.010	0.31	-42.46
Exit	640x480	-0.013	0.58	-53.00	-0.019	0.66	-62.42
Vassar	640x480	-0.011	0.44	-58.78	-0.010	0.46	-69.62
Breakdancers	1024x768	-0.032	1.29	-38.34	-0.016	0.69	-31.14
Ballet	1024x768	-0.006	0.22	-54.29	-0.004	0.11	-59.47
Doorflowers	1024x768	-0.013	0.43	-58.52	-0.007	0.20	-70.18
Dog	1280x960	-0.008	0.27	-52.04	0.010	-0.35	-55.07
Champanger	1280x960	-0.015	0.42	-56.95	-0.011	0.31	-67.23
Pantomime	1280x960	-0.013	0.34	-46.59	-0.001	0.02	-43.07
Average		-0.013	0.41	-52.40	-0.008	0.25	-58.17

mode and choose the mode between DIRECT mode and Inter16x16 mode that have less RD cost. The minimum RD cost is denoted as T_{D_16}, if $T_{D\_16} \leq T_{s\_v}$ or $MH_{s\_v} = 0$, early terminate the DIRECT mode decision. Otherwise, encode it with the remaining modes and go to step 5.

5) Encode next MB.

IV. EXPERIMENTAL RESULTS AND ANALYSES

A. Test Conditions

To evaluate the coding performance of the proposed approach, twelve typical video sequences including Flomenco1 (320x240), Golf1 (320x240), Race1 (320x240), Ballroom (640x480), Exit (640x480), Vassar (640x480), Ballet (1024x768), Breakdancers (1024x768), Doorflowers (1024x768), Dog (1280x960), Champanger (1280x960) and Pantomime (1280x960) are selected for experiments because they have quite different spatial resolutions and motion activities. The MVC reference software (JMVC8.3.1) is used as the software platform for experiments. That is, the proposed early DIRECT mode decision approach is integrated into JMVC8.3.1. The test conditions are summarized in Table I. Three views are used in the experiment, the even views are treated as the reference views and the odd view is used as the implementation of the proposed algorithm. To verify the validity of the proposed method, it is compared with state-of-the-art algorithms including ZhangTB [24], PanIET [25] and PanTB [12]. ΔPSNR, ΔBR, BDPSNR, BDBR [26] and the total consumed CPU time are used as the metrics of performance evaluation. Experimental results are summarized and compared in Table VIII and IX, where TS is the percentage of encoding time reduction compared with JMVC. It is defined as:

$$TS = \frac{T_\theta - T_{JMVC}}{T_{JMVC}} \times 100\% \tag{14}$$

where T_θ represents the total encoding time, $\theta \in$ {ZhangTB [24], PanIET [25], PanTB [12], proposed}. T_{JMVC} is the total encoding time of the original JMVC.

B. Experimental Results of the Proposed Approach and PanIET

Since PanIET [25] compares experimental results with [22] and [23], Table VIII only reports the experimental results of PanIET and the proposed MH+RDCC approach. Please note that the experimental results of PanIET and the proposed approach are provided by making comparisons with the same ground-truth, i.e., the original JMVC encoder. PanIET reduces encoding time from 38.34% to 60.58% (52.40% on average), its BDPSNR degrades from 0 dB to 0.028 dB (0.013 dB on average) and the BDBR increases from -0.01% to 1.29% (0.41% on average). For the proposed approach, the saving of total coding time varies from 31.14% to 78.57% (58.17% on average), BDPSNR degrades from 0.001 dB to 0.019 dB (0.008 dB on average) and the BDBR increases from -0.35% to 0.69% (0.25% on average). Thus, compared with PanIET, the proposed approach achieves 5.77% time reduction at the costs of 0.005 dB BDPSNR increase and 0.16% BDBR decrease. Specifically, the proposed approach achieves similar RD performance and encoding time reduction with PanIET for video sequences with fast and medium motions. For slow motion video sequences, it achieves more desirable RD performance and reduces more encoding time.

C. Comparisons Between the Proposed Approach and the State-of-the-Art Algorithms

Table IX compares the RD performance and coding efficiency among ZhangTB [24], PanIET [25], PanTB [12] and the proposed approach. Their performance are reported by making comparisons with the same original JMVC, respectively. From Table IX, it can be observed that ZhangTB reduces computational complexity from 58.03% to 90.69% (76.67% on average). Meanwhile, the BDPSNR degrades from 0.018 dB to 0.163 dB (0.069 dB on average) and the BDBR increases 0.62% to 5.36% (2.24% on average). PanTB reduces the total encoding time from 47.73% to 75.06% (68.16% on average), and its BDPSNR degrades from -0.013 dB to 0.325 dB (0.097 dB on average) and the BDBR increases from -0.45% to 6.05%

TABLE IX
COMPARISONS AMONG THE PROPOSED APPROACH WITH ZHANGTB [24], PANIET [25] AND PANTB [12]

		ZhangTB[24] vs JMVC			PanIET[25] vs JMVC			PanTB[12] vs JMVC			Proposed vs JMVC		
Sequences	QP	ΔPSNR (dB)	ΔBR (%)	TS (%)	ΔPSNR (dB)	ΔBR (%)	TS (%)	ΔPSNR (dB)	ΔBR (%)	TS (%)	ΔPSNR (dB)	ΔBR (%)	TS (%)
Flamenco1	24	-0.081	0.01	-64.07	-0.027	-0.26	-44.16	-0.035	0.97	-70.92	-0.069	0.06	-66.84
	28	-0.086	-0.24	-69.31	-0.029	-0.52	-45.60	-0.038	1.79	-73.50	-0.096	-1.32	-71.26
	32	-0.105	-0.33	-74.20	-0.011	-0.11	-47.48	-0.051	1.24	-75.90	-0.102	-0.04	-76.04
	36	-0.172	-1.32	-78.32	-0.020	0.01	-50.16	-0.194	1.49	-77.28	-0.115	-1.27	-79.80
	Average	-0.111	-0.47	-71.48	-0.022	-0.22	-46.85	-0.080	1.37	-74.40	-0.095	-0.64	-73.49
BDPSNR(dB)/BDBR(%)		**-0.080**	**1.53**	**-71.48**	**-0.007**	**0.14**	**-46.85**	**-0.138**	**2.51**	**-74.40**	**-0.056**	**1.18**	**-73.49**
Race1	24	-0.065	-0.42	-87.00	-0.035	-0.32	-62.06	-0.011	2.95	-73.77	-0.051	-0.38	-89.64
	28	-0.028	0.01	-87.01	-0.011	-0.27	-59.09	-0.032	5.46	-74.72	-0.022	0.17	-89.21
	32	-0.010	0.40	-86.90	-0.004	-0.18	-60.36	-0.028	5.99	-76.00	-0.008	1.08	-88.54
	36	-0.008	0.53	-86.25	-0.001	0.00	-58.70	-0.073	7.00	-75.76	-0.010	0.20	-88.09
	Average	-0.028	0.13	-86.79	-0.012	-0.19	-60.05	-0.036	5.35	-75.06	-0.023	0.27	-88.87
BDPSNR(dB)/BDBR(%)		**-0.034**	**0.62**	**-86.79**	**0.000**	**-0.01**	**-60.05**	**-0.325**	**6.05**	**-75.06**	**-0.040**	**0.78**	**-88.87**
Golf1	24	-0.122	-0.11	-89.54	-0.095	0.08	-64.20	0.000	0.71	-64.60	-0.093	-0.08	-91.18
	28	-0.042	-0.51	-90.46	-0.036	-0.12	-61.77	-0.006	0.33	-65.58	-0.037	-0.46	-91.69
	32	-0.023	-0.53	-91.10	-0.016	0.01	-58.04	-0.003	1.04	-67.59	-0.021	-0.45	-91.51
	36	-0.007	-0.28	-91.68	-0.002	-0.01	-58.31	0.005	3.44	-70.61	-0.005	-0.21	-91.43
	Average	-0.049	-0.36	-90.69	-0.037	-0.01	-60.58	-0.001	1.38	-67.10	-0.039	-0.30	-91.45
BDPSNR(dB)/BDBR(%)		**-0.018**	**0.33**	**-90.69**	**-0.028**	**0.56**	**-60.58**	**-0.047**	**0.99**	**-67.10**	**-0.014**	**0.27**	**-91.45**
Ballroom	24	-0.048	-0.46	-56.32	-0.029	-0.58	-41.65	-0.012	1.19	-63.66	-0.042	-0.05	-64.21
	28	-0.032	0.10	-60.83	-0.021	-0.29	-43.15	-0.019	1.77	-66.35	-0.041	0.40	-68.81
	32	-0.049	0.25	-64.25	-0.016	-0.28	-43.37	-0.049	2.10	-66.76	-0.050	0.42	-70.95
	36	-0.067	1.19	-67.02	-0.007	0.01	-43.18	-0.044	2.14	-69.51	-0.054	0.16	-73.81
	Average	-0.049	-0.27	-62.10	-0.018	-0.29	-42.84	-0.031	1.80	-66.57	-0.047	0.23	-69.45
BDPSNR(dB)/BDBR(%)		**-0.054**	**1.50**	**-62.10**	**-0.010**	**0.25**	**-42.84**	**-0.099**	**2.80**	**-66.57**	**-0.058**	**1.62**	**-69.45**
Exit	24	-0.056	-1.30	-72.84	-0.035	-0.88	-51.01	-0.036	1.44	-68.57	-0.051	-0.62	-78.69
	28	-0.064	-0.27	-78.50	-0.020	-0.39	-52.93	-0.051	2.22	-70.52	-0.057	-0.23	-82.58
	32	-0.054	0.55	-81.20	-0.014	-0.09	-53.65	-0.035	3.06	-73.12	-0.038	0.39	-84.45
	36	-0.073	1.58	-83.67	-0.013	-0.10	-54.42	-0.073	3.25	-75.50	-0.050	0.17	-86.28
	Average	-0.062	0.14	-79.05	-0.020	-0.36	-53.00	-0.049	2.49	-71.93	-0.049	-0.07	-83.00
BDPSNR(dB)/BDBR(%)		**-0.069**	**2.71**	**-79.05**	**-0.013**	**0.58**	**-53.00**	**-0.111**	**4.55**	**-71.93**	**-0.051**	**2.07**	**-83.00**
Vassar	24	-0.064	-1.43	-75.75	-0.052	-0.83	-55.26	-0.014	0.39	-66.43	-0.048	-1.29	-82.81
	28	-0.060	-0.67	-85.87	-0.030	-0.27	-60.93	-0.011	1.02	-66.41	-0.035	-0.73	-87.50
	32	-0.025	-0.59	-87.38	-0.005	-0.25	-60.31	-0.043	0.30	-66.09	-0.013	-0.49	-88.61
	36	-0.010	-0.27	-88.48	-0.002	-0.10	-58.64	0.005	1.72	-65.90	-0.009	-0.19	-88.97
	Average	-0.040	-0.74	-84.37	-0.022	-0.36	-58.78	-0.016	0.86	-66.21	-0.026	-0.68	-86.97
BDPSNR(dB)/BDBR(%)		**-0.032**	**1.51**	**-84.37**	**-0.011**	**0.44**	**-58.78**	**-0.035**	**1.91**	**-66.21**	**-0.015**	**0.72**	**-86.97**
Breakdancers	24	-0.061	-0.66	-44.66	-0.045	-1.10	-32.92	-0.019	1.74	-66.31	-0.060	-0.48	-58.25
	28	-0.082	-0.39	-56.00	-0.048	-0.63	-37.81	-0.041	2.58	-68.48	-0.056	0.07	-63.50
	32	-0.105	-0.22	-63.01	-0.039	-0.54	-40.35	-0.050	3.32	-69.52	-0.049	-0.30	-67.21
	36	-0.145	-0.90	-68.42	-0.035	-0.37	-42.27	-0.123	2.93	-69.69	-0.064	-0.97	-71.35
	Average	-0.099	-0.54	-58.03	-0.042	-0.66	-38.34	-0.058	2.64	-68.50	-0.058	-0.42	-65.08
BDPSNR(dB)/BDBR(%)		**-0.082**	**3.69**	**-58.03**	**-0.032**	**1.29**	**-38.34**	**-0.113**	**5.00**	**-68.50**	**-0.049**	**2.13**	**-65.08**
Ballet	24	-0.059	0.52	-74.20	-0.019	-0.57	-54.97	-0.004	4.49	-69.38	-0.032	-0.69	-77.60
	28	-0.096	1.60	-77.39	-0.011	-0.25	-54.20	-0.010	4.51	-70.86	-0.031	-0.61	-79.26
	32	-0.142	1.72	-79.69	-0.010	-0.14	-53.85	-0.059	3.89	-71.59	-0.031	-0.24	-80.65
	36	-0.171	1.80	-81.62	-0.008	0.03	-54.14	-0.075	2.63	-72.62	-0.040	-0.58	-82.78
	Average	-0.117	1.41	-78.22	-0.012	-0.23	-54.29	-0.037	3.88	-71.11	-0.034	-0.53	-80.07
BDPSNR(dB)/BDBR(%)		**-0.163**	**5.36**	**-78.22**	**-0.006**	**0.22**	**-54.29**	**-0.150**	**5.11**	**-71.11**	**-0.018**	**0.64**	**-80.07**
Doorflowers	24	-0.104	0.55	-82.89	-0.036	-0.61	-59.52	-0.031	0.22	-70.23	-0.050	-0.01	-86.82
	28	-0.098	1.01	-85.13	-0.023	-0.28	-58.89	-0.029	0.66	-70.84	-0.054	0.32	-88.52
	32	-0.089	1.13	-86.22	-0.018	-0.67	-57.88	-0.018	-0.17	-71.74	-0.042	-0.05	-89.20
	36	-0.092	1.62	-86.84	-0.015	-0.19	-57.76	-0.045	-1.37	-71.62	-0.039	0.32	-89.49
	Average	-0.096	1.08	-85.27	-0.023	-0.44	-58.52	-0.031	-0.16	-71.11	-0.046	0.14	-88.51
BDPSNR(dB)/BDBR(%)		**-0.124**	**4.92**	**-85.27**	**-0.013**	**0.43**	**-58.52**	**-0.029**	**1.05**	**-71.11**	**-0.054**	**2.01**	**-88.51**
Dog	24	-0.068	-1.55	-68.46	-0.038	-1.01	-51.36	-0.013	-0.63	-43.36	-0.043	-1.40	-76.22
	28	-0.063	-1.11	-73.00	-0.026	-0.56	-52.30	-0.013	-0.76	-46.18	-0.040	-1.53	-78.20
	32	-0.064	-1.69	-76.27	-0.017	-0.50	-52.18	-0.010	-0.81	-49.41	-0.035	-1.17	-79.65
	36	0.064	-1.40	-78.47	-0.014	-0.41	-52.31	-0.019	-1.67	-51.95	-0.034	-1.59	-81.25
	Average	-0.065	-1.43	-74.05	-0.024	-0.62	-52.04	-0.014	-0.97	-47.73	-0.038	-1.42	-78.83
BDPSNR(dB)/BDBR(%)		**-0.025**	**0.82**	**-74.05**	**-0.008**	**0.27**	**-52.04**	**0.013**	**-0.45**	**-47.73**	**0.002**	**-0.02**	**-78.83**
Champanger	24	-0.119	1.66	-79.87	-0.047	-0.43	-55.69	-0.076	0.07	-74.11	-0.065	-0.59	-84.06
	28	-0.115	-0.31	-83.59	-0.034	-0.30	-56.23	-0.239	-0.56	-72.80	-0.064	-0.76	-86.02
	32	-0.086	-0.25	-86.47	-0.015	-0.22	-56.86	-0.042	-0.21	-72.81	-0.039	-1.01	-87.87
	36	-0.100	-0.31	-88.57	-0.013	-0.50	-59.04	-0.012	0.01	-74.29	-0.031	-1.20	-89.84
	Average	-0.105	0.20	-84.62	-0.027	-0.36	-56.95	-0.092	-0.17	-73.50	-0.050	-0.89	-86.95
BDPSNR(dB)/BDBR(%)		**-0.097**	**2.79**	**-84.62**	**-0.015**	**0.42**	**-56.95**	**-0.108**	**2.93**	**-73.50**	**-0.018**	**0.49**	**-86.95**
Pantomime	24	-0.091	-1.31	-59.82	-0.045	-0.44	-44.17	-0.100	0.27	-67.07	-0.057	-0.69	-68.46
	28	-0.128	-2.00	-62.59	-0.055	-1.15	-43.94	-0.074	-0.67	-63.66	-0.068	-1.71	-69.57
	32	-0.062	-1.70	-67.67	-0.041	-1.24	-47.22	-0.018	-1.11	-62.48	-0.054	-2.64	-73.91
	36	-0.071	0.29	-71.57	-0.035	-0.23	-51.04	-0.012	-1.87	-65.47	-0.054	-1.56	-77.88
	Average	-0.088	-1.18	-65.41	-0.044	-0.76	-46.59	-0.051	-0.85	-64.67	-0.058	-1.65	-72.46
BDPSNR(dB)/BDBR(%)		**-0.044**	**1.14**	**-65.41**	**-0.013**	**0.34**	**-46.59**	**-0.025**	**0.65**	**-64.67**	**0.005**	**-0.20**	**-72.46**
Average		**-0.069**	**2.24**	**-76.67**	**-0.013**	**0.41**	**-52.40**	**-0.097**	**2.76**	**-68.16**	**-0.031**	**0.97**	**-80.43**

(2.76% on average). For the proposed approach, the reduction of encoding time varies from 65.08% to 91.45% (80.43% on average), the BDPSNR degrades from -0.002 dB to 0.058 dB (0.031 dB on average), and the BDBR increases from -0.20% to 2.13% (0.97% on average). Compared with ZhangTB and PanTB, the proposed approach achieves 3.76% and 12.27%

encoding time reduction on average, the BDPSNR increases 0.038 dB and 0.066 dB on average, the BDBR degrades 1.27% and 1.79% on average, respectively. Especially, the proposed approach achieves about 28% encoding time saving without sacrificing RD performance compared with PanIET.

For video sequences with low motion activities such as Golf1, Vassar, Doorflowers and Champanger, the proposed approach reduces total encoding time more than 86.95% when it is compared with the original JMVC. For low motion video sequences, there are more stationary and homogeneous regions. It is apparent that MBs in the regions are more likely to choose DIRECT mode as their optimal mode. Meanwhile, the proposed approach also saves about 65.08%-69.45% encoding time for video sequences with complex motions such as Flamenco1, Ballroom and Breakdancers. Please note that compared with the proposed method, PanTB reduces more encoding time (about 3.42%) for Breakdancers sequence. However, the proposed method achieves 0.064 dB BDPSNR increase and 2.87% BDBR decrease, respectively. That is, the proposed method achieve much better RD performance for Breakdancers sequence. The reasons can be explained as follows. Breakdancers has a highly complex motion activity, which is captured by an arc cameras array. There are large differences between neighboring frames even captured by different cameras at the same time instant, and thus the inter-view correlation is further reduced. Since spatial and inter-view correlation is exploited for fast mode decision, the percentage of MBs whose modes are DIRECT mode is decreased, and thus the reduction of encoding time is decreased as well. In summary, the proposed method outperforms the state-of-the-art approaches, and its superior RD performance is achieved by early DIRECT mode decision as accurate as possible, which jointly exploits MB mode homogeneity and RD cost correlation.

V. Conclusion

In this paper, an early DIRECT mode decision approach is proposed to reduce the computational complexity of MVC encoder. Specifically, mode homogeneity and RD cost correlation are jointly exploited to early terminate mode decision for MVC. Extensive experiments on typical video sequences with different motion activities and spatial resolutions show that the proposed approach outperforms the state-of-the-art approaches. It can significantly speed up mode decision with a negligible loss of RD performance. Moreover, the proposed early DIRECT mode decision approach is suitable for various video sequences, especially video sequences with low motion activities. In future work, we will investigate the learning algorithm [26], [27] for the classification of all candidate modes in the training stage, so as to further reduce the computational complexity.

Acknowledgment

The authors greatly appreciate the kind help from Dr Pan Zhaoqing and Dr Zhang Yun for providing the source codes for experimental comparisons. The authors would also like to thank the anonymous reviewers for their valuable comments.

References

[1] A. Vetro, A. M. Tourapis, K. Muller, and T. Chen, "3D-TV content storage and transmission," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 384–394, Jun. 2011.

[2] C. Zhu, Y. Zhao, L. Yu, and M. Tanimoto, Eds., "3D-TV system with depth-image-based rendering: Architecture, techniques and challenges," New York, NY, USA: Springer, 2013.

[3] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo, "Free-viewpoint TV," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 67–76, Jan. 2011.

[4] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard," *Proc. IEEE*, vol. 99, no. 4, pp. 626–642, Apr. 2011.

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

[6] Z.-P. Deng, Y.-L. Chan, K.-B. Jia, C.-H. Fu, and W.-C. Siu, "Fast motion and disparity estimation with adaptive search range adjustment in stereoscopic video coding," *IEEE Trans. Broadcast.*, vol. 58, no. 1, pp. 24–33, Mar. 2012.

[7] L. Q. Shen, Z. Liu, S. X. Liu, Z. Y. Zhang, and P. An, "Selective disparity 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.

[8] L. Q. Shen, Z. Liu, T. Yao, Z. Y. Zhang, and P. An, "View-adaptive motion estimation and disparity estimation for low complexity multiview video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 925–930, Jun. 2010.

[9] C. Zhu, X. Lin, and L.-P. Chau, "Hexagon-based search pattern for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 5, pp. 349–355, May 2002.

[10] C. Zhu, X. Lin, L. Chau, and L.-M. Po, "Enhanced hexagonal search for fast block motion estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 10, pp. 1210–1214, Oct. 2004.

[11] Z. Q. Pan, S. Kwong, and Y. Zhang, "A multiple hexagon search algorithm for motion and disparity estimation in multiview video coding," in *The Era of Interactive Media*. New York, NY, USA: Springer, 2012, pp. 113–122.

[12] Z. Q. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcast.*, vol. 61, no. 2, pp. 166–176, Jun. 2015.

[13] Z. J. Peng, G. Y. Jiang, M. Yu, and Q. Dai, "Fast macroblock mode selection algorithm for multiview video coding," *EURASIP J. Image Video Process.*, vol. 58, no. 1, pp. 24–33, Mar. 2008.

[14] W. Zhu, X. Tian, F. Zhou, and Y. W. Chen, "Fast inter mode decision based on textural segmentation and correlations for multiview video coding," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1696–1704, Aug. 2010.

[15] L. Q. Shen, Z. Liu, P. An, R. Ma, and Z. Y. Zhang, "Low-complexity mode decision for MVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 837–843, Jun. 2011.

[16] H. Q. Zeng, K.-K. Ma, and C. H. Cai, "Fast mode decision for multiview video coding using mode correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 11, pp. 1659–1666, Nov. 2011.

[17] Y. Zhang, S. Kwong, G. Y. Jiang, X. Wang, and M. Yu, "Statistical early termination model for fast mode decision and reference frame selection in multiview video coding," *IEEE Trans. Broadcast.*, vol. 58, no. 1, pp. 10–23, Mar. 2012.

[18] C. H. Yeh *et al.*, "Fast mode decision algorithm through inter-view rate-distortion prediction for multiview video coding system," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 594–603, Feb. 2014.

[19] T. S. Zhao *et al.*, "Multiview coding mode decision with hybrid optimal stopping model," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1598–1609, Apr. 2013.

[20] H. Q. Zeng, X. L. Wang, C. H. Cai, J. Chen, and Y. Zhang, "Fast multiview video coding using adaptive prediction structure and hierarchical mode decision," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1566–1578, Sep. 2014.

[21] Y. Li, G. B. Yang, Y. P. Zhu, C. Liu, and K. Liu, "Adaptive mode decision for multiview video coding based on macroblock position constraint model," *J. Real Time Image Process.*, pp. 1–8, Aug. 2015, doi: 10.1007/s11554-015-0527-1.

[22] 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.

[23] F. S. Wang, H. Q. Zeng, Q. H. Shen, and S. D. Du, "Efficient early direct mode decision for multi-view video coding," *Signal Process. Image Commun.*, vol. 28, no. 7, pp. 736–744, Aug. 2013.

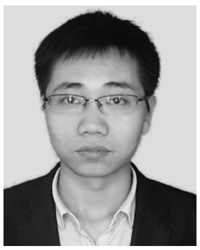
[24] Y. Zhang, S. Kwong, L. Xu, and G. Y. Jiang, "DIRECT mode early decision optimization based on rate distortion cost property and inter-view correlation," *IEEE Trans. Broadcast.*, vol. 59, no. 2, pp. 390–398, Jun. 2013.

[25] Z. Q. Pan, Y. Zhang, J. J. Lei, L. Xu, and X. M. Sun, "Early DIRECT mode decision based on all-zero block and rate distortion cost for multiview video coding," *IET Image Process.*, vol. 10, no. 1, pp. 9–15, Jan. 2016.

[26] G. Bjontegaard, *Calculation of Average PSNR Differences Between RD-Curves*, document VCEG-M33, ITU-T VCEG, Geneva, Switzerland, Apr. 2001.

[27] B. Gu, V. S. Sheng, K. Y. Tay, W. Romano, and S. Li, "Incremental support vector learning for ordinal regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 7, pp. 1403–1416, Jul. 2015.

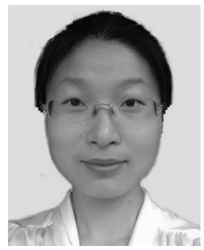
[28] B. Gu *et al.*, "Incremental learning for *v*-support vector regression," *Neural Netw.*, vol. 67, pp. 140–150, Jul. 2015.



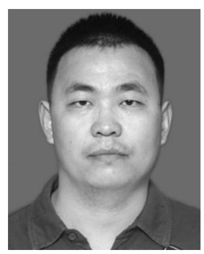
Yue Li received the B.S. degree from Hengyang Normal University, in 2010, and the M.S. degree from Central South University, in 2013. He is currently pursuing the Ph.D. degree with Hunan University, China. His current research interests include video coding and image processing.

Gaobo Yang received the master's degree from East China Jiaotong University, in 2001, and the Ph.D. degree from Shanghai University, in 2004. He is a Professor with the School of Information Science and Engineering, Hunan University, China. From 2010 to 2011, he made an academic visit to the University of Surrey, U.K. He has published over 60 papers in international journals. His research interests include video information security and multimedia communication.

Ning Chen received the master's degree from Hunan University, in 2011. He is a Senior Engineer with Hunan Mobile Corporation Ltd. His research interest is video communication over wireless networks.



Yapei Zhu received the B.S. degree from Hengyang Normal University, in 2010, and the M.S. degrees from Ningbo University, in 2013. She is currently a Teaching Assistant with Hengyang Normal University, China. Her research interests concentrate on image/video compression.



Xiangling Ding received the B.S. and M.S. degrees from Hunan Normal University, in 2003 and 2006, respectively. He is currently pursuing the Ph.D. degree with Hunan University, China. His current research interests include video information security and image processing.