DIRECT Mode Early Decision Optimization Based on Rate Distortion Cost Property and Inter-view Correlation

Yun Zhang, Sam Kwong, *Senior Member, IEEE,* Long Xu, and Gangyi Jiang, *Member, IEEE*

***Abstract*—In this paper, an Efficient DIRECT Mode Early Decision (EDMED) algorithm is proposed for low complexity multiview video coding. Two phases are included in the proposed EDMED: 1) early decision of DIRECT mode is made before doing time-consuming motion estimation/disparity estimation, where adaptive rate-distortion (RD) cost threshold, inter-view DIRECT mode correlation and coded block pattern are jointly utilized; and 2) false rejected DIRECT mode macroblocks of the first phase are then successfully terminated based on weighted RD cost comparison between 16×16 and DIRECT modes for further complexity reduction. Experimental results show that the proposed EDMED algorithm achieves 11.76% more complexity reduction than that achieved by the state-of-the-art SDMET for the temporal views. Also, it achieves a reduction of 50.98% to 81.13% (69.15% on average) in encoding time for inter-view, which is 29.31% and 15.03% more than the encoding time reduction achieved by the state-of-the-art schemes. Meanwhile, the average Peak Signal-to-Noise Ratio (PSNR) degrades 0.05 dB and average bit rate increases by −0.37%, which is negligible.**

***Index Terms*—Digital video broadcasting, multiview video coding, mode decision, early termination, DIRECT mode.**

I. Introduction

Multiview video can provide real depth perception, interactivity and novel visual enjoyment and it would be useful for many new multimedia applications, such as Three Dimensional Television (3DTV) [1] broadcasting, Free-viewpoint TeleVision (FTV) [2], immersive teleconference and virtual reality. Since multiview video is captured from multiple cameras from different viewpoints or angles simultaneously, it has a huge amount of data due to high spatial-view-temporal redundancies, and shall be efficiently encoded [3]. To this end, Mutliview Video Coding (MVC) was developed by the Joint Video Team (JVT) of Video Coding Experts Group (VCEG) and Moving Pictures Experts Group (MPEG), where hierarchical MVC prediction structure, intensive Disparity Estimation (DE) and Motion Estimation (ME) prediction technologies are adopted to exploit inter-view and temporal redundancies. In addition, variable block-size mode decision and Multi-Reference Frame (MRF) techniques are utilized to further improve the coding efficiency. However, these advanced technologies cause extremely high computational complexity which has been a bottleneck of applying multiview video into real-time realistic media applications, such as 3D live broadcasting and interactive FTV.

To reduce the computational complexity of variable block-size DE/ME, a number of efforts have been devoted to develop low complexity algorithms in different levels, including fast DE/ME [4], fast MRF selection [5] and Fast Mode Decision (FMD) [6]–[10]. In terms of low complexity mode decision, Zhao *et al.* [6] proposed a FMD algorithm for H.264/AVC based on mode priority. Wang *et al.* [7] proposed All Zero Block (AZB) detection algorithm early mode decision and ME in H.264/AVC. Jung *et al.* [8] applied the AZB detection technique to FMD for scalable video coding. In [9], Pan and Ho proposed another early mode decision for H.264/AVC inter-prediction by defining a threshold based on sum of absolute transformed difference of SKIP. Its threshold linearly increases with Quantization Parameter (QP). Zeng *et al.* [10] proposed motion activity-based early mode decision by using two Rate-Distortion (RD) cost thresholds for SKIP and INTRA modes, respectively. They fitted the RD cost thresholds with exponential relationship of QP for P frames. However, these schemes were proposed for mono-view H.264/AVC and cannot be directly applied for MVC due to different statistical characteristics. Also, inter-view MB mode correlation has not been effectively exploited.

Therefore, many FMD methods [11]–[16] have been proposed aiming at reducing the MVC complexity. Peng *et al.* [11] proposed a hybrid FMD algorithm by using several dynamic RD thresholds and inter-view MB mode correlation estimated by Global Disparity Vectors (GDVs). In [12], Zhu *et al.* proposed a fast INTER mode decision based on textural segmentation and correlation. Ding *et al.* [13] proposed a content-aware inter-view mode decision algorithm by sharing RD cost, coding modes, and motion vectors among views. In

Manuscript received November 14, 2011; revised April 4, 2012; accepted March 8, 2013. Date of publication April 26, 2013; date of current version May 20, 2013. This work was supported in part by the Hong Kong RGC General Research Fund (GRF) Projects 9041495 (CityU 115109) and the Natural Science Foundation of China under Grants 61102088, 61272289, and 61071120. This paper was recommended by Associate Editor L. Zhang

Y. Zhang is with the Shenzhen Institutes of Advanced Technology (SIAT), Chinese Academy of Sciences (CAS), Shenzhen, China, and also with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (email:yunzhang@cityu.edu.hk).

S. Kwong and L. Xu are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail: cssamk@cityu.edu.hk, lxu@jdl.ac.cn).

G. Jiang is with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China (e-mail: jianggangyi@nbu.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/TBC.2013.2253033

[14], selective variable block-size ME and DE for MVC was presented based on motion homogeneity identified by motion vector deviation. Yang *et al*. [15] proposed FMD algorithm based on Coded Block Patterns (CBP), in which DIRECT or B16×16 were best mode candidates in case that CBP of coded MB modes equals to zero. In [16], hierarchical two-stage neural classifier was adopted for mode classification. In these schemes, however, the DIRECT/SKIP mode was not efficiently early terminated before checking other INTER modes.

Since most MB modes are DIRECT/SKIP mode in MVC, several FMD schemes [17]–[20] were proposed for DIRECT or SKIP mode early termination. Based on similar SKIP mode distribution among views, in [17], SKIP mode was directly selected when inter-view corresponding MB and its neighboring MBs were mostly SKIP modes. In [18], other FMD schemes were combined with this early SKIP mode decision to reduce complexity further. Similarly, in [19], SKIP mode was selected as a best mode while its RD cost was lower than a threshold, which was generated from weighted RD costs of temporal, spatial and view neighboring SKIP MBs. However, these algorithms cannot be applied to the first coded temporal view. In [20], Zhang *et al*. proposed a statistical approach for DIRECT mode early decision, in which an adaptive RD cost threshold was adaptively selected according to QP, video content and motion properties by controlling the RD degradation within pre-defined RD constraints. However, the inter-view correlation was not exploited. In addition, to maintain a low false acceptance rate, a number of DIRECT mode MBs were still not efficiently early terminated.

In this paper, we propose an Efficient DIRECT Mode Early Decision (EDMED) for DIRECT mode early decision. The rest of the paper is organized as, motivations and statistical analyses are presented in Section II. Section III presents four early termination sub-algorithms and the overall algorithm. Section IV shows the experimental results and analyses. Finally, conclusions are given in Section V.

II. Motivations and Analyses

In H.264/AVC based MVC standard, SKIP mode assumes that the current MB has the same motion vectors as median prediction and has no residual. DIRECT mode infers motion/disparity vector in spatial, temporal or inter-view prediction in MVC and may have residual. Since DIRECT is a wider concept for the hierarchical B frame in MVC, SKIP and DIRECT are denoted by DIRECT in the following sections.

To analyze the mode distribution of MVC, six multiview video test sequences with different motion properties are analyzed for their best mode distribution. Breakdancers and Ballroom are fast motion sequences, Ballet and Altmoabit are moderate motion sequences, Exit and Doorflowers are slow motion sequences. Table I shows the statistical probability of MB mode distribution for hierarchical B frames in MVC, where basis QP (denoted as *bQP* in the following sections) is 28. On one hand, we can observe that from 65.33% to 91.9% (81.86% on average) MBs selects DIRECT mode as the best MB modes for different sequences. Meanwhile, the percentage decreases as motion becomes faster or more complicate. On the other hand, from 79.25% to 97.24% (91.26% on average) MBs select large MB size blocks, i.e. DIRECT and B16×16, as the best mode due to high temporal correlation.

TABLE I
MODE DISTRIBUTION FOR MULTIVIEW VIDEO SEQUENCES [UNIT:%]

	Break.	Ballet	Ballroom	Exit	Doorflowers	Altmoabit	Avg.
DIRECT	65.33	87.56	70.06	88.95	91.90	87.37	81.86
B16×16	13.92	8.28	15.32	6.79	5.34	6.71	9.39
B8×16	2.40	1.24	3.76	1.03	0.82	1.52	8.75
B16×8	1.69	1.08	4.53	1.87	1.01	0.73	
B8×8	0.60	0.26	2.69	0.60	0.37	0.36	
INTRA	16.03	1.54	3.59	0.68	0.52	3.28	

In MVC mode decision, the MB modes are checked sequentially in the order of DIRECT, B16×16, B16×8, B8×16, B8×8 (including sub-MB modes) and INTRA modes. Since the ME/DE search times for each MB increase as MB partition becomes smaller, small block-size INTER MB modes usually consume more coding time than the large block-size INTER MB modes. Additionally, most MBs use DIRECT mode, in which no ME/DE is required and its complexity is negligible. Therefore, if we can early terminate the mode selection process by checking DIRECT mode only for those MBs selecting DIRECT mode as the best mode, significant computational complexity can be saved by skipping needless temporal and inter-view INTER prediction for small MB partitions.

III. The Proposed EDMED Algorithm

A. CBP Based DIRECT Mode Early Decision

The CBP is a syntax element in the H.264/AVC MB layer, which specifies six 8 × 8 blocks (for 4:2:0 sub-sampling) containing non-zero transform coefficient levels [15]. A CBP information will be generated after checking each MB mode, denoted as CBP_m, $m \in \{$DIRECT, B16×16, B8×16, B16×8, B8×8, I4MB, I8 MB, I16MB$\}$. When the six-bit CBP_m equals to zero, it is all zero MB for the current MB coded with mode m and the mode m is likely to be the best mode due to precise prediction. Therefore, if CBP of DIRECT mode (CBP_D) equals to zero, the current MB is AZB and DIRECT mode will probability be the best mode.

To testify the hit probability of DIRECT mode early decision for the early termination condition, Absolute False Acceptance Rate (AFAR) and Absolute False Rejection Rate (AFRR) are adopted. Let **A** be the event that DIRECT mode is set as the final best mode of one MB in the MVC mode decision, and **B** be the event of choosing non-DIRECT mode as the best mode. Let $N(\mathbf{B} \Rightarrow \mathbf{A}|\Phi)$ be the number of non-DIRECT mode MBs which are falsely classified as DIRECT mode with the early termination condition Φ, where symbol "$\Rightarrow$" indicates one event transfer to another. Similarly, let $N(\mathbf{A} \Rightarrow \mathbf{B}|\Phi)$ be the number of DIRECT MBs which are falsely classified as non-DIRECT with the early termination condition Φ. Thus, the AFAR and AFRR of **A** under condition Φ, $P_{AFAR}(\mathbf{A}|\Phi)$ and $Q_{AFRR}(\mathbf{A}|\Phi)$, can be calculated as

$$\begin{cases} P_{AFAR}(\mathbf{A}\,|\,\mathbf{\Phi}) = N(\mathbf{B} \Rightarrow \mathbf{A}\,|\,\mathbf{\Phi})\big/N \\ Q_{AFRR}(\mathbf{A}\,|\,\mathbf{\Phi}) = N(\mathbf{A} \Rightarrow \mathbf{B}\,|\,\Phi)\big/N \end{cases}, \tag{1}$$

where N is the total number of MBs in one frame.

TABLE II

AVERAGE P_{AFAR} AND Q_{AFRR} ANALYSES FOR DIRECT MODE EARLY DECISION CONDITIONS [%]

Views	Multiview video	CBP		GDV		SDMET		SDMET+ GDV+CBP		Proposed EDMED	
		P_{AFAR}	Q_{AFRR}	P_{AFAR}	Q_{AFRR}	P_{AFAR}	Q_{AFRR}	P_{AFAR}	Q_{AFRR}	P_{AFAR}	Q_{AFRR}
Even	Ballroom	4.81	2.37	/	/	0.77	32.12	0.72	32.15	1.22	1.32
	Ballet	3.92	0.4	/	/	0.39	25.66	0.38	25.68	0.38	0.74
	Breakdancers	8.02	3.37	/	/	1.31	23.16	1.14	23.48	1.58	0.96
	Doorflowers	2.47	0.13	/	/	0.6	35.36	0.59	35.36	0.59	0.69
Even views Avg.		**4.81**	**1.57**	**/**	**/**	**0.77**	**29.08**	**0.71**	**29.17**	**0.94**	**0.94**
Odd	Ballroom	3.29	1.76	0.76	34.82	0.53	26.75	0.65	15.65	1.04	1.30
	Ballet	3.45	0.3	0.31	33.79	0.23	26.30	0.66	12.60	0.66	0.76
	Breakdancers	8.13	2.17	0.44	41.16	0.78	22.87	0.92	17.47	1.00	1.49
	Doorflowers	2.01	0.06	0.44	27.28	0.5	35.24	0.72	11.74	0.72	0.71
Odd views Avg.		**4.22**	**1.07**	**0.49**	**34.26**	**0.51**	**27.79**	**0.74**	**14.36**	**0.86**	**1.06**
Total Avg.		**4.51**	**1.32**	**0.49**	**34.26**	**0.64**	**28.43**	**0.72**	**21.77**	**0.90**	**1.00**

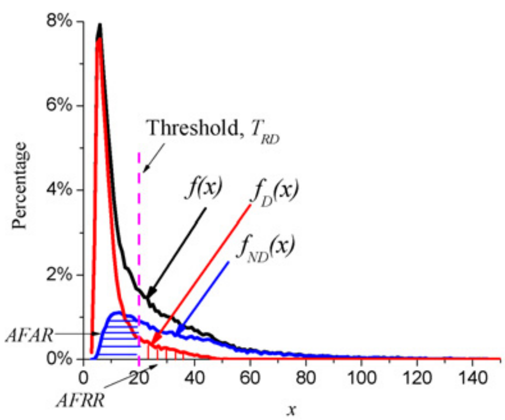

Fig. 1. Graphical explanation of probability functions (Statistical PDF for Breakdancers, bQP is 28).

Table II shows the AFAR and AFRR for four different test video sequences and two kinds of views (temporal and inter-view views). Coding conditions are: JMVC8.0, fast ME/DE is enabled and their search ranges are ±96, GOP length is 12, other parameters are default settings. These values in Table II are the average AFAR and AFRR of different bQPs, which are 24, 28, 32 and 36. The first two columns, labeled as "CBP", show the AFAR and AFRR for DIRECT mode early decision while CBP_D equals to zero, i.e. $P_{AFAR}(\mathbf{A}|CBP_D{=}0)$ and $Q_{AFRR}(\mathbf{A}|CBP_D{=}0)$. The Q_{AFRR} value is very small, 0.13%, which means almost all CBP_D of DIRECT mode MBs are zero and mode decision process for DIRECT mode MB can be efficiently early terminated. However, the P_{AFAR} is large, i.e. 4.51%, which indicates that 4.51% other INTER/INTRA modes will be selected as best mode and large RD degradation could be caused in case we use CBP_D=0 as a early termination condition. Based on this observation, it is reasonable and required to combine this CBP condition with other conditions to reduce the P_{AFAR}.

B. Statistical DIRECT Mode Early Termination (SDMET)

Let us denote the average square root of the RD cost of encoding a MB with DIRECT mode by a random variable **X.** Let $f(x)$ be the Probability Density Function (PDF) of **X** for the MBs of a frame. Let $f_D(x)$ be the probability function of **X** for those MBs who select DIRECT mode as the best mode and $f_{\mathrm{ND}}(x)$ be the probability function of **X** for those MBs who select non-DIRECT mode as the best mode, $f_{\mathrm{ND}}(x)$ equals to 1- $f_D(x)$. Fig. 1 shows statistical PDFs for Breakdancers when bQP is 28, where x-axis is **X** while each MB is encoded with DIRECT mode, y-axis is the probability. According to the figure, $f(x)$ and $f_D(x)$ are approximately Laplacian source shape when $x > 0$, thus, they are formulated as

$$f(x) = \begin{cases} \frac{1}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\} & x > \mu \\ 0 & x \le \mu \end{cases}, \tag{2}$$

$$f_D(x) = \begin{cases} P_D \frac{1}{\sigma_D}\exp\left\{-\frac{x-\mu_D}{\sigma_D}\right\} & x > x_0 \\ f(x) & x \le x_0 \end{cases}, \tag{3}$$

where $\sigma_D < \sigma, \mu_D \approx \mu$, $x_0 = \mu + \frac{\sigma\sigma_D}{\sigma-\sigma_D}\ln\frac{P_D\sigma}{\sigma_D}$, P_D is percentage of DIRECT mode MB. Thus, AFRR and AFAR values, P_{AFAR} and Q_{AFRR}, can be calculated based on the Cumulative Density Function (CDF) of $f_D(x)$. They are

$$\begin{array}{c} P_{AFAR}(T_{RD}) = \int_0^{T_{RD}} 1 - f_D(x)\,dx \\ Q_{AFRR}(T_{RD}) = \int_{T_{RD}}^{\infty} f_D(x)\,dx \end{array}, \tag{4}$$

where T_{RD} is the RD cost threshold for DIRECT mode early decision. Here, we use differential Mean Square Error (MSE), denoted by ΔMSE, to evaluate the RD degradation caused by early termination. Here, the ΔMSE can be calculated as

$$\begin{array}{l} \Delta MSE = \int_0^{T_{RD}} \left(x^2 - x_{ND}^2\right) f_{ND}(x)\,dx \\ \le \int_0^{T_{RD}} x^2 f_{ND}(x)\,dx = \Delta MSE_{Up}(T_{RD}) \end{array}, \tag{5}$$

where x and x_{ND} denote the cost to encoding an MB with DIRECT and non-DIRECT mode, respectively.$\Delta MSE_{Up}(T_{RD})$ is the upper bound of ΔMSE with threshold T_{RD}.

As the optimization target is to minimize AFRR with acceptable AFAR and ΔMSE, we obtain the optimal T_{RD} by

$$T_{RD}^{*} = \arg\min_{T_{RD}} \left(Q_{AFRR}(T_{RD}) + \phi \cdot P_{AFAR}(T_{RD})\right), \tag{6}$$

$$\Delta MSE_{Up}\left(T_{RD}^{*}\right) \le T_{\Delta MSE}, \tag{7}$$

where ϕ is a weighted coefficient indicating the trade-off between P_{AFAR} and Q_{AFRR}, $\phi > 0$, larger ϕ indicates stricter limit on RD degradation. Meanwhile, Eq. (7) is used as verification while Eq. (6) is solved. Here, ϕ is initialized to be a

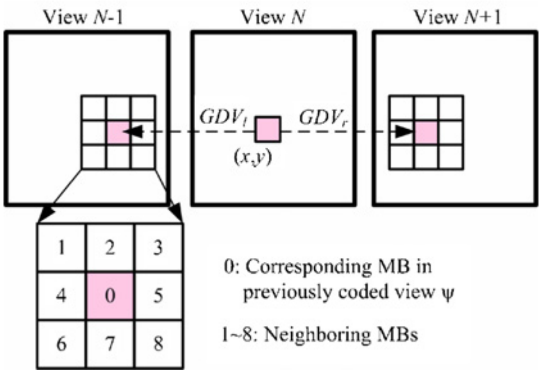

Fig. 2. DIRECT mode decision using GDV.

constant and automatically adjusted and updated till Eq. (7) is satisfied. For more details of the implementation steps, readers can refer to [20]. Here, a stricter threshold $T_{\Delta PSNR}$, which indicates the allowance for the RD degradation in SDMET, is adopted to maintain high compression efficiency and it is 0.05.

The fifth and sixth columns of Table II, labeled as "SDMET", show the average $P_{AFAR}(A|J_D < T_{RD})$ and $Q_{AFRR}(A|J_D < T_{RD})$ for early DIRECT mode decision for different video sequences. The SDMET scheme can be applied to both odd and even views. We can observe that the P_{AFAR} is smaller than 0.77% which ensures small RD degradation. However, the Q_{AFRR}s are 29.08% and 27.79% for temporal (even) and inter-view (odd) views, respectively, which means almost 30% DIRECT mode MBs are not efficiently early terminated and there is a large optimizing space for complexity reduction.

C. DIRECT Mode Decision Using GDV

Since multiview video data is captured from multiple cameras from slightly different viewpoints or angles simultaneously, in addition to spatial and temporal correlation, multiview video also has a high inter-view correlation among views [17]. Thus, similar MB partitions will probably be adopted for the inter-view corresponding MB and it is of high inter-view mode correlation. Fig. 2 shows the illustration of DIRECT mode early decision by exploiting mode correlation using GDV. As for the inter-view views, two corresponding MBs could be found from neighboring views. A weighted factor of DIRECT mode early decision $W_D(x,y)$ is defined as

$$W_D(x, y) = \min\begin{pmatrix} W_l(x + GDVx_l,\, y + GDVy_l), \\ W_r(x + GDVx_r,\, y + GDVy_r) \end{pmatrix}, \quad (8)$$

where x, y are horizontal and vertical coordinates of a MB, respectively, $GDV_\psi = \{GDVx_\psi, GDVy_\psi\}$, $\psi \in \{l, r\}$, which is left view or right view. For the situations that only one GDV is available, the W_ψ without GDV is set as the maximum value. $W_\psi(u,v)$ is defined as

$$W_\psi(u, v) = \begin{cases} \sum_{i=0}^{8} \alpha_i K_{\psi,i}(u, v) & u \in [1, W-2],\, v \in [1, H-2] \\ K_{\psi,0}(u, v) \sum_{i=0}^{8} \alpha_i & u \in \{0, W-1\},\ v \in \{0, H-1\} \\ 0 & \text{others} \end{cases}, \quad (9)$$

where α_i is constants for DIRECT mode for corresponding MBs in Fig. 2, $i \in [0,8]$. Here, α_0 is 2.0, $\alpha_1 \alpha_3 \alpha_6 \alpha_8$ are 0.25, the rest of α_i s are 1.0 [17]. W and H are image width and height measured with MB unit. The coefficient $K_{\psi,i}(u, v)$ is

$$K_{\psi,i}(u, v) = \begin{cases} 1 & if\ M(u, v, i) = DIRECT \\ 0 & else \end{cases}, \quad (10)$$

where $M(u, v, i)$ is the MB mode at (u, v) location and its surrounding MBs (0-8 in Fig. 2), $i \in [0,8]$. Here, a threshold T_W is defined for $W_D(x, y)$. When the weight factor $W_D(x, y)$ is larger than T_W, the current MB is early determined as DIRECT mode. The threshold T_W is set as 6.25 according to analyses in [17].

The third and fourth columns of Table II, labeled as "GDV", shows $P_{AFAR}(A|W_D(x, y) < T_W)$ and $Q_{AFRR}(A|W_D(x, y) < T_W)$ for early DIRECT termination. Distance between temporal views is larger and the correlation between temporal views is relative low, only the correlations between interviews and temporal views are considered and exploited. We can see that the P_{AFAR} value is sufficient low to guarantee high compression efficiency. However, its Q_{AFRR} is 34.26% which means about half of DIRECT mode MBs have not been efficiently terminated. In other words, the complexity reduction is limited. In subsections 3.1 to 3.3, we attempt to early terminate DIRECT mode decision without checking INTRA modes and other time-consuming INTER modes. The seventh and eighth columns of Table II show AFAR and AFRR of a combination of aforementioned three early termination sub-algorithms, denoted by "CBP+GDV+SDMET". The Q_{AFRR}s are reduced to 29.17% and 14.36% for even and odd views, respectively. However, a large number of DIRECT mode MBs are still falsely rejected and need to further check other INTER mode, i.e. large Q_{AFRR}, while maintaining the P_{AFAR} sufficient low. Therefore, we introduce a RD cost comparison based DIRECT mode early decision method to further reduce the Q_{AFRR} in the next subsection, where B16×16 mode is extensively checked.

D. Early DIRECT Mode Determination Based on Weighted RD Cost Comparison

In H.264/AVC based MVC, the best MB mode is selected based on minimizing the Lagrangian RD cost which is calculated as

$$m^* = \min_{m \in \mathbf{M}} J_m, \quad (11)$$

$$J_m = SSD(m) + \lambda_{MODE} R(m), \quad (12)$$

where $SSD()$ is a sum of the squared differences between a original block and its reconstruction, $R()$ is the encoding bits for the encoded block, $\mathbf{M}$ is a complete set of all mode

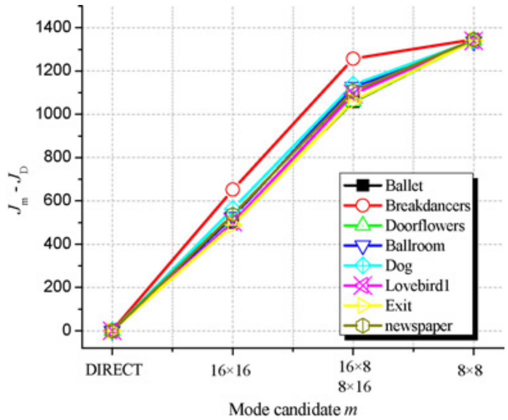

Fig. 3. Average RD cost difference between DIRECT mode and other INTER modes for the MBs that DIRECT mode is set as the best mode. (*bQP* is 28).

candidates, λ_{MODE} is Lagrangian multiplier for mode decision. Let $\mathbf{M}_{nD}$ be mode set excluding DIRECT mode, i.e. $\mathbf{M}_{nD}$=**M**-{DIRECT}. If RD cost of DIRECT mode, J_D, is smaller than the RD cost of all other MB mode $m \in \mathbf{M}_{nD}$, the best MB mode m^* will be DIRECT mode, where event **A** happens. This mode decision process can be represented by conditional probability as

$$P(\mathbf{A}|J_D < J_m, m \in \mathbf{M}_{nD}) = 100\%, \tag{13}$$

where $P(\cdot|\cdot)$ indicates conditional probability. That means its AFRR and AFAR under condition $J_D < J_m$ both equal to zero. Although $J_D < J_m$ is a sufficient condition for DIRECT mode determination, all MB mode candidates shall be checked.

To reduce the mode decision complexity, the mode candidates shall be checked as few as possible while determining the best mode. To this end, we analyze the RD cost properties for the MBs that select DIRECT mode as best mode after full mode search, where *bQP* is 28, as shown in Fig. 3. In the figure, *y*-axis is RD cost difference for each MB which is coded with mode m and DIRECT mode, *x*-axis is mode candidate m. For the MBs that select DIRECT as the best mode, the average RD cost difference generally linearly increases as the block partition getting smaller. In other words, if this increasing trend is always satisfied, we can early terminate mode decision for DIRECT mode MBs by only comparing the RD cost of DIRECT and B16×16 modes. Similar increasing trends can be found when *bQP* is 24, 32 or 36. The miss rate for the early mode decision is calculated as

$$\varepsilon = P(\mathbf{A}|J_D < J_m, m \in \mathbf{M}_{nD}) - P(\mathbf{A}|J_D < \lambda \times J_{B16\times16}), \tag{14}$$

where λ is positive coefficient, ε tends to be zero when λ is small. In fact, ε is direct proportional to False Acceptance Rate (FAR). In addition, False Rejection Rate (FRR) is also analyzed for the condition $J_D < \lambda \times J_{B16\times16}$. The FAR and FRR are defined as

$$p_{FAR}(\mathbf{A}|J_D < \lambda \times J_{B16\times16}) = \frac{N(\mathbf{B} \Rightarrow \mathbf{A}|J_D < \lambda \times J_{B16\times16})}{N(\mathbf{B})}, \tag{15}$$

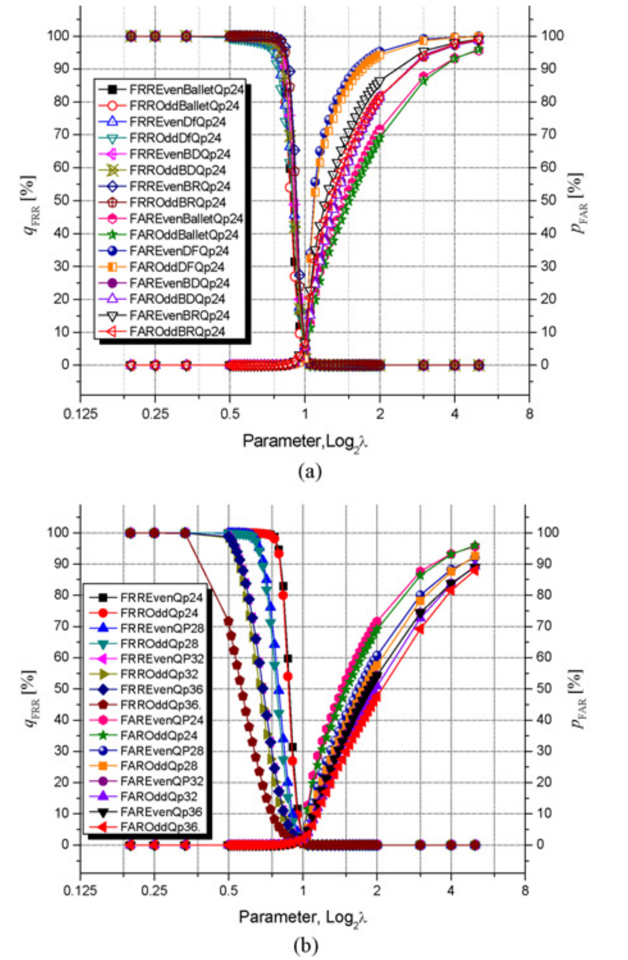

Fig. 4. FRR and FAR for cost comparison based on early DIRECT mode decision. (a) p_{FAR} and q_{FRR} for different multiview video sequences (*bQP* is 24). (b) p_{FAR} and q_{FRR} for different *bQP*s (Ballet).

$$q_{FRR}(\mathbf{A}|J_D < \lambda \times J_{B16\times16}) = \frac{N(\mathbf{A} \Rightarrow \mathbf{B}|J_D \geq \lambda \times J_{B16\times16})}{N(\mathbf{A})}, \tag{16}$$

where $N(\mathbf{A})$ and $N(\mathbf{B})$ represent the number of DIRECT mode MBs and non-DIRECT mode MBs, respectively.

To analyze the accuracy and performance of the sufficient condition assumption, we analyze the q_{FRR} and p_{FAR} with different λs, where we assign more λ test points nearby 1.0, i.e. $\mu \in \left\{\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}\right\} \cup \left\{\frac{1}{2-0.05\times k}\right\} \cup \{1\} \cup \{1 + 0.05 \times k\} \cup \{2, 3, 4, 5\}$, $k \in [1, 19]$. Four different multiview video sequences (Ballet, Breakdancers, Doorflowers and Ballroom), odd and even views, and four different *bQP*s are used in the test experiments. Fig. 4 shows the q_{FRR} and p_{FAR} for different λs. Fig. 4(a) shows q_{FRR} and p_{FAR} for different nultiview video sequences, where the left *y*-axis is the q_{FRR}, the right *y*-axis is the p_{FAR}, and the *x*-axis is λ in log scale. Here, we use q_{FRR} and p_{FAR} instead of previous mentioned AFRR and AFAR for better observation while drawing them in one figure with one scale. From the statistical analyses, we have the following three observations: 1) Firstly, the q_{FRR} is linearly decreasing from 100% to 0% when λ is in the range of 0.75 to 1.0. Also, it approximately equals to zero when λ becomes larger than 1.0. The p_{FAR} increases along with λ and it increases to 100% in exponential trend as λ becomes larger than 1.0. 2) Secondly, the q_{FRR} decreasing trends are almost the same for different video sequences, meanwhile, the increasing slopes of p_{FAR} increase as the motion of video sequences become slower or simpler. 3) Thirdly, similar results can be found for Ballet sequence with different *bQP*s, as shown in Fig. 4(b). In addition, the slopes of q_{FRR} and p_{FAR} curves decrease as *bQP* increases. For one sequence and fixed *bQP*, the RD performance is direct proportional to p_{FAR}, and the complexity reduction rate decreases as the q_{FRR} increases. Thus, to maximize speed up ratio while maintaining RD performance, both q_{FRR} and p_{FAR} shall be small when selecting optimal λ. According to the observations of Fig. 4, p_{FAR} and q_{FRR} are sufficiently low when λ equals to 1.0 for different sequences and *bQP*s. Thus, this value is chosen as the optimal λ.

E. Proposed Overall EDMED Algorithm

According to the aforementioned analysis, the flowchart of the proposed overall EDMED algorithm is shown in Fig. 5 and its detailed implementing steps are described as follow:

1) If current frame is an anchor frame, perform exhaustive full mode decision and select best mode via RD cost comparison, go to Step 6;
Otherwise, encode MB with DIRECT mode and go to Step 2.
2) If CBP_D is non-zero, go to Step 5;
else if current frame is inter-view frame, go to Step 3;
else if current frame is temporal frame, go to Step 4.
3) If conditions $J_D < T_{RD}$ or $W_D > T_W$ are satisfied, the MB mode of current MB is early determined as DIRECT mode directly and go to Step 6; Otherwise, go to Step 5.
4) If conditions $J_D < T_{RD}$ satisfies, the current MB is set as DIRECT and go to Step 6; Otherwise, go to Step 5.
5) Perform 16×16 ME/DE, compare RD cost of DIRECT and the weighted RD cost of B16×16.
If $J_D < \lambda \times J_{B16\times16}$ satisfies, select the best mode between B16×16 and DIRECT based on the minimum RD cost;
else perform other INTER and INTRA modes and select the best mode via RD cost comparison.
6) Go to Step 1 for Next MB.

The last two columns of Table II show P_{AFAR} and Q_{AFRR} of the proposed overall EDMED algorithm, the Q_{AFRR} decreases from 21.77% to 1.00% by adding the weighted RD cost comparison algorithm. It means only 1.00% out of 81.86% in total DIRECT mode MBs have not been early terminated. As 81.86% MBs will select DIRECT mode as the best mode, there are 60.09% out of 81.86% (i.e. 81.86%-21.77%) MBs will check DIRECT mode only, 20.77% MBs will check DIRECT and B16×16 modes, and the rest 1.00% MBs need check all mode candidates in oder to select the best mode. As for the P_{AFAR}, the proposed EDMED is 0.94% which means 0.94% MBs will be falsely select DIRECT mode as the best mode. The value is a little larger than that of GDV and SDMET

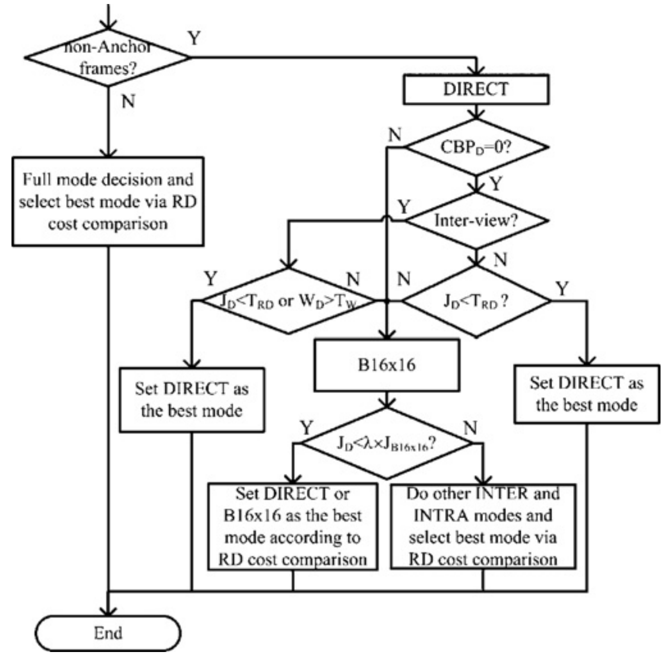

Fig. 5. Flowchart of the proposed overall EDMED.

schemes. However, the CBP_D=0 constraints might make RD degradation for each false mode selection smaller, since other INTER or INTRA mode MBs which falsely chooses DIRECT as best shall satisfy CBP_{D}=0 and guarantee good prediction.

IV. Experimental Results and Analyses

The recent H.264/AVC based MVC reference software JMVC 8.0 is utilized to evaluate the proposed fast algorithms. Fast ME/DE is enabled and their search ranges are ±96. The number of bi-prediction iteration is 4 and search range for iteration is 8. The maximum number of reference frames is 2 and GOP length is 12. Eight multiview video test sequences, including Race1, Ballroom, Exit, Lovebird1, Doorflowers, Breakdancers, Ballet and Dog, with various motion properties and camera arrangement are adopted. Eight views are encoded and four *bQP* values, 24, 28, 32 and 36, are used in our experiments. The coding parameters are also the same for the original JMVC, Shen's scheme (denoted by ShenSPIC) [17], SDMET [20] and proposed EDMED. ShenSPIC [17], SDMET [20] and EMDED are all also proposed for early DIRECT mode termination, so it is fair for us to make the comparison.

Table III shows the encoding time, Peak Signal-to-Noise Ratio (PSNR), bit rate comparison among original JMVC, ShenSPIC [17], SDMET and proposed EDMED, where the time saving ratio, PSNR difference and bit rate increment between the original JMVC encoder and compared algorithms are computed as

$$\begin{cases} \Delta T_\theta = (T_{JMVC} - T_\theta)\big/T_{JMVC} \times 100[\%] \\ \Delta PSNR_\theta = PSNR_\theta - PSNR_{JMVC} \\ \Delta R_\theta = (R_\theta - R_{JMVC})\big/R_{JMVC} \times 100[\%] \end{cases}, \quad (17)$$

where T_θ, $PSNR_\theta$ and R_θ are total encoding time, PSNR and bit rate of scheme θ, respectively, $\theta \in$ {ShenSPIC, SDMET,

TABLE III

BIT RATE, PSNR AND ENCODING TIME COMPARISONS AMONG ORIGINAL JMVC, SHENSPIC, SDMET AND THE PROPOSED EDMED

Scheme			$\Delta R_{ShenSPIC}$(Unit:%)/$\Delta PSNR_{ShenSPIC}$(Unit:dB)/ $\Delta T_{ShenSPIC}$(Unit:%)			
ShenSPIC scheme[17] VS Original JMVC		*bQP*	Ballet	Breakdancers	Doorflowers	Lovebird1
	Even Views	24,28,32,36	0.0/0.0/0.0			
	Avg.					
	BDBR(%)/BDPSNR(dB)					
	Odd Views	24	0.86/−0.04/39.39	−0.26/−0.01/11.81	0.53/−0.05/45.89	−0.44/−0.05/47.59
		28	1.49/−0.07/47.62	−0.30/−0.03/18.02	2.55/−0.05/48.31	−0.42/−0.03/62.47
		32	1.85/−0.10/52.97	−0.66/−0.04/22.18	4.53/−0.06/54.34	−0.40/−0.02/66.99
		36	2.32/−0.14/57.13	−0.94/−0.08/27.42	4.77/−0.07/58.53	−0.35/−0.02/69.99
	Avg.		1.63/−0.09/49.28	−0.54/−0.04/19.86	3.10/−0.06/51.77	−0.40/−0.03/61.76
	BDBR(%)/BDPSNR(dB)		4.46/−0.14	1.06/−0.03	5.13/−0.15	0.38/−0.01
		bQP	Dog	Ballroom	Exit	Race1
	Even Views	24,28,32,36	0.0/0.0/0.0			
	Avg.					
	BDBR(%)/BDPSNR(dB)					
	Odd Views	24	−0.90/−0.04/39.76	−0.18/−0.02/22.61	−0.40/−0.03/22.20	−0.06/−0.06/13.52
		28	−0.76/−0.03/48.45	−0.02/−0.01/33.50	0.50/−0.04/37.67	0.43/−0.11/18.42
		32	−0.81/−0.03/52.75	0.20/−0.02/38.61	1.24/−0.05/42.67	1.48/−0.17/22.36
		36	−0.90/−0.04/55.77	1.37/−0.04/24.84	1.70/−0.10/46.39	3.73/−0.31/24.68
	Avg.		−0.84/−0.04/49.18	0.34/−0.02/29.89	0.76/−0.05/37.23	1.39/−0.16/19.75
	BDBR(%)/BDPSNR(dB)		0.17/0.00	0.71/−0.03	2.72/−0.08	4.49/−0.21
SDMET Scheme[20] VS Original JMVC			ΔR_{SDMET}(Unit:%)/$\Delta PSNR_{SDMET}$(Unit:dB)/ΔT_{SDMET}(Unit:%)			
		bQP	Ballet	Breakdancers	Doorflowers	Lovebird1
	Even Views	24	−1.37/−0.06/49.34	−0.38/−0.09/28.02	−1.41/−0.13/60.14	−0.04/0.00/27.76
		28	−0.95/−0.05/53.17	−1.00/−0.09/32.29	−1.37/−0.11/65.63	−0.29/−0.02/51.65
		32	−0.59/−0.04/56.39	−0.97/−0.08/36.39	−0.97/−0.08/66.07	−0.42/−0.03/65.32
		36	−0.28/−0.02/59.18	−0.52/−0.05/40.97	−0.49/−0.04/65.46	−0.28/−0.02/67.57
	Avg.		−0.79/−0.04/54.52	−0.72/−0.08/34.42	−1.06/−0.09/64.32	−0.26/−0.02/53.08
	BDBR(%)/BDPSNR(dB)		0.95/−0.02	2.93/−0.06	2.20/−0.06	0.27/−0.01
	Odd Views	24	−1.13/−0.07/67.00	−0.50/−0.08/39.09	−2.12/−0.16/67.52	0.00/−0.03/39.34
		28	−0.73/−0.04/66.55	−0.81/−0.07/41.28	−1.20/−0.10/67.59	−0.22/−0.02/50.14
		32	−0.27/−0.03/64.99	−0.75/−0.06/43.34	−0.67/−0.07/65.16	−0.25/−0.02/66.26
		36	−0.53/−0.02/63.87	−0.59/−0.04/45.92	−0.45/−0.05/62.23	−0.06/−0.01/64.54
	Avg.		−0.66/−0.04/65.60	−0.66/−0.06/42.41	−1.11/−0.09/65.62	−0.13/−0.02/55.07
	BDBR(%)/BDPSNR(dB)		0.97/−0.02	2.14/−0.05	2.55/−0.07	0.35/−0.01
		bQP	Dog	Ballroom	Exit	Race1
	Even Views	24	−0.44/−0.04/33.41	−0.66/−0.07/29.92	−1.09/−0.06/37.64	−0.62/−0.11/38.79
		28	−0.66/−0.05/47.98	−0.42/−0.04/33.15	−0.53/−0.03/43.05	−0.81/−0.10/39.45
		32	−1.08/−0.07/58.85	−0.26/−0.03/36.23	−0.19/−0.02/46.11	−0.89/−0.09/39.86
		36	−0.65/−0.05/60.55	−0.24/−0.04/37.16	−0.11/−0.01/49.43	−0.76/−0.07/39.78
	Avg.		−0.71/−0.05/50.20	−0.40/−0.04/34.12	−0.48/−0.03/44.06	−0.77/−0.09/39.47
	BDBR(%)/BDPSNR(dB)		0.74/−0.03	0.68/−0.03	0.82/−0.02	1.34/−0.06
	Odd Views	24	−0.26/−0.03/40.76	−0.35/−0.05/50.37	−0.82/−0.07/54.56	0.15/−0.15/44.14
		28	−0.34/−0.04/56.83	−0.21/−0.03/52.38	−0.46/−0.04/53.11	−1.36/−0.12/43.45
		32	−0.42/−0.05/66.56	−0.01/−0.02/52.95	−0.12/−0.02/60.21	−0.69/−0.09/41.34
		36	−0.27/−0.04/66.20	−0.11/−0.04/50.82	−0.05/−0.01/60.88	−0.15/−0.09/40.69
	Avg.		−0.32/−0.04/57.59	−0.17/−0.04/51.63	−0.36/−0.03/57.19	−0.51/−0.11/42.40
	BDBR(%)/BDPSNR(dB)		0.86/−0.03	0.78/−0.03	0.97/−0.02	1.82/−0.08
Proposed EDMED Scheme VS Original JMVC			ΔR_{EDMED}(Unit:%)/$\Delta PSNR_{EDMED}$(Unit:dB)/ΔT_{EDMED}(Unit:%)			
		bQP	Ballet	Breakdancers	Doorflowers	Lovebird1
	Even Views	24	−0.31/−0.03/51.05	−0.62/−0.06/31.17	−0.37/−0.06/65.98	0.10/−0.01/63.92
		28	−0.27/−0.03/56.34	−0.48/−0.07/37.67	−0.31/−0.06/70.39	−0.02/−0.01/70.91
		32	−0.03/−0.03/61.21	−0.31/−0.07/43.29	−0.21/−0.05/73.31	−0.15/−0.02/76.26
		36	−0.10/−0.03/66.03	−0.23/−0.07/49.70	−0.04/−0.04/75.88	−0.04/−0.02/79.90
	Avg.		−0.18/−0.03/58.66	−0.41/−0.07/40.46	−0.23/−0.05/71.39	−0.03/−0.01/72.75
	BDBR(%)/BDPSNR(dB)		0.82/−0.02	2.33/−0.05	1.51/−0.05	0.36/−0.01
	Odd Views	24	−1.05/−0.04/71.28	−0.82/−0.06/43.82	−1.94/−0.10/79.87	−0.25/−0.06/75.48
		28	−1.13/−0.06/74.16	−0.91/−0.07/49.12	−1.20/−0.09/80.33	−0.41/−0.03/82.10
		32	−1.30/−0.08/75.97	−1.13/−0.09/53.28	−1.07/−0.08/81.60	−0.42/−0.03/83.25
		36	−1.87/−0.11/77.50	−1.81/−0.11/57.69	−0.63/−0.09/82.21	−0.40/−0.02/83.71
	Avg.		−1.34/−0.08/74.73	−1.17/−0.08/50.98	−1.21/−0.09/81.00	−0.37/−0.04/81.13
	BDBR(%)/BDPSNR(dB)		1.11/−0.03	2.71/−0.06	2.00/−0.06	0.63/−0.02

	bQP	Dog	Ballroom	Exit	Race1
Even Views	24	−0.31/−0.03/53.28	0.30/−0.04/38.02	0.04/−0.04/45.53	−0.01/−0.07/54.46
	28	−0.20/−0.03/58.78	0.34/−0.03/46.36	0.26/−0.03/52.64	0.24/−0.07/58.64
	32	−0.24/−0.04/63.48	0.22/−0.04/43.37	0.38/−0.03/57.76	0.40/−0.09/60.82
	36	−0.27/−0.04/66.77	0.48/−0.06/45.00	0.07/−0.04/62.06	0.41/−0.09/62.40
Avg.		−0.25/−0.03/60.58	0.34/−0.04/43.19	0.19/−0.03/54.50	0.26/−0.08/59.08
BDBR(%)/BDPSNR(dB)		0.74/−0.03	1.27/−0.05	1.47/−0.04	2.18/ −0.09
Odd Views	24	−1.19/−0.06/66.81	0.33/−0.05/59.25	−0.49/−0.05/64.09	−0.52/−0.12/56.89
	28	−1.05/−0.06/69.87	0.14/−0.04/66.95	−0.24/−0.05/70.86	−0.69/−0.15/60.82
	32	−1.01/−0.06/71.04	0.48/−0.04/64.46	0.14/−0.06/73.27	−0.86/−0.21/63.65
	36	−0.91/−0.06/71.83	0.41/−0.07/61.78	−0.21/−0.10/75.35	−0.37/−0.34/64.61
Avg.		−1.04/−0.06/69.89	0.34/−0.05/63.11	−0.20/−0.07/70.89	−0.61/−0.20/61.49
BDBR(%)/BDPSNR(dB)		0.81/−0.03	1.57/−0.06	2.06/−0.06	3.56/−0.16

EDMED}, T_{JMVC}, $PSNR_{JMVC}$ and R_{JMVC} are total encoding time, PSNR and bit rate of the original JMVC, respectively. Also, Bjonteggard Delta PSNR (BDPSNR) and Bjonteggard Delta Bit Rate (BDBR) [21] are used to measure the average PSNR and bit rate differences between RD curves.

From Table III, we can see that ShenSPIC reduces by 19.75% to 61.76% (39.84% on average) computational complexity for the inter-view (odd) views. Meanwhile, the average PSNR degrades 0.03 dB to 0.16 dB and bit rate increases from -0.84% to 3.10%. The RD degrades mainly due to inaccurate mode mapping from MB unit GDV. Also, the assumption of video objects having the same displacement among views is usually incorrect. For fast motion video sequences, such as Breakdancers and Ballroom, ShenSPIC can only achieve 19.86% and 29.89% complexity reduction because of sparse DIRECT mode distribution and high rejection rate caused by inaccurate mode mapping from GDV. In addition, ShenSPIC scheme cannot be applied to some temporal views without GDV and no complexity reduction is achieved for these views. The BDPSNR between ShenSPIC and the original JMVC is 0.0 to -0.21 dB (−0.08 dB on average) for the odd views. As for the SDMET scheme, PSNR degrades 0.02 dB to 0.11 dB (0.05 dB on average) for the test sequences and bit rate decreases from 0.13% to 1.11%. The BDPSNR between SDMET and JMVC ranges from −0.01 to −0.08 dB (−0.04 dB on average) on both odd and even views. Meanwhile, the SDMET can reduce complexity from 34.12% to 64.32% (46.78% on average) for even views and from 42.40% to 65.62% (54.69% on average) for odd views.

As for the proposed EDMED scheme, the PSNR degrades from 0.01 dB to 0.08 dB (0.04 dB on average) for even views and from 0.04 dB to 0.08 dB (0.07 dB on average) for most sequences of odd views. Meanwhile, bit rate increases from −1.34% to 0.34 % (-0.37% on average) which means 0.37% bit reduction for test sequences. As for the BDPSNR between the proposed EDMED and JMVC, we can see that it is from −0.01 dB to −0.09 dB (−0.04 dB on average) for even views and from −0.02 dB to −0.16 dB (−0.06 dB on average) for odd views, which is a little bit superior to the ShenSPIC scheme and a little bit inferior to the SDMET scheme. However, in terms of the complexity, the proposed EDMED can reduce total encoding time by 40.46% to 72.75% (57.57% on average) for even views, which reduces 11.76% more complexity than SDMET. Also, EDMED reduces 50.98% to 81.13% (69.15% on average) complexity for odd views, which is 29.31% and 15.03% more than the average complexity reduction achieved by ShenSPIC and SDMET schemes, respectively. The proposed EDMED achieves significant complexity reduction while maintaining sufficient high coding performance. In addition, it is DIRECT mode early decision and can be easily combined with other FMD algorithms for further coding complexity reduction.

V. Conclusions

An Efficient DIRECT Mode Early Decision (EDMED) algorithm is proposed to reduce MVC complexity. Firstly, DIRECT mode is directly determined through statistical RD cost threshold, inter-view mode correlation and CBP before checking time-consuming ME/DE. To maintain sufficient low AFAR, large AFRRs (29.17% for even views and 15.03% for odd views) are caused in the first phase optimization. Thus, a weighted RD cost comparison between DIRECT and B16×16 is presented to reduce the AFRRs to 1.00%. Experimental result proves that the proposed EDMED is efficient.

References

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

[2] 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.

[3] 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," in *Proc. IEEE*, vol. 99, no. 4, Apr. 2011, pp. 626–642.

[4] Z. Deng, Y. L. Chan, K. Jia, C. Fu, and W. 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.

[5] Y. Zhang, S. Kwong, G. Jiang, and H. Wang, "Efficient multi-reference frame selection algorithm for hierarchical B frames in multiview video coding," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 15–24, Mar. 2011.

[6] T. Zhao, H. Wang, S. Kwong, and C. C. J. Kuo, "Fast mode decision based on mode adaptation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 5, pp. 697–705, May 2010.

[7] H. Wang, S. Kwong, and C. W. Kok, "An efficient mode decision algorithm for H.264/AVC encoding optimization," *IEEE Trans. Multimedia*, vol. 9, no. 4, pp. 882–888, Jun. 2007.

[8] S. W. Jung, S. J. Baek, C. S. Park, and S. J. Ko, "Fast mode decision using all-zero block detection for fidelity and spatial scalable video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 2, pp. 201–206, Feb. 2010.

[9] L. J. Pan and Y. S. Ho, "Fast mode decision algorithm for H.264 inter-prediction," *Electronic Lett.*, vol. 43, no. 24, pp. 761–766, Nov. 2007.

[10] 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.

[11] Z. Peng, G. Jiang, M. Yu, and Q. Dai, “Fast macroblock mode selection algorithm for multiview video coding,” *EURASIP J. Image Video Process.*, vol. 2008, article 393727, p. 14, 2008.

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

[13] 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. 1553–1563, Dec. 2008.

[14] L. Shen, Z. Liu, S. Liu, Z. 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.

[15] S. H. Yang and Y. S. Liou, “Fast reference frame and mode decision for multi-view video coding based on coded block patterns,” in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2010, pp. 1102–1107.

[16] W. C. Chen, L. M. Liu, K. F. Hsu, and W. N. Lie, “A fast H.264/AVC based stereo video encoding algorithm based on hierarchical two-stage neural classification,” *IEEE J. Select. Topics Signal Process.*, vol. 5, no. 2, pp. 309–320, Apr. 2011.

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

[19] H. Q. Zeng, K. K. Ma, and C. 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.

[20] Y. Zhang, S. Kwong, G. 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.

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