Statistical Early Termination Model for Fast Mode Decision and Reference Frame Selection in Multiview Video Coding

Yun Zhang, Sam Kwong, *Senior Member, IEEE*, Gangyi Jiang, Xu Wang, and Mei Yu

***Abstract*—Multiview Video Coding (MVC) adopts exhaustive variable size mode decision and multiple reference frame selection to significantly improve high compression efficiency at each macroblock. However, these two technologies increase the computational complexity of MVC encoders tremendously. In this paper, we propose an efficient Statistical DIRECT Mode Early Termination (SDMET) model which estimates the rate distortion degradation, false acceptance rate and false reject rate of early DIRECT mode decision. It can adaptively adjust the rate distortion cost threshold not only according to the quantization parameter, but also the video content and motion properties. Experimental results show that SDMET can reduce 42.40% to 65.60% computation complexity for fast mode decision. When it is jointly optimized with fast multi-reference frame selection, the proposed overall algorithm can achieve 79.57% to 89.21% computational complexity reduction with unnoticeable rate distortion degradation. Additionally, the proposed SDMET and the overall fast mode decision algorithm can be applied to both temporal views and inter-view views in MVC.**

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

I. Introduction

An increasing demand is currently witnessed for Three-Dimensional (3D) video since it provides more real perception, interactivity and novel visual enjoyment. It would be useful for many multimedia applications, such as Free-viewpoint TeleVision (FTV) [1], 3D TeleVision (3DTV) [2] broadcasting, immersive teleconference and virtual reality. However, multiview videos captured simultaneously from multiple cameras at slightly different views or angles are required for representing genuine 3D world video content, and a tremendously huge amount of storage space, transmission bandwidth and large computing power for compression are needed as compared with traditional mono-view videos. To compress large volume of multiview video data efficiently, it is necessary to develop Multiview Video Coding (MVC) with high compression efficiency and low complexity for interactive 3D video applications, such as live 3D broadcasting and interactive 3D video communication.

Currently, Joint Video Team (JVT), which was organized by ISO/IEC Moving Picture Experts Group (MPEG) and ITU-T Video Coding Experts Group (VCEG), devotes their efforts on MVC standardization and develops novel coding algorithms based on the state-of-the-art H.264/AVC standard. Merkle *et al.* proposed a MVC scheme using Hierarchical B Pictures (MVC-HBP) [3] which has been adopted into MVC standardization draft and Joint Multiview Video Coding (JMVC) reference software due to its superior compression efficiency and temporal scalability. Additionally, variable block-size Motion/Disparity Estimation (ME/DE) [4] and Multi-Reference Frame (MRF) prediction techniques are used to reduce the temporal and inter-view redundancies. These techniques improve the coding efficiency at the cost of extremely high complexity, which is the bottleneck of enabling MVC into practical use. As recognizing from the fact that MVC-HBP is too complex, Merkle *et al.* then proposed several improved prediction structures, including KS_IPP, KS_IBP, KS_PIP, [5] to lower coding complexity by reducing reference frames of non-anchor frames and designing anchor prediction structure. However, their Rate-Distortion (RD) performance degrades for high inter-view correlation sequences due to reducing inter-view prediction. In [6], prediction structures are adaptively switched according to inter-view correlation of multiview video sequences to lower coding complexity and improve interactive functionalities, such as fast random accessibility and flexible view scalability. However, the encoding time saving ratio is still limited and unstable for multiview videos with different spatio-temporal correlations.

As variable block-size mode decision is adopted in video coding to improve compression efficiency by significantly increasing the coding complexity, a number of efforts [7]–[11] have been devoted to develop fast algorithms in Fast Mode Decision (FMD). Wang *et al.* [7] proposed a joint early terminate mode decision and ME scheme using sufficient condition to detect all-zero blocks in H.264/AVC. In [8], Pan and Ho proposed another early terminate mode decision for H.264/AVC inter-prediction by defining a threshold based on Sum of Absolute Transformed Difference (SATD) cost of SKIP mode. The threshold linearly increases with Quantization Parameter (QP). Zeng *et al.* [9] proposed motion activity-based early mode decision where an early termination is performed by using two RD cost thresholds for SKIP and INTRA mode. They fitted the RD

Manuscript received May 09, 2011; revised June 14, 2011; accepted October 23, 2011. Date of publication December 21, 2011; date of current version February 23, 2012. This work was supported in part by Hong Kong RGC General Research Fund (GRF) Projects 9041495 (CityU 115109) and in part by the National Natural Science Foundation of China under Grants 61071120, 60872094, and 61102088.

Y. Zhang is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail: yunzhang@cityu.edu.hk).

S. Kwong and X. Wang are with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong (e-mail:cssamk@cityu.edu.hk; xuwang6@student.cityu.edu.hk).

G. Jiang and M. Yu are with the Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China (e-mail: jianggangyi@nbu.edu.cn; yumei@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.2011.2174282

cost thresholds with exponential relationship of QP. Jing and Chau [10] presented a FMD scheme in which the Mean Absolute Frame Difference (MAFD) of the current frame and Mean Absolute Difference (MAD) of the current MB are compared to selectively early terminate mode decision. Utilizing motion information of spatial-temporal neighboring blocks, Zhao *et al.* [11] proposed a 2D map and then a priority-based mode candidate list is constructed for best mode selection. However, these schemes are proposed for P frames or traditional B frames in mono-view H.264/AVC. They can not be directly applied for MVC using hierarchical prediction structure and hierarchical B pictures. Additionally, when both the Disparity Compensation Prediction (DCP) and the Motion Compensation Prediction (MCP) are adopted in MVC, they might have different statistical properties from traditional temporal prediction.

Therefore, several FMD methods for MVC [12]–[19] have been proposed aiming at reducing computational complexity more. Zhu *et al.* [12] proposed a fast INTER mode decision scheme, meanwhile, inter-view prediction of the small-block size Macroblock (MB) is selectively reduced based on B16×16 prediction information for inter-view views. Peng *et al.* [13] proposed FMD for MVC with dynamic multi-threshold early termination. Ding *et al.* [14] proposed content-aware inter-view mode decision algorithm by sharing RD cost, coding modes, and motion vectors among views. Shen *et al.* [15] proposed selective variable size ME and DE for MVC based on motion homogeneity, which is identified by motion vector deviation among spatial neighboring MBs and inter-view corresponding MBs estimated by Global Disparity Vectors (GDVs) [16]. Based on similar mode distribution among views, they also proposed early SKIP mode decision for MVC by exploring inter-view mode correlation [17]. Additionally, Han and Lee [18] and Yu *et al.* [19] explore MB mode similarity among neighboring views to facilitate FMD process by using GDVs. The GDV concept assumes a globally unique displacement among inter-views. Unfortunately, this hypothesis is usually not true for toed-in camera arrangement and video with large Depth-Of-Fields (DOF). Additionally, it not accurate enough as it is measured with MB unit.

In this paper, we propose an early DIRECT termination model to adaptively adjust the RD cost threshold not only according to QP, but also to the video contents and motion properties. Then, it is applied to FMD and fast MRF to significantly reduce the computational complexity of MVC. The rest of the paper is organized as follow, the Statistical DIRECT Mode Early Termination (SDMET) algorithm for hierarchical B pictures in MVC is presented in Section II. Then, fast MRF selection and overall optimization algorithm are presented in Sections III and IV. Section V presents both the theoretical and the experimental verification of the proposed SDMET model and overall FMD algorithm. Finally, conclusions are given.

II. SDMET Model for Hierarchical B Pictures in MVC

A. Statistical Analyses and Motivation

Four multiview video test sequences with different motion properties are encoded and analyzed for their best mode distribution. Breakdancers [20] and Ballroom [21] are fast motion sequences, Ballet [20] moves moderate and Doorflowers [22] is moves slowly. Fig. 1 shows the statistical probability of MB mode distribution for hierarchical B frame in MVC-HBP,

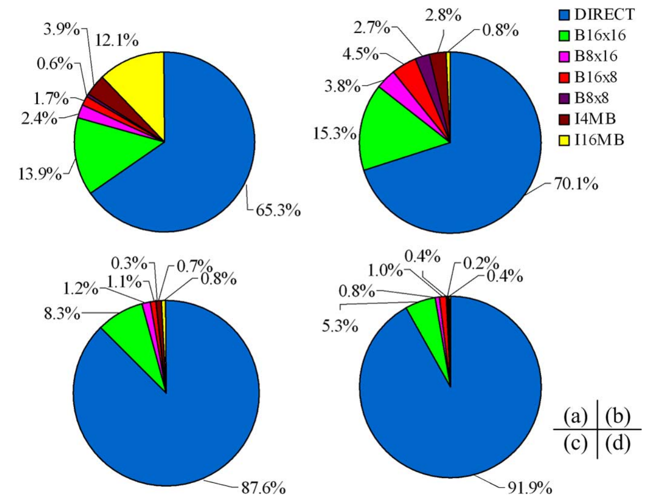


Fig. 1. Statistical results of MB mode distribution for hierarchical B frames in MVC-HBP: (a) Breakdancers (1024 × 768 at 15 fps); (b) Ballroom (640 × 480 at 25 fps); (c) Ballet (1024 × 768 at 15 fps); (d) Doorflowers (1024 × 768 at 30 fps).

where basis QP (denoted as bQP in the following sections) is 28. We can observe that 65.3% to 91.9% MBs select DIRECT mode and the percentage decreases as motion becomes fast or complicate. For Breakdancers sequence, 16.0% MBs select the INTRA mode as the best mode due to its very fast motion and low capture frame rate. For all of the four test sequences, 78.7% MBs on average select DIRECT mode as the best mode and very few MBs select small MB mode, e.g. B8×8 mode, as the best mode. Also, the percentage decreases as the MB partition becomes smaller. However, because the ME/DE search times for each MB increase as MB partition becomes smaller, small block-size mode MB usually consumes more coding time than large block-size mode MB. Additionally, most MBs are using DIRECT mode, in which no ME/DE and MRF are required, thus, their complexity can be omitted. Therefore, if we early terminate the mode selection process by checking only DIRECT mode or large block size MB modes, significant coding complexity reduction could be achieved.

While encoding one MB in hierarchical B frames, the optimal INTER mode, optimal reference direction and reference frame, $\{m^*, \psi^*, \mathbf{r}^*\}$, are determined by RD optimization. It can be mathematically expressed as

$$\{m^*, \psi^*, \mathbf{r}^*\} = \arg \min_{m\in\mathbf{M}, \psi\in\{FWD,BWD,BI\},\mathbf{r}} J\left(\mathbf{C}, \mathbf{R}(m,\psi,\mathbf{r})\right), \quad (1)$$

where $\mathbf{C}$ and $\mathbf{R}$ are the current and reference block, respectively; $\mathbf{M}$ is a set of INTER modes, including B16×16, B16×8, B8×16, B8×8Frext, B8×8, SubB8×4, SubB4×8 and SubB4×4. The reference frame indicator $\mathbf{r}$ is composed of two elements: the reference index in List0 (RefIdx0) and the reference index in List1 (RefIdx1), RefIdx0, $\mathrm{RefIdx1} \in [-1, n_{NRF} - 1]$, where n_{NRF} is number of active frame in each memory list. ψ indicates reference prediction direction, $\psi \in \{\mathrm{FWD}, \mathrm{BWD}, \mathrm{BI}\}$ which indicates forward, backward and bi-directional prediction. In this paper, we intend to optimize the MVC in two levels. One is the fast variable block size mode decision loop in which the mode selection process is selectively early terminated. The other is MRF

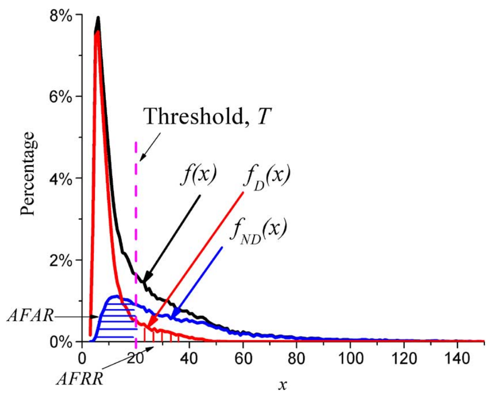


Fig. 2. Graphical explanation of the probability functions.

selection loop that selects the best reference frame by checking each active reference frame and direction repeatedly for each variable block size mode. Thus, this loop can also be optimized to further reduce coding complexity by exploring reference similarity among MB modes which has not been early terminated.

B. Proposed SDMET Model

As aforementioned that DIRECT mode has a high percentage in mode distribution and it is with lowest computational cost, we thus propose a SDMET algorithm to early terminate mode decision before checking other MB modes with DE/ME so that the computational complexity can be reduced as much as possible. Related works [8], [9] have been proposed for early SKIP/DIRECT mode termination based on RD cost thresholds. However, their thresholds are linearly or exponentially fitted from QP and have not taken different video content, motion, texture characteristic of video sequences into account. Additionally, they are proposed for P/B frames in mono-view video coding and can not be directly applied to MVC with inter-view prediction and hierarchical B pictures. In this section, a new DIRECT early termination model is proposed based on statistical information and sensitive to both QP and multiview video content, including motion, texture and inter-view correlation.

Let us denote the average square root of the RD cost of encoding a MB using DIRECT mode by a random variable $\mathbf{X}$. Let $f(x)$ be the Probability Density Function (PDF) of $\mathbf{X}$ for the MBs of a frame. Let $f_{\mathrm{D}}(x)$ be the probability function of $\mathbf{X}$ for those MBs who select DIRECT mode as the best mode and $f_{\mathrm{ND}}(x)$ be the probability function of $\mathbf{X}$ for those MBs who select non-DIRECT mode as the best mode. Fig. 2 shows an example of $f(x)$, $f_{\mathrm{ND}}(x)$ and $f_{\mathrm{D}}(x)$, which is obtained from statistical analyses of Breakdancers sequence, and these functions satisfy

$$\begin{aligned} &f_{ND}(x) = f(x) - f_D(x) \\ &\int_0^\infty f(x)dx = 1 \qquad \forall x, f(x) \geq f_D(x). \end{aligned} \tag{2}$$

Accordingly, we obtain the total percentage of the MBs who select DIRECT mode as best mode as

$$P_D = \int_0^\infty f_D(x)dx. \tag{3}$$

Here, we use square root of RD cost for low complexity because it can be directly obtained from encoding a MB with DIRECT mode. To analyze the performance of DIRECT mode early termination, we define two performance indices: Absolute False Reject Rate (AFRR) is the percentage of DIRECT mode MBs being falsely classified as non-DIRECT mode MBs, Absolute False Acceptance Rate (AFAR) is the percentage of non-DIRECT mode MBs being falsely classified as DIRECT mode MBs. Small AFRR indicates the computational complexity is reduced efficiently, meanwhile, small AFAR indicates negligible RD degradation. According to (2), AFRR and AFAR value can be calculated as an integration of $f_{\mathrm{ND}}(x)$ and $f_{\mathrm{D}}(x)$ at interval $[0, T]$, respectively. They are

$$\begin{aligned} F_{AFAR}(T) &= \int_0^T f_{ND}(x)dx \\ F_{AFRR}(T) &= \int_T^\infty f_D(x)dx, \end{aligned} \tag{4}$$

where T is the threshold for DIRECT mode early termination.

Let $P_{\mathrm{ET}}(T)$ be the early terminate rate of the DIRECT mode decision, it can be represented by Cumulative Density Function (CDF) of $f(x)$ as

$$P_{ET}(T) = \int_0^T f(x)dx. \tag{5}$$

Accordingly, the speedup ratio R_{SU} by the DIRECT mode early termination can be obtained as

$$R_{SU} = 1 - \frac{t_{DIRECT} \times P_{ET} + t_{ALL} \times (1 - P_{ET})}{t_{ALL}}, \tag{6}$$

where t_{DIRECT} and t_{ALL} are the coding time of checking DIRECT mode only and checking all MB modes, respectively. Generally, the coding time of DIRECT mode t_{DIRECT} is much smaller than t_{ALL} and thus negligible because the DIRECT mode does not have ME/DE process. Therefore, (6) can be simplified as

$$R_{SU} \approx P_{ET}. \tag{7}$$

As early termination is implemented for mode decision, incorrect mode decision may cause RD degradation whose value is direct proportional to the AFAR. Here, we use difference of Mean Square Error (MSE), denoted by ΔMSE, to evaluate the RD degradation caused. Thus, the ΔMSE is calculated as

$$\begin{aligned} \Delta MSE &= \int_0^T \left(x^2 - x_{ND}^2\right) f_{ND}(x)dx \\ &\leq \int_0^T x^2 f_{ND}(x)dx = \int_0^T x^2 \left(f(x) - f_D(x)\right) dx \\ &= \Delta MSE_{Up}, \end{aligned} \tag{8}$$

where x and x_{ND} denote the average square root of RD cost value for coding a MB with DIRECT and non-DIRECT mode, respectively. However, in the optimized coding process, value

x_{ND} is unknown in early termination. Therefore, x_{ND} is assumed to be zero which indicates the worst case of RD degradation. We denote the maximum value of estimated ΔMSE as ΔMSE_{Up}. Obviously, ΔMSE_{Up} increases as T increases.

As the objective for early termination of DIRECT mode is to maximize speed-up ratio R_{SU} subject to endurable RD degradation, i.e. small $F_{\mathrm{AFAR}}(T)$ and ΔMSE_{Up}. The optimization target can be mathematically expressed as

$$max(R_{SU}) \quad s.t. \quad \begin{cases} F_{AFAR}(T) < T_{AFAR} \\ \Delta MSE_{Up} \leq T_{\Delta MSE}, \end{cases} \tag{9}$$

where T_{AFAR} and $T_{\Delta\mathrm{MSE}}$ are thresholds for AFAR and ΔMSE_{Up}, respectively. In fact, maximizing R_{SU} is equivalent to minimizing AFRR, which is decreasing the number of false reject MBs as much as possible. Therefore, the conditional equation (9) can be rewritten as

$$T^* = \arg_T \min\left(F_{AFRR}(T) + \phi \cdot F_{AFAR}(T)\right), \tag{10}$$

where ϕ is a weighted coefficient indicating the relative importance between AFAR and AFRR, $\phi > 0$, larger ϕ indicates AFAR is more important and stricter on limiting RD degradation. Meanwhile, $\Delta MSE_{\mathrm{Up}} \leq T_{\Delta\mathrm{MSE}}$ is used as verification while (10) solved.

Since DC coefficients for motion compensation prediction errors obey Laplace distribution [23] and the PDF of $\mathbf{X}$, $f(x)$, is in Laplacian shape while $x > 0$, as shown in Fig. 2, according to the statistical analyses of Breakdancers, we assume $f(x)$ is a Laplacian source, that is

$$f(x) = \begin{cases} \frac{1}{\sigma}\exp\left\{-\frac{x-\mu}{\sigma}\right\} & x > \mu \\ 0 & x \leq \mu, \end{cases} \tag{11}$$

where σ and μ are parameters. Similar to $f(x)$, $f_{\mathrm{D}}(x)$ is also assumed to be a Laplacian source. According to constraints in (2), probability function of DIRECT mode, $f_{\mathrm{D}}(x)$, can be represented as

$$f_D(x) = \begin{cases} P_D\frac{1}{\sigma_D}\exp\left\{-\frac{x-\mu_D}{\sigma_D}\right\} & x > x_0 \\ f(x) & x \leq x_0, \end{cases} \tag{12}$$

where $\sigma_D < \sigma$, $\mu_D \approx \mu$, $x_0 = \mu + (\sigma\sigma_D/\sigma - \sigma_D)\ln(P_D\sigma/\sigma_D)$. Applying (11) and (12) to (8), we obtain ΔMSE_{Up} as

$$\Delta MSE_{Up} = \begin{cases} \Phi(T) - \Phi(x_0) & T > x_0 \\ 0 & T \leq x_0, \end{cases} \tag{13}$$

where

$$\begin{aligned}\Phi(y) = {} & \left[(\mu+\sigma)^2 + \sigma^2\right]\left(1 - e^{-\frac{y-\mu}{\sigma}}\right) \\ & - (y-\mu)(\mu + 2\sigma + y)e^{-\frac{y-\mu}{\sigma}} \\ & - P_D\left[(\mu+\sigma_D)^2 + \sigma_D^2\right]\left(1 - e^{-\frac{y-\mu}{\sigma_D}}\right) \\ & + P_D\,(y-\mu)\,(\mu_D + 2\sigma_D + y)e^{-\frac{y-\mu}{\sigma_D}}. \end{aligned} \tag{14}$$

On the other hand, the AFAR and AFRR can be obtained as

$$F_{AFAR}(T) = \begin{cases} \left(e^{-\frac{x_0-\mu}{\sigma}} - P_D e^{-\frac{x_0-\mu_D}{\sigma_D}}\right) & \\ \quad - \left(e^{-\frac{T-\mu}{\sigma}} - P_D e^{-\frac{T-\mu_D}{\sigma_D}}\right) & T > x_0 \\ 0 & T \leq x_0, \end{cases} \tag{15}$$

$$F_{AFRR}(T) = \begin{cases} P_D e^{-\frac{T-\mu_D}{\sigma_D}} & T \geq x_0 \\ P_D & T < x_0. \end{cases} \tag{16}$$

Also, early terminate rate for DIRECT mode, P_{ET}, is

$$P_{ET}(T) = 1 - e^{-\frac{T-\mu}{\sigma}}. \tag{17}$$

From (13)–(17), $F_{\mathrm{AFAR}}(T)$, $F_{\mathrm{AFRR}}(T)$, ΔMSE_{Up}, and $P_{\mathrm{ET}}(T)$ are functions of T.

To obtain the optimal T for minimizing $F_{\mathrm{AFAR}}(T) + \phi F_{\mathrm{AFRR}}(T)$, its deviation is calculated and set to be zero. That is

$$\frac{\partial\left(F_{AFRR}(T) + \phi \cdot F_{AFAR}(T)\right)}{\partial T} = 0. \tag{18}$$

Applying (15)–(18) and then solving (18), we can obtain the optimal T as

$$T = \frac{\sigma\mu_D - \sigma_D\mu}{\sigma - \sigma_D} + \frac{\sigma_D\sigma\ln\left(\frac{1+\phi}{\phi}P_D\frac{\sigma}{\sigma_D}\right)}{\sigma - \sigma_D}, \tag{19}$$

where σ, μ, μ_{D}, σ_{D} and P_{D} are statistical parameters. These parameters are used to early terminate mode decision of current MB, but they can only be obtained after the encoding process. To solve this chicken-egg dilemma, σ, μ, μ_{D}, σ_{D} and P_{D} are predicted from previous coded frames, that is $\sigma = \sqrt{E(X^2) - E(X)^2}$, $\mu = E(X) - \sqrt{E(X^2) - E(X)^2}$, $\sigma_{\mathrm{D}} \approx E(X_{\mathrm{D}}) - \mu$ and $\mu_{\mathrm{D}} \approx \mu$, where $E(\cdot)$ is the symbol of mathematical expectation. As $\mu_{\mathrm{D}} \approx \mu$, (19) is simplified as

$$T = x_0 + \frac{\sigma_D\sigma}{\sigma - \sigma_D}\ln\left(\frac{1+\phi}{\phi}\right). \tag{20}$$

Based on the optimization constraints of MSE in (9), $\Delta MSE_{\mathrm{Up}} \leq T_{\Delta\mathrm{MSE}}$ shall also be satisfied with the threshold T applied to (13). To adaptively select the optimal ϕ and maximize the speed up ratio, the estimated ΔMSE_{Up} shall be close to $T_{\Delta\mathrm{MSE}}$ as much as possible. However, in the video coding optimization, $T_{\Delta\mathrm{MSE}}$ is not directly given by user because it is an absolute distortion and independent with QP. Instead, $T_{\Delta\mathrm{PSNR}}$ is provided and it can be converted to $T_{\Delta\mathrm{MSE}}$ as

$$T_{\Delta MSE} = \left(10^{\frac{T_{\Delta PSNR}}{10}} - 1\right) \times 255^2 \Big/ \left(P_D \times 10^{\frac{PSNR_{Org}}{10}}\right), \tag{21}$$

where $PSNR_{\mathrm{org}}$ and P_{D} are Peak Signal-to-Noise Ratio (PSNR) value of the frame and percentage of DIRECT mode used in encoding current frame, respectively. We can obtain the $T_{\Delta\mathrm{MSE}}$ from the given endurable PNSR degradation, threshold $T_{\Delta\mathrm{PSNR}}$, for the early termination. For example, $T_{\Delta\mathrm{PSNR}} = 0.1$ dB, $PSNR_{\mathrm{org}} = 37$ dB, $P_{\mathrm{D}} = 60\%$, $T_{\Delta\mathrm{MSE}} = 0.508$. Based on this $T_{\Delta\mathrm{MSE}}$, we can get the speed up ratio R_{SU} of the proposed SDMET using (5) and (7). Given larger endurable PSNR degradation, $T_{\Delta\mathrm{MSE}}$ is enlarged and then more R_{SU} could be achieved. However, due to the fact that P_{D} and $PSNR_{\mathrm{org}}$ can not be obtained before coding the current frame, they are estimated from P_{D} and average PSNR values of previous encoded frames.

Therefore, the proposed SDMET algorithm is illustrated in Fig. 3 and the steps are described as follows.

Step 1). If it is the first n_{wo} blocks of the first non-anchor frame of view S0 or S1 of a Group-Of-Picture (GOP), encode the n_{wo} blocks without early termination optimization and do parameters initialization, then go to Step 2. Otherwise, if it is the first n_{wo} blocks of the first frame of view S2 to S7, statistical information σ, μ, μ_{D}, σ_{D} and P_{D} are referenced from neighboring views, i.e. S2→S0, S4→S2, S6→S4; S3→S1, S5→S3, S7→S5, where symbol "→" is a reference direction, then go to Step 3;

Step 2). Calculate σ, μ, μ_{D}, σ_{D} and P_{D} by using $E(\mathrm{X})$ and $E(\mathrm{X}^2)$ from previous coded blocks;

Step 3). Load previous ϕ, calculate T_ϕ and $T_{\phi-\Delta\phi}$ by using (20) with ϕ and $\phi - \Delta\phi$;

Step 4). Apply T_ϕ and $T_{\phi-\Delta\phi}$ to (13) and calculate the estimated theoretical ΔMSE_{Up}, $\Delta MSE_{\mathrm{Up}}(T_{\phi-\Delta\phi})$ and $\Delta MSE_{\mathrm{Up}}(T_\phi)$, they satisfies $\Delta MSE_{\mathrm{Up}}(T_{\phi-\Delta\phi}) > \Delta MSE_{\mathrm{Up}}(T_\phi), T_{\phi-\Delta\phi} > T_\phi$;

Step 5). If $\Delta MSE_{\mathrm{Up}}(T_{\phi-\Delta\phi}) < T_{\Delta\mathrm{MSE}}$, then, $\phi = \phi - \Delta\phi$, $n = n + 1$, go to Step 6. If $\Delta MSE_{\mathrm{Up}}(T_\phi) > T_{\Delta\mathrm{MSE}}$, then $\phi = \phi + \Delta\phi, n^+ = n^+ + 1$, go to Step 6. Otherwise, if the theoretical ΔMSE_{Up} satisfies $\Delta MSE_{\mathrm{Up}}(T_\phi) \le T_{\Delta\mathrm{MSE}} \le \Delta MSE_{\mathrm{Up}}(T_{\phi-\Delta\phi})$, go to Step 7;

Step 6). If $n^- > N^-$or $n^+ > N^+$, then go to Step 7; otherwise, go to Step 3 to recalculate T_ϕ and $T_{\phi-\Delta\phi}$;

Step 7). Store ϕ, T_ϕ and calculate the final threshold for the early termination as $T_{\mathrm{DIRECT}} = T_\phi^2 \times 256$;

Step 8). Encode next m MBs with early termination threshold T_{DIRECT}. If last MB of one frame is encoded, update $T_{\Delta\mathrm{MSE}}$ according to (21). If this is last frame of the current GOP, go to Step 1 for next GOP; else, go to Step 2 update statistical information and thresholds for next frame.

Here, T_ϕ is originally denoted by T in (3) to (20), $\Delta\phi$ measures the fidelity and indicates the difference between ΔMSE_{Up} and target $T_{\Delta\mathrm{MSE}}$, $\Delta\phi \times N^+ \times m$ and $\Delta\phi \times N^- \times m$ are the response steps which indicate how fast to get optimal T_{DIRECT} from initial ϕ setting, where N^+ and N^- are the maximum recalculation times in increasing and decreasing updating cycle. n_{wo} and m are SDMET model parameters. Coefficient m is number of MBs for each updating cycle for threshold T_ϕ. Coefficient n_{wo} is the number of MBs encoded without optimization. These n_{wo} MBs are encoded in order to collect statistical characteristic information of the current multiview video sequences. The value n_{wo} is the trade-off between stability and speed-up ratio of the proposed algorithm. As n_{wo} increases, the stability of the proposed algorithm is improved, but the speed-up ratio may decrease.

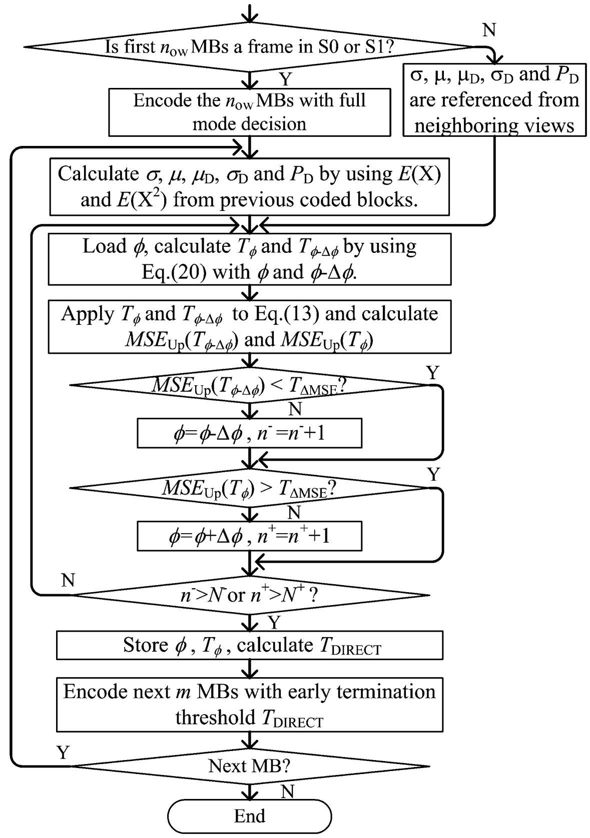


Fig. 3. Diagram of the proposed SDMET.

III. Joint Optimization With Fast MRF Selection Algorithm

Due to similar properties and high spatial correlation of blocks within one MB, smaller MB partition modes, e.g. B8×8 and B16×8, will probably also select the same prediction direction and reference indices as B16×16 does. It is of high probability that $\psi^* = \psi_{\mathrm{B16}\times 16}$ and $\mathbf{r}^* = \mathbf{r}_{\mathrm{B16}\times 16}$ for encoding modes with partition size smaller than 16×16 [24], where ψ^* indicates the best reference direction, $\mathbf{r}^*$ indicates the best reference frame index in each direction. This probability is statistically analyzed with MVC experiments. Three multiview video sequences, Breakdancers (fast motion), Ballet (moderate motion) and Doorflowers (slow motion), are analyzed. Fast ME/DE is enabled and the parameter n_{NRF} is set to 2. Eight views, including four temporal coded views (even views) and four inter-view/temporal joint coded views (odd views), are encoded.

Fig. 4 shows the statistical analyses for probability of $\psi^* = \psi_{\mathrm{B16}\times 16}$ and $\mathbf{r}^* = \mathbf{r}_{\mathrm{B16}\times 16}$, denoted as ρ, via exhaustive search method, which searches all reference frames, directions and MB modes. The x-axis of the figures shows different view section and different frames in each view section. The y-axis is the hit probability ρ when bQP is 28. We can see that the probability ρ is higher than 93% for different views and tested multiview video sequences with different motion properties. Usually, the probabilities of temporal views are higher than those of inter-view views. Additionally, Table I shows average probability ρ for different bQPs by searching all reference frames, directions and MB modes. We can see that over 93.78% reference frame indices and direction of MB and sub-MB partition

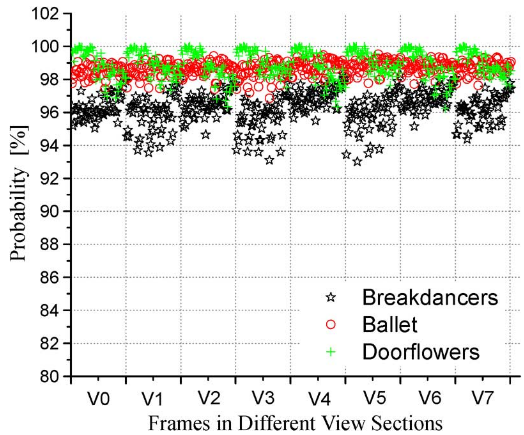


Fig. 4. Statistical analyses on probability ρ when bQP is 28.

TABLE I
AVERAGE PROBABILITY ρ FOR DIFFERENT BASIS QPs [Unit: %]

bQP	Breakdancers	Ballet	Doorflowers
24	93.78	97.76	99.17
28	96.08	98.51	99.60
32	97.42	99.04	99.81
36	98.35	99.37	99.89

are the same as that of B16×16 MB mode. Additionally, similar statistical results can be found for full motion/disparity search and different number of reference frames.

IV. OVERALL FMD ALGORITHM AND COMPLEXITY ANALYSES

Since anchor frames usually have different statistical information with non-anchor frames and have much less DIRECT mode MBs, the proposed overall FMD algorithm is only applied for non-anchor hierarchical B pictures. The flowchart of the proposed overall FMD algorithm is illustrated in Fig. 5, and it is described as

Step1) Encode the current MB with DIRECT mode and store the RD cost of the DIRECT mode as J_{RD}.

Step2) Calculate the statistical threshold, T_{DIRECT}, for the DIRECT mode with SDMET model, then compare J_{RD} with T_{DIRECT}. If $J_{\mathrm{RD}} \geq T_{\mathrm{DIRECT}}$, go to Step3; else if $J_{\mathrm{RD}} < T_{\mathrm{DIRECT}}$, set DIRECT mode as the best mode and go to Step 6.

Step3) Encode the current MB with B16×16 mode and obtain the best reference frames RefIdx_B16×16_List0 and RefIdx_B16×16_ List1 in the memory List0 and List1, respectively. Save prediction directions of the B16×16 mode as flags BIPred- B16×16, FWDPredB16×16 and BWDPredB16×16.

Step4) Encode the current MB with other MB partitions, i.e. B16×8, B8×16, B8×8, B8×8Frext, and subMB partitions with the same reference indices/direction as the indices/direction of B16×16 mode.

Step5) Encode the current MB with INTRA modes.

Step6) Store the coding parameters with the smallest RD cost and write coded bitstream. Then go to Step 1 for next MB.

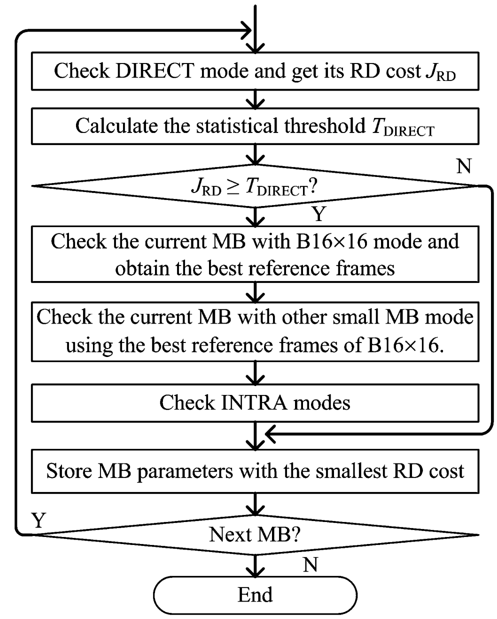


Fig. 5. Diagram of the proposed overall FMD algorithm.

To theoretically analyze to complexity of the proposed algorithms, let α be the computational complexity of encoding MB mode with B16×16 in the original multiview video encoder, therefore, computational complexity of DIRECT, B16×8, B8×16, B8×8 (including SubDIRECT, B8×4, B4×8, B4×4) and INTRA mode (including I16 MB, I8 MB, I4 MB and PCM) can be presented as $\lambda_1\alpha$, $\lambda_2\alpha$, $\lambda_3\alpha$, $\lambda_4\alpha$, and $\lambda_5\alpha$, respectively, where λ_{i} are multiplication coefficients, $0 < \lambda_{\mathrm{i}}$, $i \in \{1, 2, 3, 4, 5\}$. Thus, the total complexity of original JMVC multiview video encoder C_{JMVC} can be calculated as the summation of coding complexity of all the MB modes

$$C_{JMVC} = \left(1 + \sum_{i=1}^{5} \lambda_i\right)\alpha. \tag{22}$$

Accordingly, the computational complexity of MVC using SDMET, C_{SDMET}, can be calculated as

$$C_{SDMET} = P_{ET}\lambda_1\alpha + (1 - P_{ET})\left(1 + \sum_{i=1}^{5} \lambda_i\right)\alpha, \tag{23}$$

where P_{ET} is the early DIRECT mode terminate rate. Because DIRECT mode do not have ME/DE search, its computational complexity is negligible while compared with other INTER modes, i.e. $\lambda_1 \approx 0$. Thus, the complexity reduction ratio of SDMET is $1 - C_{\mathrm{SDMET}}/C_{\mathrm{JMVC}} \approx P_{\mathrm{ET}}$, where P_{ET} has been defined in (5).

Let β_1 and β_2 be the complexity of multi-reference search for forward and backward of B16×16, respectively. Based on our empirical experiences, β_1 can be considered approximately equal to β_2, so only β is used, i.e. $\beta_1 \approx \beta_2 = \beta$. Let $\lambda_{\mathrm{bi}}\beta$ be the complexity of bi-directional iterative search, where λ_{bi} is a coefficient that depends on the number of iterations, search algorithm and iterative search range. Because INTER modes, except DIRECT mode, is composed of forward/backward multi-reference search and bi-directional iterative search, α equals to

TABLE II
TEST MULTIVIEW VIDEO SEQUENCES

Multiview video	Provider	Resolution	Frame rate(fps)/ Spacing(cm)/ Camera Array	Features
Race1	KDDI	640×480	30 /20/1D	Global fast motion
Ballroom	MERL	640×480	25 /20/1D	Large disparity, rotated motion
Exit		640×480	25/20/1D	Large disparity
Lovebird1	ETRI	1024×768	30 /3.5/1D	Outdoor scene, small disparity
Dooflowers	HHI	1024×768	16.7 /6.5/1D	Slow motion
Ballet	MSR	1024×768	15 /20/1D-arc	Fast and small motion
Breakdancers		1024×768	15 /20/1D-arc	Very fast motion
Dog	Nagoya Univ.	1280×960	30 /5/1D	Slow motion, large image size

$\beta + \beta + \lambda_{\rm bi}\beta$. Let p_1, p_2 and p_3 be the probabilities of selecting forward, backward and bi-directional iterative search as the best direction in MVC coding process, where $p_1+p_2+p_3=100\%$. Thus, the total computational complexity of the proposed overall FMD algorithms can be calculated as

$$C_{ALL} = \lambda_1\alpha + (1-P_{ET})\left(\alpha + \sum_{i=2}^{4}\lambda_i\gamma + \lambda_5\alpha\right), \quad (24)$$

where $\gamma = (p_1\beta + p_2\beta + p_3\alpha)/n$, n is the number of active reference frames in each memory list. On the basis of complexity analyses of MVC coding process, we obtain $\lambda_1 \approx 0$, $\lambda_2+\lambda_3+\lambda_4 \approx 8.75$, $\lambda_5 \approx 0.25$, $\lambda_{\rm bi} \approx 2$ when search range is ±96, number of bi-prediction iteration is 4 and search range for iterations is 8, fast ME/DE is enabled and number of reference frames equals 2. Hence, the total complexity of encoding one MB via original JMVC and proposed fast overall FMD algorithm are 10α and $(2.34\alpha+3.28\alpha p_3)(1-P_{\rm ET})$, respectively. For different multiview video sequences, p_3 usually ranges from 5% to 35%, and 20% on average. The complexity of proposed fast algorithm will approximately be $3.0\alpha(1-P_{\rm ET})$, which means 85.0% complexity reduction can be achieved if $P_{\rm ET}$ is 50%.

V. EXPERIMENTAL RESULTS AND ANALYSES

To verify the accuracy and effectiveness of the proposed algorithms, two-phase experiments are performed in this section. One is the theoretical model verification for SDMET, and the other is the MVC video coding experiment to testify the effectiveness of SDMET and the proposed overall FMD algorithm. Eight multiview video test sequences, including Race1, Ballroom, Exit, Lovebird1, Doorflowers, Breakdancers, Ballet and Dog [20]–[22], [25]–[27], with various motion properties and camera arrangements are adopted in MVC experiments and related theoretical analysis. Detailed information of the test sequences is shown in Table II.

A. Theoretical Verification for SDMET

To verify the accuracy and effectiveness of the proposed SDMET model, five typical multiview video sequences, including Ballet (moderate motion) [20], Breakdancers (very fast motion) [20], Dog (very slow motion) [25], Ballroom (fast motion) [21], and Doorflowers (slow motion) [22] sequences, are statistically and theoretically analyzed.

Figs. 6 and 7 show the statistical $\Delta MSE_{\rm Up}$ originated from MVC and theoretical $\Delta MSE_{\rm Up}$ calculated by the proposed SDMET model, respectively, for different video sequences and bQP. The x-axis is the threshold T_ϕ used in SDMET and the y-axis is the theoretical/statistical average $\Delta MSE_{\rm Up}$ or their $\Delta MSE_{\rm Up}$ difference for each MB. Fig. 6(a) and (b) shows the average $\Delta MSE_{\rm Up}$ for Breakdancers with different bQP, in which each line stands for one bQP. Since the target of the proposed SDMET is to early terminate DIRECT mode with endurable $\Delta MSE_{\rm Up}$, it means only small average $\Delta MSE_{\rm Up}$ is allowed. Thus, we enlarge the dash rectangles in Fig. 6(a)–(c) for better observation, as shown in Fig. 6(d)–(f). We obtain the following three facts from the Fig. 6(a)–(f). 1) The average $\Delta MSE_{\rm Up}$ is zero when threshold T_ϕ is small. As the threshold T_ϕ increases, the average $\Delta MSE_{\rm Up}$ increases. 2) When T_ϕ is fixed, the magnitude of average $\Delta MSE_{\rm Up}$ increases as the bQP increases. 3) The increasing trend and magnitude of the curves in the two figures (statistical and theoretical) are very similar correspondingly for different bQPs. Fig. 6(c) shows the $\Delta MSE_{\rm Up}$ difference where the y-axis is calculated from the statistical $\Delta MSE_{\rm Up}$ minus theoretical $\Delta MSE_{\rm Up}$. The $\Delta MSE_{\rm Up}$ difference is much smaller than the actual and theoretical values. Also, it increases as bQP increases. As shown in Fig. 6(c) and (f), most points are negative values which indicate that the theoretical curves are a little larger than the statistical data so that stricter constraints are adopted to guarantee RD degradation is within pre-defined constraints.

Fig. 7 shows the statistical, predicted $\Delta MSE_{\rm Up}$ and $\Delta MSE_{\rm Up}$ difference as T_ϕ increases for different test sequences when bQP is fixed as 28. We can observe that the magnitudes of the $\Delta MSE_{\rm Up}$ increases as motion characteristic of the video sequences becomes faster and more complicate. Also, the increasing trend and magnitude of the curves in the statistical and theoretical figures are very similar correspondingly for videos with different motion properties and camera arrays.

From Fig. 7(c) and (f), we can see that the $\Delta MSE_{\rm Up}$ difference is small compared with the theoretical and statistical curves. In addition, the $\Delta MSE_{\rm Up}$ difference increases as the motion becomes fast. Though there are still a number of mismatches between the predicted and statistical $\Delta MSE_{\rm Up}$ values, it is also accurate enough to estimate the potential $\Delta MSE_{\rm Up}$ caused by increasing T_ϕ for DIRECT mode early termination.

In addition to $\Delta MSE_{\rm Up}$, AFRR and AFAR values of SDMET model are also analyzed. Fig. 8 shows the AFRR and AFAR for different multiview video sequences, where "line" is AFAR curve and "line with symbol" is AFRR. Fig. 8(a) shows the statistical AFRR and AFAR while threshold T_ϕ is adopted in early termination. Fig. 8(b) shows the theoretical AFRR and AFAR which are calculated using the proposed SDMET model. As the threshold T_ϕ increases, more and more DIRECT mode MBs will be correctly early terminated so that the AFRR decreases till zero, which indicates computational complexity of mode decision is significantly reduced. However, as T_ϕ increases, AFAR will increase because more non-DIRECT mode MBs will be falsely early terminated as the DIRECT mode, which may cause RD degradation. Therefore, low AFRR indicates significant computational complexity reduction, and

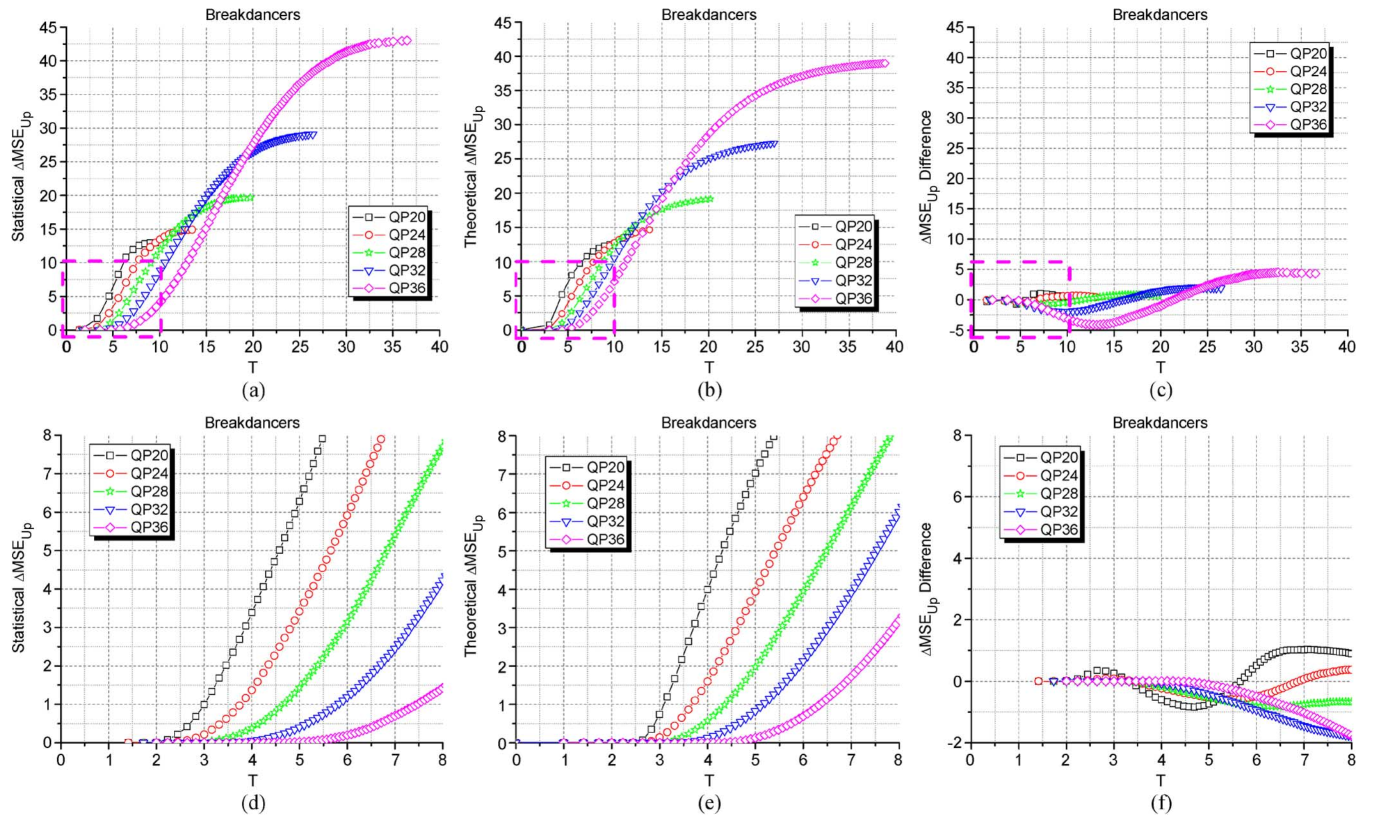


Fig. 6. Statistical and theoretical ΔMSE_{Up} for different bQPs (Breakdancers): (a) statistical ΔMSE_{Up}; (b) theoretical ΔMSE_{Up}; (c) ΔMSE_{Up} difference between statistical data and theoretical value; (d) enlarged figure of rectangle in (a); (e) enlarged figure of rectangle in (b); (f) enlarged figure of rectangle in (c).

low AFAR indicates small RD degradation. In order to achieve the MVC optimization target, which is significantly reducing the complexity with unnoticeable RD degradation, T_ϕ is determined as the value keeping both AFAR and AFRR in a low level. From the comparison of the theoretical and statistical curves in Fig. 8, we can see that the proposed model can predict the AFAR/AFRR value well for multiview video sequences with different motion properties, camera arrangements and resolutions. Fig. 8(c) shows AFAR and AFRR difference between statistical data and theoretical data. The average AFAR and AFRR differences are 0.40% and -2.05%, respectively. The mismatches mainly occur in AFRR prediction when T_ϕ is small due to large AFRR value and approximation operations in the SDMET algorithm.

B. MVC Coding Experiments

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 iterations is 8. The maximum number of reference frame is 2 for each memory list and GOP length is 12. Eight different multiview video test sequences with various motion properties and camera arrangement are encoded. Table II shows the properties of the test multiview video sequences. Eight views for each multiview video sequences and 5 GOPs for each view are encoded. Four bQP values, 24, 28, 32 and 36, are used in our experiments. The coding parameters are also consistent for original JMVC, Zhu09 [12], ShenTB09 [15], and ShenSPIC10 [17] for fair comparison. In the proposed SDMET, $T_{\Delta\mathrm{PSNR}}$ is set as 0.2 dB in order to maximize the speed-up ratio. That is the PSNR degradation constraint is 0.2 dB while the computational complexity is reduced. Additionally, initial ϕ is set as 5 and step $\Delta\phi$ is set as 0.01. Model parameter n_{wo} is set as quarter of the number of MBs in one frame for good stability and speed-up ratio according to our experiment testing. The model parameter m is set as 100 which means (13) and (20) are updated every 100 MBs. Video coding experiments are performed on Dell OPTIPLEX GX620 computer, Intel Pentium IV dual Core 3.20 GHz and 3.19 GHz CPU, 2 GB memory, Microsoft Windows XP Professional operating system. The MVC coding experiments has two phases: the MVC coding performances of SDMET is analyzed in phase one; then, the proposed overall FMD algorithm is analyzed in phase two.

Tables III and IV shows the encoding time, PSNR, bit rate comparison among original JMVC, ShenSPIC10, Zhu09, ShenTB09, proposed SDMET and proposed overall FMD scheme (denoted by ‘ALL’), where the time saving ratio, PSNR difference and bit rate increment between the original JMVC encoder and compared algorithms are computed as

$$\begin{cases} \Delta T_\theta = \frac{T_{JMVC} - T_\theta}{T_{JMVC}} \times 100[\%] \\ \Delta PSNR_\theta = PSNR_\theta - PSNR_{JMVC} \\ \Delta R_\theta = \frac{R_\theta - R_{JMVC}}{R_{JMVC}} \times 100[\%], \end{cases} \tag{25}$$

where T_θ, $PSNR_\theta$ and R_θ are total encoding time, PSNR and bit rate of scheme θ, $\theta \in \{\mathrm{Zhu09}, \mathrm{ShenSPIC10}, \mathrm{ShenTB09}, \mathrm{SDMET}, \mathrm{ALL}\}$, T_{JMVC},

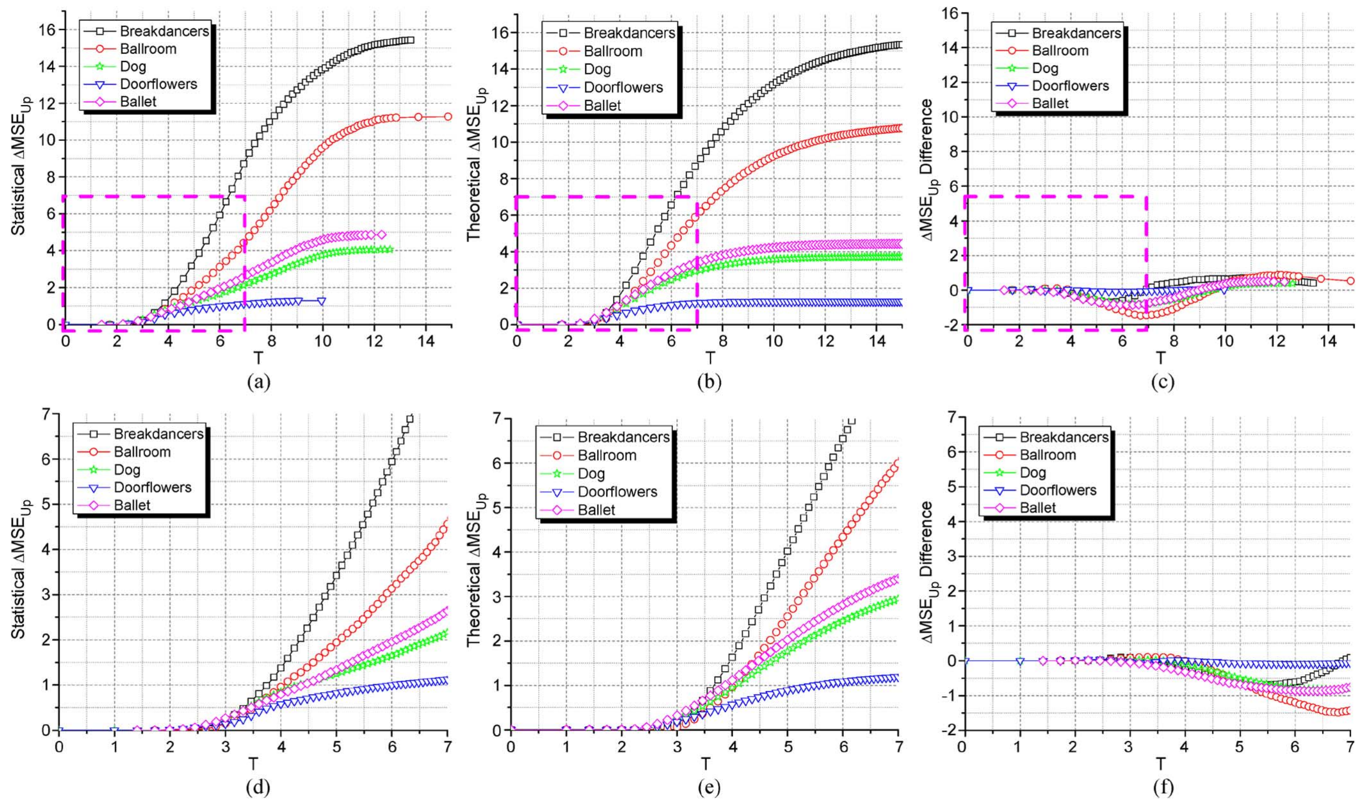


Fig. 7. Statistical and theoretical ΔMSE_{Up} for five different multiview video sequences (bQP is 28): (a) statistical ΔMSE_{Up}; (b) theoretical ΔMSE_{Up}; (c) ΔMSE_{Up} difference between statistical data and theoretical value; (d) enlarged figure of rectangle in (a); (e) enlarged figure of rectangle in (b); (f) enlarged figure of rectangle in (c).

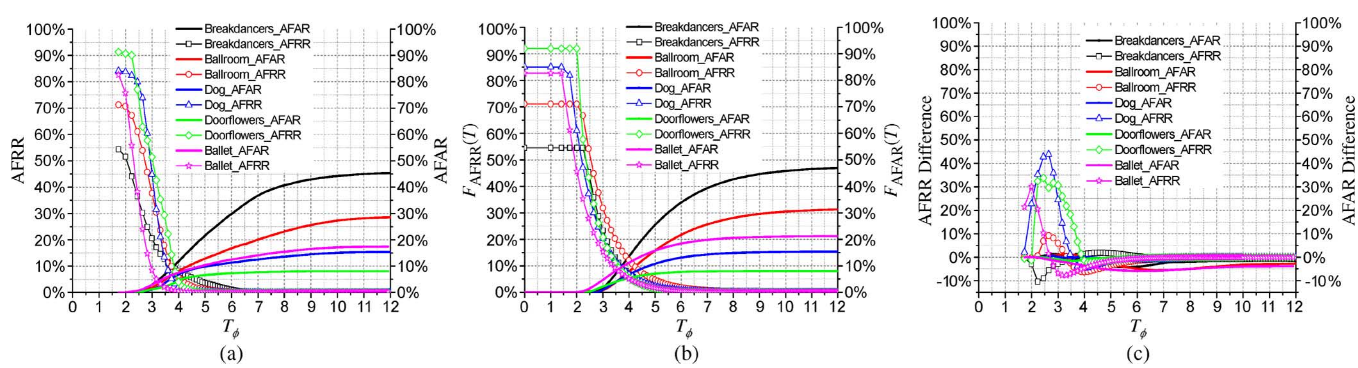


Fig. 8. Statistical and theoretical AFAR/AFRR for different video sequences: (a) statistical AFAR and AFRR; (b) theoretical AFAR and AFRR; (c) AFAR and AFRR difference between statistical and theoretical data.

$PSNR_{\mathrm{JMVC}}$ and R_{JMVC} are total encoding time, PSNR and bit rate of the original JMVC. Also, Bjonteggard Delta PSNR (BDPSNR) and Bjonteggard Delta Bit Rate (BDBR) [28] are used to measure the average PSNR and bit rate differences between RD curves.

Table III shows bit rate, PSNR and encoding time comparisons among original JMVC, ShenSPIC10 and proposed SDMET. ShenSPIC10 is early DIRECT mode termination for MVC based on the analysis of prediction mode distribution regarding the corresponding MBs in the neighboring views. SDMET is also proposed for early DIRECT mode termination, so we compare ShenSPIC10 scheme with proposed SDMET for fair comparison. We can observe that ShenSPIC10 reduces computational complexity from 19.75% to 61.76%, 39.84% on average, for the odd views. Meanwhile, the average PSNR degrades from 0.03 dB to 0.16 dB and bit rate increases from −0.84% to 3.10%. The RD degradation is mainly due to inaccurate MB unit GDV, also, the assumption of video objects having the same displacement among views is usually not correct, especially for sequences with large DOF and captured by toed-in camera arrangement. For fast motion video sequences, such as Breakdancers and Ballroom, ShenSPIC10 can only achieve 19.86% and 29.89% complexity reduction because of sparse distribution of DIRECT mode in even views. In addition, ShenSPIC10 scheme can only be applied to inter-view (odd) views and no complexity reduction is achieved for temporal (even) views. As for the proposed SDMET, though PSNR degradation is from 0.02 dB to 0.11 dB, 0.05 dB on average, for different kinds of video sequences, bit rate increases from −0.13% to −1.11%, which means bit rate is saved. The

TABLE III
BIT RATE, PSNR, AND ENCODING TIME COMPARISONS AMONG ORIGINAL JMVC, ShenSPIC10, AND THE PROPOSED SDMET

			$\Delta R_{ShenSPIC10}$ (UNIT:%) / $\Delta PSNR_{ShenSPIC10}$ (UNIT:dB) / $\Delta T_{ShenSPIC10}$ (UNIT:%)			
ShenSPIC10 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
			$\Delta R_{ShenSPIC10}$ (UNIT:%) / $\Delta PSNR_{ShenSPIC10}$ (UNIT:dB) / $\Delta T_{ShenSPIC10}$ (UNIT:%)			
		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 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
			ΔR_{SDMET} (UNIT:%) / $\Delta PSNR_{SDMET}$ (UNIT:dB) / ΔT_{SDMET} (UNIT:%)			
		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

BDPSNR between SDMET and JMVC ranges from -0.01 to -0.08 dB (-0.04 dB on average) which is smaller than that of ShenSPIC10 scheme. Meanwhile, as far as the computational complexity is concerned, the proposed SDMET can reduce computational 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, which is 14.85% more complexity reduction than ShenSPIC10 scheme while maintaining better RD performance. While the early terminate rate analyses of the two schemes is concerned, the AFARs of SDMET and ShenSPIC10 is around 0.5%, and SDMET's AFRR is 27.8%, which is 6.47% smaller than that of ShenSPIC10. Additionally, the proposed SDMET can be applied for optimizing both inter-view and temporal views.

Table IV shows bit rate, PSNR and encoding time comparisons among original JMVC, Zhu09, ShenTB09 and proposed

TABLE IV
BIT RATE, PSNR, AND ENCODING TIME COMPARISONS AMONG ORIGINAL JMVC, Zhu09, ShenTB09, AND THE PROPOSED OVERALL FMD SCHEME

Scheme						
Zhu09 Scheme [12] VS Original JMVC			ΔR_{Zhu09} (UNIT:%) / $\Delta PSNR_{Zhu09}$ (UNIT:dB) / ΔT_{Zhu09} (UNIT:%)			
		bQP	Ballet	Breakdancers	Doorflowers	Lovebird1
	Even Views	24,28, 32,36	0.0/0.0/0.0			
	Avg.					
	BDBR(%)/BDPSNR(dB)					
	Odd Views	24	0.24/-0.01/51.51	0.11/-0.02/49.87	0.31/-0.01/63.80	0.28/-0.01/65.37
		28	-0.03/-0.01/54.45	0.05/-0.02/49.84	0.55/-0.01/57.96	0.08/-0.01/58.17
		32	0.20/-0.01/51.56	0.03/-0.01/47.37	0.55/-0.01/55.34	0.03/-0.01/54.15
		36	0.22/-0.01/50.52	0.09/-0.01/45.30	0.46/0.00/54.08	0.00/0.00/51.47
	Avg.		0.16/-0.01/52.01	0.07/-0.01/48.10	0.47/-0.01/57.80	0.10/-0.01/57.39
	BDBR(%)/BDPSNR(dB)		0.45/-0.01	0.59/-0.01	0.66/-0.02	0.97/-0.02
			ΔR_{Zhu09} (UNIT:%) / $\Delta PSNR_{Zhu09}$ (UNIT:dB) / ΔT_{Zhu09} (UNIT:%)			
		bQP	Dog	Ballroom	Exit	Race1
	Even Views	24,28, 32,36	0.0/0.0/0.0			
	Avg.					
	BDBR(%)/BDPSNR(dB)					
	Odd Views	24	0.08/-0.01/48.59	1.17/-0.02/57.22	1.01/-0.01/56.29	0.33/-0.02/56.58
		28	0.04/-0.01/53.83	1.07/-0.03/55.86	0.80/-0.02/56.62	0.00/-0.03/56.08
		32	-0.05/-0.01/55.98	1.14/-0.03/53.92	0.89/-0.01/55.67	0.54/-0.03/53.82
		36	0.15/-0.01/55.16	0.95/-0.02/51.49	0.42/-0.02/54.83	0.79/-0.06/50.93
	Avg.		0.05/-0.01/53.39	1.08/-0.02/54.62	0.78/-0.02/55.85	0.42/-0.03/54.35
	BDBR(%)/BDPSNR(dB)		0.33/-0.01	1.79/-0.07	0.31/-0.01	1.10/-0.05
ShenTB09 Scheme [15] VS Original JMVC			$\Delta R_{ShenTB09}$ (UNIT:%) / $\Delta PSNR_{ShenTB09}$ (UNIT:dB) / $\Delta T_{ShenTB09}$ (UNIT:%)			
		bQP	Ballet	Breakdancers	Doorflowers	Lovebird1
	Even Views	24,28, 32,36	0.0/0.0/0.0			
	Avg.					
	BDBR(%)/BDPSNR(dB)					
	Odd Views	24	-0.02/-0.02/60.94	-0.12/-0.01/23.19	-0.06/-0.03/50.56	-0.29/-0.04/76.51
		28	0.18/-0.02/64.03	0.15/-0.02/29.77	0.41/-0.03/55.21	-0.38/-0.02/78.16
		32	0.30/-0.03/56.93	0.35/-0.04/35.72	0.57/-0.04/57.41	-0.29/-0.02/77.33
		36	0.23/-0.05/68.02	0.68/-0.07/42.02	0.65/-0.06/61.27	-0.20/-0.01/77.13
	Avg.		0.17/-0.03/62.48	0.27/-0.04/32.68	0.39/-0.04/56.11	-0.29/-0.02/77.28
	BDBR(%)/BDPSNR(dB)		1.15/-0.03	1.79/-0.04	1.94/-0.05	0.25/-0.01
			$\Delta R_{ShenTB09}$ (UNIT:%) / $\Delta PSNR_{ShenTB09}$ (UNIT:dB) / $\Delta T_{ShenTB09}$ (UNIT:%)			
		bQP	Dog	Ballroom	Exit	Race1
	Even Views	24,28, 32,36	0.0/0.0/0.0			
	Avg.					
	BDBR(%)/BDPSNR(dB)					
	Odd Views	24	-0.58/-0.03/56.42	0.35/-0.02/46.68	-0.20/-0.03/71.98	-0.23/-0.02/3.50
		28	-0.39/-0.03/57.91	0.62/-0.02/49.98	0.55/-0.03/68.51	-0.19/-0.03/4.19
		32	-0.29/-0.04/58.49	0.71/-0.03/50.86	1.28/-0.04/68.81	0.01/-0.03/5.74
		36	0.05/-0.04/58.87	0.90/-0.03/44.73	0.84/-0.04/70.19	0.70/-0.09/6.46
	Avg.		-0.30/-0.03/57.92	0.65/-0.03/48.06	0.62/-0.03/69.87	0.07/-0.04/4.97
	BDBR(%)/BDPSNR(dB)		0.50/-0.02	1.34/-0.05	1.69/-0.05	0.85/-0.04

overall FMD algorithm. Zhu09 and ShenTB09 schemes adopted both FMD and reference frame selection optimization, thus, they are compared with the proposed overall FMD algorithm for fair comparison. We can observe that Zhu09 scheme is only proposed for optimizing odd views and reduces total encoding time from 48.10% to 57.80%, 54.19% on average, for these odd views; meanwhile, the average bit rate increase within 0.05% to 1.08% and the average PSNR degradation is from 0.01 dB to 0.03 dB. The BDPSNR between Zhu09 and original JMVC is from -0.01 to -0.07. The RD performance degradation for the Zhu09 scheme is negligible, but the complexity reduction is limited for the odd views and no complexity reduction could be achieved for even views. ShenTB09 scheme reduces computational complexity from 4.97% to 77.28%, 51.17% on average, for odd views, meanwhile, the bit rate increases from -0.25% to 0.7% and PSNR degradation is from 0.02 to 0.04 dB. The BDPSNR between ShenTB09 and original JMVC is from -0.01 to -0.05, -0.04 on average. However, for Race1 sequence, only 4.97% computational complexity can be reduced. It is because globally motion, unequal vertical/horizontal displacement and focus mismatch [29] of Race1 make early termination not so efficient. Additionally, the fixed threshold in

			ΔR_{ALL} (Unit:%) / $\Delta PSNR_{ALL}$ (Unit:dB) / ΔT_{ALL} (Unit:%)			
		bQP	Ballet	Breakdancers	Doorflowers	Lovebird1
Proposed overall FMD algorithm VS Original JMVC	Even Views	24	-0.80/-0.07/76.66	-0.22/-0.10/68.24	-0.81/-0.14/82.24	0.04/-0.02/68.55
		28	-0.55/-0.07/79.08	-0.71/-0.11/72.04	-0.80/-0.12/84.90	-0.23/-0.03/78.61
		32	-0.24/-0.05/80.84	-0.95/-0.10/74.92	-0.47/-0.09/85.42	-0.47/-0.04/84.21
		36	-0.13/-0.04/82.13	-0.36/-0.07/77.23	-0.13/-0.06/85.47	-0.49/-0.02/85.71
	Avg.		-0.43/-0.06/79.68	-0.56/-0.10/73.10	-0.55/-0.10/84.51	-0.29/-0.03/79.27
	BDBR(%)/BDPSNR(dB)		1.50/-0.04	3.73/-0.09	3.02/-0.09	0.49/-0.02
	Odd Views	24	-0.61/-0.09/87.27	-0.01/-0.10/77.84	-1.43/-0.17/89.91	0.34/-0.04/84.66
		28	-0.42/-0.07/87.27	-0.39/-0.11/79.43	-0.58/-0.13/89.74	0.14/-0.03/85.62
		32	-0.29/-0.05/86.76	-0.75/-0.09/80.26	-0.16/-0.10/89.12	-0.11/-0.03/87.90
		36	0.01/-0.04/86.07	-0.67/-0.08/80.76	-0.12/-0.07/88.07	-0.08/-0.03/87.80
	Avg.		-0.33/-0.06/86.84	-0.45/-0.09/79.57	-0.57/-0.11/89.21	0.07/-0.03/86.49
	BDBR(%)/BDPSNR(dB)		1.80/-0.05	4.11/-0.08	3.55/-0.10	0.89/-0.03
			ΔR_{ALL} (Unit:%) / $\Delta PSNR_{ALL}$ (Unit:dB) / ΔT_{ALL} (Unit:%)			
		bQP	Dog	Ballroom	Exit	Race1
	Even Views	24	-0.22/-0.05/72.31	-0.14/-0.09/67.36	-0.80/-0.07/71.97	-0.15/-0.13/77.73
		28	-0.41/-0.06/78.58	0.31/-0.06/70.51	-0.23/-0.05/76.51	-0.15/-0.12/79.49
		32	-0.76/-0.09/82.96	0.34/-0.06/73.33	0.02/-0.05/78.77	-0.03/-0.12/80.54
		36	-0.32/-0.07/84.29	0.51/-0.08/76.30	0.13/-0.05/80.12	-0.47/-0.11/81.18
	Avg.		-0.43/-0.07/79.53	0.26/-0.07/71.87	-0.22/-0.05/76.84	-0.20/-0.12/79.73
	BDBR(%)/BDPSNR(dB)		1.44/-0.06	1.94/-0.08	1.81/-0.05	3.14/-0.12
	Odd Views	24	-0.10/-0.05/78.99	1.35/-0.09/82.60	0.21/-0.08/83.53	0.49/-0.19/82.74
		28	-0.23/-0.06/81.52	1.20/-0.08/83.35	0.32/-0.06/85.86	-0.65/-0.15/82.57
		32	-0.40/-0.07/83.82	1.62/-0.07/83.69	0.65/-0.04/86.77	0.22/-0.13/82.01
		36	-0.20/-0.07/84.34	1.57/-0.09/83.77	0.68/-0.05/86.85	-1.23/-0.13/81.53
	Avg.		-0.23/-0.06/82.17	1.43/-0.08/83.35	0.47/-0.06/85.75	-0.29/-0.15/82.21
	BDBR(%)/BDPSNR(dB)		1.50/-0.05	3.42/-0.14	2.51/-0.07	2.70/-0.13

the ShenTB09 scheme is strict which causes large false reject rate. Similar to Zhu09 scheme, ShenTB09 can only be applied to odd views. In addition, its complexity reduction varies with sequences.

As for the proposed overall FMD algorithm, the average bit rate increases -0.30% for even views and 0.01% for odd views for all eight test multiview video sequences. Meanwhile, the PSNR degradation is 0.07 dB on average for even views and 0.08 dB on average for odd views, respectively. Also, the average BDPSNR is -0.07 dB for even views and -0.08 dB for odd views. Generally, the proposed FMD retains almost the same RD performance as Zhu09 scheme and ShenTB09 scheme. As far as the computational complexity is concerned, the proposed FMD can reduce total encoding time from 71.87% to 84.51% and 78.07% on average for even views for the eight test sequences; meanwhile, it achieves 79.57% to 89.21% and 84.45% on average complexity reduction for odd views, which is 22.23% to 45.02% more than the average complexity reduction achieved by Zhu09 and ShenTB09 schemes.

For better observation, Fig. 9 shows the encoding time saving ratio achieved by Zhu09, ShenTB09 and the proposed FMD algorithm while comparing with original JMVC. In summary, from the above analyses, we obtain the following three facts, 1) the proposed overall FMD can achieve 6% more complexity reduction for odd views than even views because these odd views maintain more coding time originally. 2) Additionally, the overall results indicate the proposed FMD retains reliable complexity reduction, from 71.87% to 89.21%, for all test sequences and all views, even with various video contents, camera arrangements and motion properties. 3) The proposed FMD reduces computational complexity 22.23% to 45.02% more than Zhu09 and ShenTB09 schemes achieve for odd (inter-view) views while maintaining almost the same RD performance, meanwhile, it reduces 78.07% more for even (temporal) views.

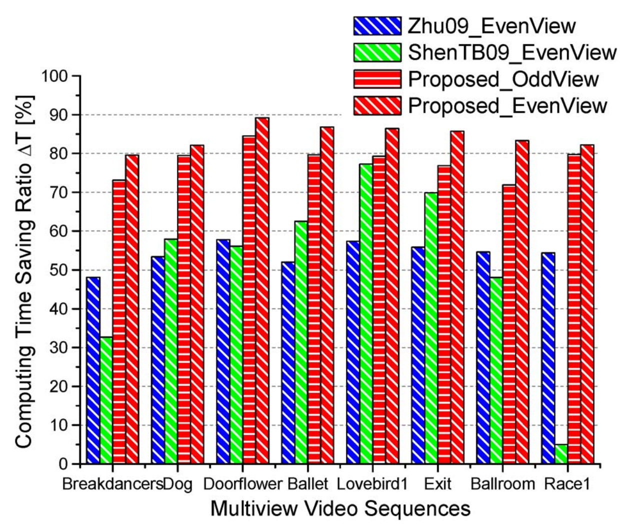


Fig. 9. Encoding time saving ratio achieved by Zhu09, ShenTB09, and the proposed FMD algorithm.

VI. Conclusions

This paper proposes efficient Statistical DIRECT Mode Early Termination (SDMET) model which estimates the RD degradation, speedup ratio, false acceptance rate and false reject rate of early DIRECT mode decision. It can adaptively adjust the RD cost threshold not only according to with quantization parameter, but also according to the video contents and motion properties. Experimental results show that SDMET can reduce 42.40%

to 65.60% computation complexity for early mode decision. When it is jointly optimized with fast multi-reference frame selection, the overall FMD algorithm can achieve 79.57% to 89.21% computational complexity reduction with unnoticeable RD degradation. Compared with the state-of-the-art FMD algorithms, the proposed FMD algorithm reduces computational complexity 22.23% to 45.02% more than Zhu09 and ShenTB09 schemes do. Additionally, it can be applied to both temporal views and inter-view views.

REFERENCES

[1] M. Tanimoto, "Overview of FTV (free-viewpoint television)," in *Proc. IEEE ICME'09*, Cancun, Mexico, Jul. 2009, pp. 1552–1553.

[2] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3DTV," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 191–199, Jun. 2005.

[3] P. Merkle, A. Smolic, K. Müller, and T. Wiegand, "Efficient prediction structures for multi-view video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1461–1473, Nov. 2007.

[4] Y. Luo, Z. Zhang, and P. An, "Stereo video coding based on frame estimation and interpolation," *IEEE Trans. Broadcast.*, vol. 49, no. 1, pp. 14–21, Mar. 2003.

[5] P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Comparative study of MVC prediction structures," in *JVT of ISO/IEC MPEG & ITU-T VCEG*, Marrakech, Morocco, Jan. 2007, Doc. JVT-V132.

[6] Y. Zhang, G. Jiang, M. Yu, and Y. S. Ho, "Adaptive multiview video coding scheme based on spatio-temporal correlation analyses," *ETRI Journal*, vol. 31, no. 2, pp. 151–161, Apr. 2009.

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

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

[10] X. Jing and L. P. Chau, "Fast approach for H.264 inter mode decision," *Electron. Lett.*, vol. 40, no. 17, pp. 1050–1051, Aug. 2004.

[11] 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.

[12] W. Zhu, W. Jiang, and Y. Chen, "A fast inter mode decision for multiview video coding," in *Proc. ICIECS'09*, Qiangdao, China, Dec. 2009, pp. 1–4.

[13] Z. Peng, G. Jiang, and M. Yu, "A fast multiview video coding algorithm based dynamic multi-threshold," in *Proc. IEEE ICME'09*, Cancun, Mexico, Jul. 2009, pp. 113–116.

[14] 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.

[15] 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.

[16] H. S. Koo, Y. J. Jeon, and B. M. Jeon, "MVC motion skip mode," in *JVT of ISO/IEC MPEG & ITU-T VCEG*, San Jose, California, USA, Apr. 2007, Doc.JVT-W081.

[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] D. H. Han and Y. L. Lee, "Fast mode decision using global disparity vector for multiview video coding," in *in Proc. FGCNS'08*, 2008, vol. 3, pp. 209–213.

[19] M. Yu, Z. Peng, W. Liu, F. Shao, G. Jiang, and Y. D. Kim, "Fast macroblock selection algorithm for multiview video coding based on interview global disparity," in *Proc. CISP'08*, Sanya, China, May 2008, pp. 575–578.

[20] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," in *Proc. ACM SIGGRAPH Trans. Graphics*, Los Angeles, USA, Aug. 2004, pp. 600–608.

[21] A. Vetro, M. McGuire, W. Matusik, A. Behrens, J. Lee, and H. Pfister, "Multiview video test sequences from MERL for the MPEG multiview working group," in *ISO/IEC JTC1/SC29/WG11*, Busan, Korea, Apr. 2005, Doc. M12077.

[22] I. Feldmann, M. Mueller, F. Zilly, R. Tanger, K. Mueller, A. Smolic, P. Kauff, and T. Wiegand, "HHI test material for 3D video," in *ISO/IEC JTC1/SC29/WG11*, Archamps, France, Apr. 2008, Doc. M15413.

[23] K. Takagi, Y. Takishima, and Y. Nakajima, "A study on rate distortion optimization scheme for JVT coder," in *Proc. VCIP'03*, 2003, vol. 5150, pp. 914–923.

[24] 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. 15–23, Mar. 2011.

[25] M. Tanimoto, T. Fujii, and N. Fukushima, "1D parallel test sequences for MPEG-FTV," in *ISO/IEC JTC1/SC29/WG11*, Archamps, France, Apr. 2008, Doc.M15378.

[26] R. Kawada, "KDDI multiview video sequences for MPEG 3DAV use," in *ISO/IEC JTC1/SC29/WG11*, Munich, Mar. 2004, Doc. M10533.

[27] G. M. Um, G. Bang, N. Hur, J. Kim, and Y. S. Ho, "3D Video. Test material of outdoor scene," in *ISO/IEC JTC1/SC29/WG11*, Archamps, France, Apr. 2008, Doc. M15371.

[28] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in *ITU-T VCEG, VCEG-M33*, April 2001.

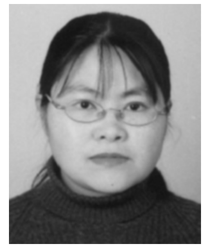
[29] Y. S. Ho, K. J. Oh, C. Lee, B. Choi, and J. H. Park, "Observations of multi-view test sequences," in *JVT of ISO/IEC MPEG & ITU-T VCEG*, San Jose, California, USA, Apr. 2007, Doc.JVT-W084.

Yun Zhang received the B.S. and M.S. degrees in electrical engineering from Ningbo University, Ningbo, China, in 2004 and 2007, respectively, and the Ph.D. degree in computer science from Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, China, in 2010. From 2009 to 2011, he was a Post-doc Researcher with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. In 2010, he joined in the Shenzhen Institutes of Advanced Technology (SIAT), CAS, as an Assistant Professor. His research interests are multiview video coding, video object segmentation and content based video processing.

Sam Kwong (M'93–SM'04) received the B.S. and M.S. degrees in electrical engineering from the State University of New York at Buffalo in 1983, the University of Waterloo, Waterloo, ON, Canada, in 1985, and the Ph.D. degree from the University of Hagen, Germany, in 1996. From 1985 to 1987, he was a Diagnostic Engineer with Control Data Canada. He joined Bell Northern Research Canada as a Member of Scientific Staff. In 1990, he became a Lecturer in the Department of Electronic Engineering, City University of Hong Kong, where he is currently a Professor in the Department of Computer Science. His research interests are video and image coding and evolutionary algorithms.

Gangyi Jiang received his M.S. degree from Hangzhou University, Hangzhou, China, in 1992, and received the Ph.D. degree from Ajou University, Korea, in 2000. In 2000, he joined the Faculty of Information Science and Engineering, Ningbo University, China, as a Professor. His research interests include digital video compression and communications, multi-view video coding and image processing.

Xu Wang received the B.S. degree from South China Normal University, Guangzhou, China, in 2007, and M.S. degree from Ningbo University, Ningbo, China in 2010. He is now working toward the Ph.D. degree from the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. His current research interests include stereoscopic image/video quality assessment, bit allocation and rate control for 3D video.



Mei Yu received the M.S. degree from Hangzhou Institute of Electronics Engineering, Hangzhou, China, in1993, and the Ph.D. form Ajou University, Korea, in 2000. Then, she joined in Ningbo University, Ningbo, China. Since 2005, she works as a Professor in Faculty of Information Science and Engineering, Ningbo University, China. Her research interests include image/video coding and video perception.