Fast Inter Mode Decision Based on Textural Segmentation and Correlations for Multiview Video Coding

Wei Zhu, Xiang Tian, Fan Zhou and Yaowu Chen

Abstract —*Multiview video coding (MVC) plays a critical role in reducing the ultra high data bandwidth of 3-D video. However, the heavy computational complexity of MVC slows down its applications. In this paper, a fast Inter mode decision based on textural segmentation and correlations is presented: Firstly, each frame is segmented into three textural regions based on the correlation between textural complexities and Intra mode Rate-Distortion (RD) costs, and the correlation of Intra mode RD costs between views. Then, by using textural region types and the correlation of Skip mode RD costs between views, an early decision of Skip mode is introduced. Thirdly, by utilizing Inter16×16 mode prediction results and textural region types, the disparity estimation of other Inter modes is selected. Finally, the estimation of Inter8×8 mode can be reduced according to textural region types and motion activity. As compared to the full mode decision in MVC reference software, experimental results show that the proposed algorithm can reduce 55~70% of the computational time, while maintaining high RD performance*[1].

Index Terms —Multiview video coding, 3-D video, textural segmentation, mode decision, disparity estimation.

I. INTRODUCTION

With the advances in display and camera technology, many new applications for 3-D scene communication have emerged such as 3D-TV [1] and free viewpoint TV (FTV) [2]. Multiview video which is widely used as the signal for 3-D applications is a group of views captured by a set of cameras from different positions on the same scene. However, with the increasing number of views, it consumes an ultra large amount of video data bandwidth. For efficient storage and transmission of the huge video data, Multiview Video Coding (MVC) tries to compress multiple video data efficiently, and the Joint Video Team (JVT) has developed a Joint Multiview Video Model (JMVM) [3] which is based on H.264/AVC. Since views are captured on the same scene at the same time, multiview video data have both temporal redundancy and inter-view redundancy. MVC employs both the temporal prediction in conventional video coding and the inter-view prediction among views to improve coding efficiency [4], [5]. Fig.1 shows an illustration of prediction structure employed in JMVM; it uses a hierarchical B prediction structure [6] for each view. All views can be classified into two categories; primary views (such as S0 and S2) which mainly employ temporal prediction, and auxiliary views (such as S1) which can be referred to the neighboring primary views.

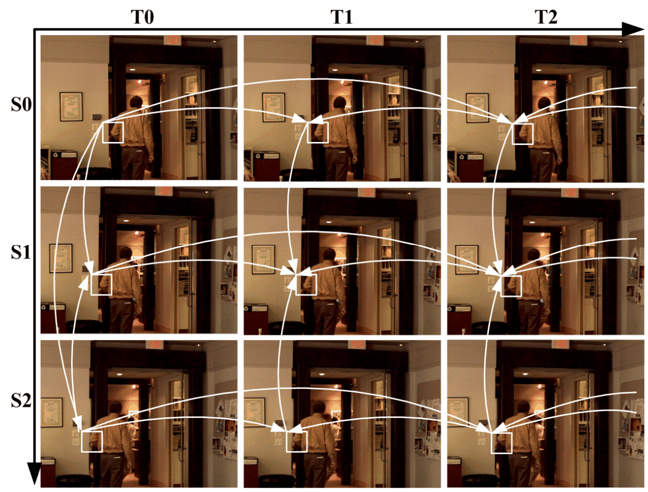

Fig. 1. Illustration of prediction structure for MVC.

MVC uses Rate-Distortion Optimization (RDO) technique [7] to select the optimum coding mode. For each MB, the MVC encoder calculates Rate-Distortion (RD) costs of all MB modes and selects the optimum mode having the minimum RD cost. Because the process of RDO is repeatedly carried out for each MB, the computational complexity of mode decision is enormous.

Some fast mode decision algorithms for H.264/AVC can be adopted to speed up the mode decision of MVC. Pan *et al* proposed a fast Intra mode decision based on the local edge information [8]. Wu *et al* presented a fast Inter mode decision by using of the spatial homogeneity and the temporal stationarity characteristics of video objects [9]. You *et al* presented a fast Inter mode decision for P slices by utilizing the results of P16×16 motion estimation [10]. However, the coding structures in MVC are more complex than that in H.264/AVC, especially for auxiliary views employing inter-view prediction. The non-anchor frames in auxiliary views (such as S1T1 shown in Fig.1) need to perform motion estimation for temporal prediction and disparity estimation for inter-view prediction, while the non-anchor frames in primary views (such as S0T1) only need to perform motion estimation. Thus, the prediction directions could be decided to reduce the computational complexity of auxiliary views. In addition, as compared to H.264/AVC, the correlations between views can be exploited to further speed up the mode decision of MVC.

Recently, several fast mode decision algorithms have been proposed for MVC. Shen *et al* proposed a selective disparity

[1] W. Zhu, X. Tian, F. Zhou and Y. Chen are with the Institute of Advanced Digital Technology and Instrument, Zhejiang University, Hangzhou 310027 China (e-mail: tianx@mail.bme.zju.edu.cn)
Contributed Paper
Manuscript received 06/21/10
Current version published 09/23/10
Electronic version published 09/30/10.

estimation and variable size motion estimation based on the motion homogeneity [11]. Shen *et al* later proposed a early Skip mode decision by using the inter-view correlation [12]. An inter-view mode decision was presented in [13] to choose the most probable coding mode by exploiting the correlation of coding information between views. An early termination scheme was proposed in [14] to skip the motion estimation or disparity estimation. In [15], disparity estimation was decided by using the minimum sum of absolute differences of motion estimation. In [16], motion estimation and disparity estimation were decided according to the prediction results of collocated MBs in temporal reference frames. And in [17], the motion estimation and disparity estimation were selected by using the segmentation results of foreground and background objects. By using above mode decision algorithms, the computational complexity can be reduced effectively. However, correlations between views, correlations among Inter modes, and textural complexities can be further exploited to reduce the complexity, while maintaining the coding efficiency. In this paper, a fast Inter mode decision based on the textural segmentation and correlations is proposed: Considering that RD costs are sensitive to textural complexities, each frame is segmented into three textural regions by using Intra mode RD costs and segmentation thresholds. Then, segmentation results are used in subsequent Inter mode decision to save the computation time while maintaining picture quality. And Skip mode RD costs classified by Intra mode are used to exploit the RD cost correlation between views instead of using disparity vectors; Skip mode is early decided by comparing its RD cost with the RD cost threshold which is calculated according to different textural regions. For the optimization of the prediction direction, the correlation of inter-view prediction among Inter modes is employed to select the disparity estimation, and the selection frequency is adjusted in different textural regions accordingly. In addition, the correlation between Inter8×8 mode and textural regions is combined with the motion activity to reduce the estimation of Inter8×8 mode.

The rest of paper is organized as follows: Section II describes the observation and analysis of modes' features and the correlation between textural complexities and RD costs. Section III presents our fast algorithm. Experimental results and conclusions are given in Section IV and V respectively.

II. Observation and Analysis

The MVC reference software JMVC 4.0 [18] with the full mode decision was used to analyze modes' features and the correlation between textural complexities and RD costs. MB modes are classified into Intra mode and Inter mode according to the prediction algorithm. Intra mode uses the boundary pixels of reconstructed neighboring MBs in the current frame to predict the current MB. Inter mode uses reconstructed pixels of previous encoded frames to predict. In our analysis of modes' features, Inter mode is further classified into following five categories; Skip, Inter16×16, Inter16×8, Inter8×16, and Inter8×8 which includes four sub-modes [19].

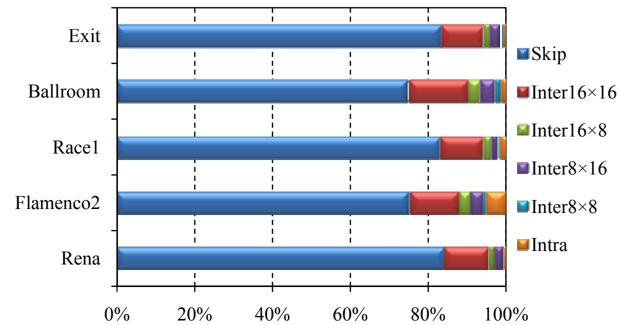

Fig. 2. Proportion of MB modes.

Skip mode utilizes the motion vector predictor to predict the current MB, and it needn't to perform inter estimation. Other Inter modes need to perform inter estimation by referring to either temporal or inter-view reference frames. For auxiliary view S1, proportions of MB modes for different sequences are shown in Fig.2. It can be seen that Skip mode occupies the largest proportion among all MB modes. This is because that there is a large proportion of background for each sequence, and Skip mode is more likely to be selected as the optimum mode in the background. Because of no inter estimation, the calculation time of Skip mode is much faster than other Inter modes. Thus, if Skip mode could be determined to be the optimum mode as earlier as possible, it would avoid the estimation operation of other modes which consumes most of the encoding time. In Fig.2, it can also be found that Inter8×8 mode occupies the smallest proportion among Inter modes. However, from our statistical calculation, the estimation of Inter8×8 mode consumes about 50% percent of the encoding time which is the largest among Inter modes. Thus, it is important to reduce the estimation of Inter8×8 mode. From the aspect of prediction direction, Inter mode also can be classified into Inter_ME, Inter_DE, and Inter_ME_DE. Inter_ME mode only employs temporal prediction by using motion estimation, Inter_DE mode only employs inter-view prediction by using disparity estimation, and Inter_ME_DE mode employs both temporal prediction and inter-view prediction. From our statistics, Inter_DE mode and Inter_ME_DE mode only occupy a small proportion, while the computation time consumed by disparity estimation for inter-view prediction occupies more than 50% of the encoding time. Therefore, if we could reduce disparity estimation, the encoding time would be effectively reduced. Based on the analysis above, the early decision of Skip mode, the selection of disparity estimation, and the reduction of Inter8×8 mode estimation will be our main focuses.

In Fig.2, Intra mode occupies a small proportion among all MB modes, and the computation time of Intra mode decision is also small from our statistics. Thus, Intra mode decision is not our focus, and it is implemented before Inter mode decision. Then the results of Intra mode can be used for the optimization of subsequent Inter mode decision. To exploit the correlation between Intra mode and Inter mode, and the correlation between textural complexities and RD costs,

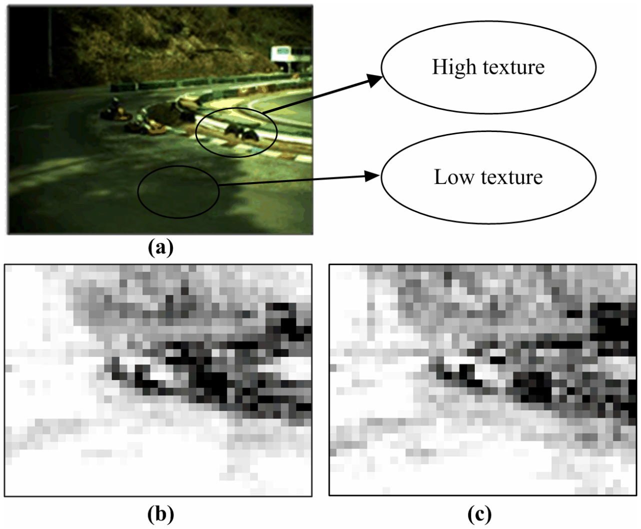

Fig. 3. Distribution of normalized RD costs in "Race1" sequence. (a) Original frame. (b) Normalized Intra mode RD costs. (c) Normalized Inter mode RD costs.

distributions of RD costs were analyzed. For the frame S1T8 in "Race1" sequence, all textural regions have almost the same motion activity because of the global motion of camera. Distributions of normalized Intra mode RD costs and Inter mode RD costs are illustrated in Fig.3: Blocks with higher values of RD costs are darker and blocks with lower values of RD costs are lighter. It can be seen that there is a similar distribution of normalized RD costs between Intra mode and Inter mode. Since both Intra mode and Inter mode have higher RD costs in regions with higher textural complexity, and both Intra mode and Inter mode have lower RD costs in regions with lower textural complexity, it demonstrates that both Intra and Inter mode RD costs are sensitive to textural complexities. Because Inter mode RD costs have different sensitivities to different textural regions, textural complexities should be taken into consideration for reducing the computation time while maintaining the picture quality. And Intra mode RD costs can be used for the segmentation of textual regions.

III. Proposed Fast Inter Mode Decision

According to the analysis in the previous section, a fast Inter mode decision based on the textural segmentation and correlations is proposed. The system flow diagram of MVC encoder with the proposed algorithm is shown in Fig.4. After encoding of the frames in neighboring reference views, information of RD costs are transferred to the frame in current view. The proposed algorithm is circled with the red dotted line. Based on the correlation between textural complexities and Intra mode RD costs and the correlation of Intra mode RD costs between views, current frame is segmented into three textural regions. Then, the estimation of Intra, Skip, and Inter16×16 modes are performed before other modes. The purpose of estimating Intra mode is to get its RD cost which is used to obtain the textural region type of MB. The purpose of estimating Skip mode is to get its RD cost for the early decision of Skip mode. The purpose of estimating Inter16×16 mode is to get results of motion estimation and disparity estimation, and these results will be used to select the disparity estimation and reduce the estimation of Inter8×8 mode. Based on textural region types and correlations, detailed procedures are introduced respectively. Focuses are being placed on the early decision of Skip mode, the selection of disparity estimation, and the reduction of Inter8×8 mode estimation.

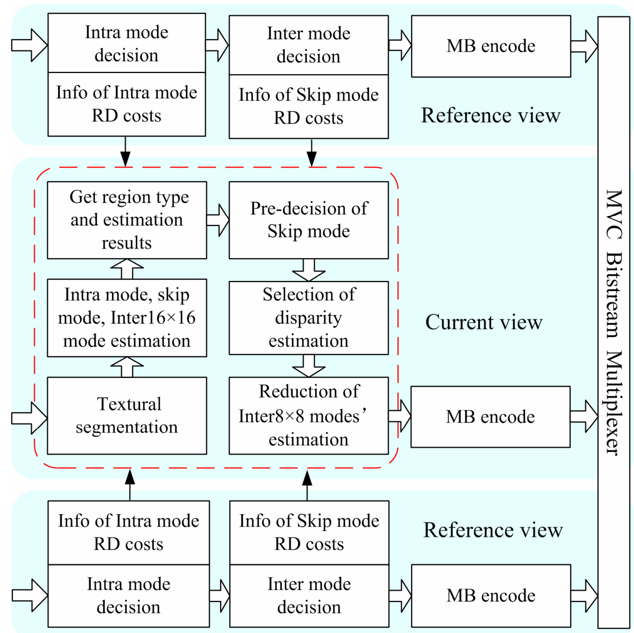

Fig. 4. System flow diagram of MVC encoder with the proposed algorithm.

A. Segmentation of Textural Regions

Based on the correlation between textural complexities and Intra mode RD costs which have been analyzed in Section II, Intra mode RD costs are used to segment textural regions. From the point of H.264/AVC codec [19], the area with high textural features often selects Intra4×4 mode, and the homogeneous area which contains low textural features often selects Intra16×16 mode. Thus, the Intra mode obtained from Intra mode decision is classified into INTRA4 and INTRA16 in our approach. INTRA4 mode which is more likely to appear in high texture regions includes Intra4×4, Intra8×8, and PCM modes, and INTRA16 mode which is more likely to appear in low texture regions includes Intra16×16 mode. To analyze RD cost distributions, RD costs of INTRA4 mode and INTRA16 mode for frames S0T4, S1T4, and S2T4 in "Race1" sequence are sorted respectively from low to high, and curves of RD costs are shown in Fig.5. It can be seen that INTRA16 mode RD costs are in a relative low level, while INTRA4 mode RD costs are in a relative high level, and the variation of RD costs with the same Intra mode among views are consistent. Thus, the distribution of all Intra mode RD costs can be appropriately divided into three segments; the low, the high, and the in-between values by utilizing Intra mode RD costs of frames in neighboring reference views.

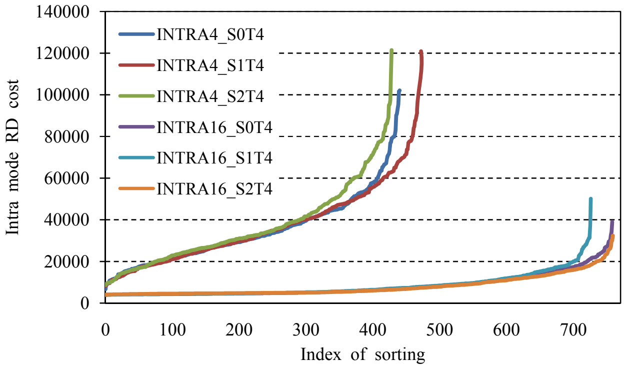

Fig. 5. Sorting of INTRA4 mode RD costs and INTRA16 mode RD costs for frames S0T4, S1T4, and S2T4 in "Race1" sequence.

Medians of INTRA16 mode RD costs calculated in (1) and medians of INTRA4 mode RD costs calculated in (2) are used to calculate thresholds of the textural segmentation.

$$MedianCost_{\text{INTRA16}}(k) = \text{median}\{Cost_{\text{INTRA16}}(n) \mid n = 1,2,\cdots,N_{\text{INTRA16}}\} \quad (1)$$

$$MedianCost_{\text{INTRA4}}(k) = \text{median}\{Cost_{\text{INTRA4}}(n) \mid n = 1,2,\cdots,N_{\text{INTRA4}}\} \quad (2)$$

where k is the index of neighboring reference views, $Cost_{\text{INTRA16}}(n)$ is INTRA16 mode RD cost of *nth* MB with INTRA16 as the Intra mode, N_{INTRA16} is the number of $Cost_{\text{INTRA16}}$, $Cost_{\text{INTRA4}}(n)$ is INTRA4 mode RD cost of *nth* MB with INTRA4 as the Intra mode, and N_{INTRA4} is the number of $Cost_{\text{INTRA4}}$. After encoding of the frames in neighboring reference views, the medians of RD costs are transferred to the frame in current view. Then, the segmentation threshold of low textural region ($SegTH_{\text{L}}$) is calculated in (3) by averaging the medians of INTRA16 mode RD costs obtained from neighboring reference views, and the segmentation threshold of high textural region ($SegTH_{\text{H}}$) is calculated in (4) by averaging the medians of INTRA4 mode RD costs.

$$SegTH_{\text{L}} = \frac{1}{N}\sum_{k\in \text{R}} MedianCost_{\text{INTRA16}}(k) \quad (3)$$

$$SegTH_{\text{H}} = \frac{1}{N}\sum_{k\in \text{R}} MedianCost_{\text{INTRA4}}(k) \quad (4)$$

where R is a set of neighboring reference views' index, and N is the number of neighboring reference views. For each frame, it needs to calculate the thresholds once at beginning, and the frame is segmented into three textural regions by using these two thresholds. For each MB of the frame, its textural region type is obtained in (5) by comparing the Intra mode RD cost with segmentation thresholds.

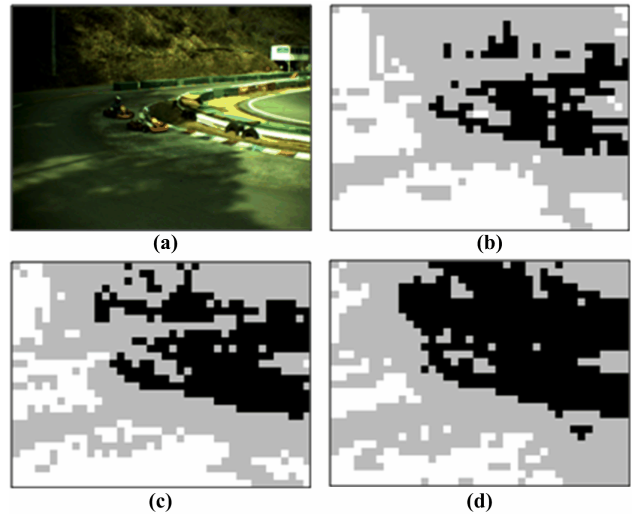

Fig. 6. Distribution of textural regions in "Race1" sequence. (a) Original frame. (b) With QP=37. (c) With QP=32. (d) With QP=22.

$$Region(n) = \begin{cases} \text{LOW}, & \text{if } Cost_{\text{INTRA}} <= SegTH_{\text{L}} \\ \text{HIGH}, & \text{if } Cost_{\text{INTRA}} >= SegTH_{\text{H}} \\ \text{MEDIUM}, & \text{others} \end{cases} \quad (5)$$

where n is the index of MB, and $Cost_{\text{INTRA}}$ is Intra mode RD cost obtained from Intra mode decision. MBs with $Cost_{\text{INTRA}}$ less than or equal to $SegTH_{\text{L}}$ belong to the low textural region (LOW), MBs with $Cost_{\text{INTRA}}$ larger than or equal to $SegTH_{\text{H}}$ belong to the high textural region (HIGH), and MBs with $Cost_{\text{INTRA}}$ between $SegTH_{\text{L}}$ and $SegTH_{\text{H}}$ belong to the medium textural region (MEDIUM).

A distribution of textural regions in "Race1" sequence is shown in Fig.6. The high textural region is in black color; the medium textural region is in grey color; and the low textural region is in white color. It can be seen that regions are segmented approximately based on textural complexity, and distributions of textural regions are varying with different basis QPs setting; area of the high textural region with the lower QP is larger than with the higher QP, while area of the low textural region with the lower QP is smaller than with the higher QP. The calculation of RD cost is relevant to QP [7], and the textural segmentation is based on Intra mode RD costs. These demonstrate that segmentation regions are adaptive to QP. Because the textural segmentation takes the QP factor into consideration, it is more suitable for Inter mode decision.

B. Early decision of Skip Mode

As the analysis in Section II, the early decision of Skip mode is our focus. To early decide Skip mode as the optimum mode, RD cost of Skip mode is compared with the RD cost threshold. Considering that RD costs have different sensitivities to textural complexities, so the RD cost threshold should be calculated for different textural regions.

For frames S0T6, S1T6, and S2T6 which are at the time instance T6 with different views in "Exit" sequence, numbers of Skip mode selected by the full mode decision are shown in

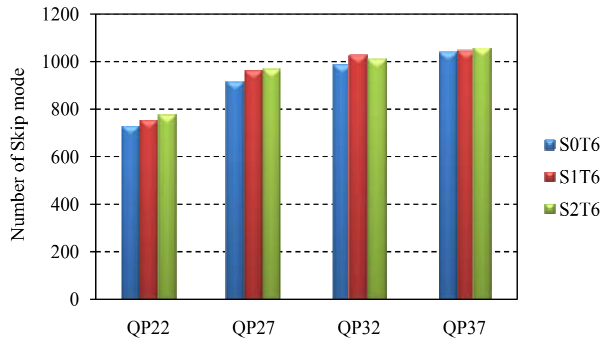

Fig. 7. Number of Skip mode for frames S0T6, S1T6, and S2T6 under different basis QPs in "Exit" sequence.

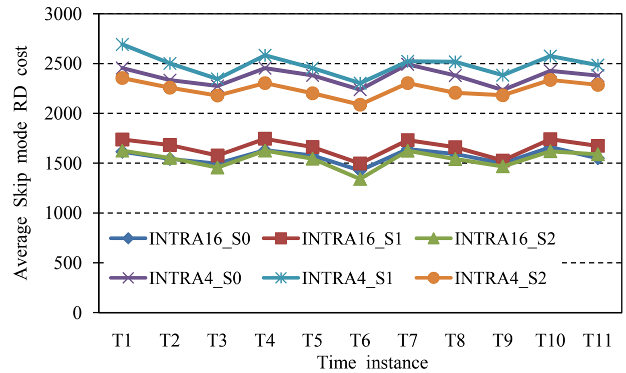

Fig. 8. Average Skip mode RD cost curves of $\boldsymbol{INTRA16\_Cost_{\mathrm{SKIP}}}$ **and** $\boldsymbol{INTRA4\_Cost_{\mathrm{SKIP}}}$ **in "Exit" sequence.**

Fig.7. It can be seen that the numbers are similar between frames with the same QP. This is because that there are similar background scenes between views, and Skip mode is likely to be selected as the optimum mode in background scenes. As the analysis in Section II, Inter mode RD costs are sensitive to textural complexities, so the statistical average of Skip mode RD costs classified by INTRA16 mode and INTRA4 mode respectively are presented as follows:

$$INTRA16\_Cost_{\mathrm{SKIP}}(k)=\frac{\sum_{n=1}^{N_{\mathrm{SKIP}}} Cost_{\mathrm{SKIP}}(n)\times B_{\mathrm{INTRA16}}(n,m)}{\sum_{n=1}^{N_{\mathrm{SKIP}}} B_{\mathrm{INTRA16}}(n,m)} \quad (6)$$

$$INTRA4\_Cost_{\mathrm{SKIP}}(k)=\frac{\sum_{n=1}^{N_{\mathrm{SKIP}}} Cost_{\mathrm{SKIP}}(n)\times B_{\mathrm{INTRA4}}(n,m)}{\sum_{n=1}^{N_{\mathrm{SKIP}}} B_{\mathrm{INTRA4}}(n,m)} \quad (7)$$

where $Cost_{\mathrm{SKIP}}(n)$ is Skip mode RD cost of *nth* MB with Skip mode as the optimum mode, N_{SKIP} is the number of $Cost_{\mathrm{SKIP}}$, m represents the Intra mode of the *nth* MB. The B_{INTRA16} and B_{INTRA4} are calculated as follows:

$$B_{\mathrm{INTRA16}}(n,m)=\begin{cases}1, & \text{if } m=\mathrm{INTRA16}\\ 0, & \text{otherwise}\end{cases} \quad (8)$$

$$B_{\mathrm{INTRA4}}(n,m)=\begin{cases}1, & \text{if } m=\mathrm{INTRA4}\\ 0, & \text{otherwise}\end{cases} \quad (9)$$

In Fig.8, $INTRA16\_Cost_{\mathrm{SKIP}}$ and $INTRA4\_Cost_{\mathrm{SKIP}}$ of view S0, S1, and S2 at different time instances in "Exit" sequence are shown. It can be seen that curves of $INTRA4\_Cost_{\mathrm{SKIP}}$ are always above curves of $INTRA16\_Cost_{\mathrm{SKIP}}$ for different views. Values of $INTRA4\_Cost_{\mathrm{SKIP}}$ are similar at the same time instance, so as values of $INTRA16\_Cost_{\mathrm{SKIP}}$.

As seen from the analysis above, there is a correlation of Skip mode RD costs between views, the $INTRA16\_Cost_{\mathrm{SKIP}}$ are in relative low level, and the $INTRA4\_Cost_{\mathrm{SKIP}}$ are in relative high level. Thus, the RD cost threshold in the low textural region ($SkipTH_{\mathrm{L}}$) is calculated in (10) by averaging $INTRA16\_Cost_{\mathrm{SKIP}}$ obtained from neighboring reference views. The RD cost threshold in the high textural region ($SkipTH_{\mathrm{H}}$) is calculated in (11) by averaging $INTRA4\_Cost_{\mathrm{SKIP}}$.

$$SkipTH_{\mathrm{L}}=\alpha\times\frac{1}{N}\sum_{k\in R} INTRA16\_Cost_{\mathrm{SKIP}}(k) \quad (10)$$

$$SkipTH_{\mathrm{H}}=\alpha\times\frac{1}{N}\sum_{k\in R} INTRA4\_Cost_{\mathrm{SKIP}}(k) \quad (11)$$

where parameter α is inserted as a tradeoff between the quality and the encoding time, and the selection of optimum α value will be analyzed in the next section.

Because the rate part of Skip mode in the RD cost formula [7] is zero, Skip mode RD cost includes only the distortion part. The value of distortion is more relevant to the textural complexity of MB, so in our proposed algorithm, $SkipTH_{\mathrm{L}}$, $SkipTH_{\mathrm{H}}$, $Cost_{\mathrm{INTRA}}$ and segmentation thresholds are used to linearly interpolate the RD cost threshold in the medium textural region ($SkipTH_{\mathrm{M}}$). The equation is shown in (12).

$$SkipTH_{\mathrm{M}}=SkipTH_{\mathrm{L}}+\frac{Cost_{\mathrm{INTRA}}-SegTH_{\mathrm{L}}}{SegTH_{\mathrm{H}}-SegTH_{\mathrm{L}}}\times(SkipTH_{\mathrm{H}}-SkipTH_{\mathrm{L}}) \quad (12)$$

The relationship between the RD cost threshold of Skip mode ($SkipTH$) and $Cost_{\mathrm{INTRA}}$ is shown in Fig.9. Thresholds in the low and high textural regions are fixed to $SkipTH_{\mathrm{L}}$ and $SkipTH_{\mathrm{H}}$ respectively, while the threshold in the medium textural region needs to be obtained dynamically. For each MB, if Skip mode has minimum RD cost among Intra, Skip, and Inter16×16 modes, and its RD cost is less than the corresponding threshold according to textural region types, then the Skip mode will be early decided as the optimum mode. In order to prevent error propagation of the prediction, the early decision of Skip mode is not used for anchor frames.

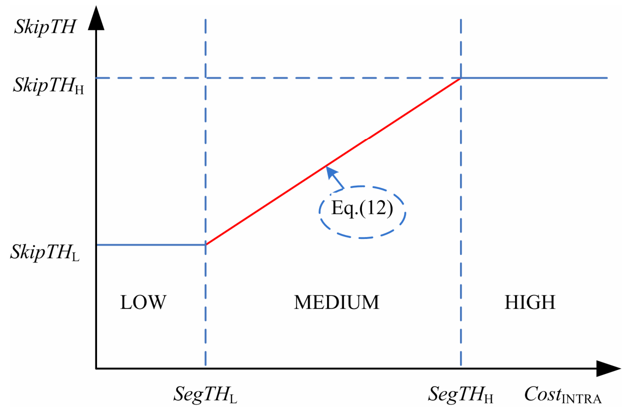

Fig. 9. The relationship between RD cost threshold of Skip mode (*SkipTH*) and Intra mode RD Cost (*Cost*$_{\text{INTRA}}$).

(a) (b)

Fig. 10. Distribution of inter-view prediction in "Exit" sequence. (a) After estimation of Inter16×16 mode. (b) After estimation of all Inter modes.

C. Selection of Disparity Estimation

For non-anchor frames in auxiliary views, there are temporal reference frames for the temporal prediction and inter-view reference frames for the inter-view prediction. Except for Skip mode, other Inter modes are required to estimate same predictions. Thus, the correlation of selected inter-view prediction between Inter16×16 mode and all Inter modes was analyzed: For the same frame, the distribution of the inter-view prediction selected after the estimation of Inter16×16 mode is shown in Fig.10 (a) as pink blocks, and so does the distribution of the inter-view prediction selected after the estimation of all Inter modes shown in Fig.10 (b). It can be seen that distributions of these two cases are similar. If Inter16×16 mode employs the inter-view prediction, other Inter modes with variable block sizes are also likely to select the same prediction. Therefore, prediction results of Inter16×16 mode can be utilized to guide the selection of disparity estimation for other Inter modes.

Because the optimum prediction is selected by comparing RD costs of motion estimation and disparity estimation, so Inter16×16 mode RD costs of motion estimation and disparity estimation are used to select the disparity estimation of other Inter modes.

For each MB, the enabling of disparity estimation (EN_{DE}) for other Inter modes is determined in (13) based on the analysis above.

$$EN_{\text{DE}}(n) = \begin{cases} 1, & \text{if } \ Cost_{\text{DE}} < \beta \times Cost_{\text{ME}} \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where $Cost_{\text{DE}}$ is the RD cost of Inter16×16 mode estimated from the forward inter-view reference frame, and $Cost_{\text{ME}}$ is the RD cost of Inter16×16 mode estimated from the forward temporal reference frame. In (13), the parameter β is a tradeoff between the quality and the encoding time, and it is set to different values in (14) according to textural region types. Since RD costs are most sensitive to the high textural region, a larger β value in the region is preferable as compared to other regions in order to increase the selection frequency of disparity estimation. As a result, a larger β value, $\beta 3$, will be selected for our simulation.

$$\beta = \begin{cases} \beta 1, & Region(n) = \text{LOW} \\ \beta 2, & Region(n) = \text{MEDIUM} \\ \beta 3, & Region(n) = \text{HIGH} \end{cases} \tag{14}$$

D. Reduction of Inter8×8 Mode Estimation

A distribution of Inter8×8 mode selected by the full mode decision is shown in Fig.11. MBs with Inter8×8 mode as the optimum mode are represented in red color blocks. It can be seen that most of Inter8×8 mode appear in regions with more textural complexity. Thus, proportions of Inter8×8 mode in different textural regions are provided in Table I. It can be seen that most of Inter8×8 mode are in the high textural region, and there are very small proportions of Inter8×8 mode in the low textural region. In Fig.11, it also can be found that there are few Inter8×8 mode in backgrounds with no motion activity, so the motion activity of MB can also be used to reduce the estimation of Inter8×8 mode.

Based on the analysis above, the estimation of Inter8×8 mode is reduced according to textural region types and motion activity. In the low textural region, the estimation of Inter8×8 mode is disabled. In the medium textural region, the estimation of Inter8×8 mode for the MBs with no motion activity is also disabled. In our algorithm, the motion activity of MB is determined as follow:

$$MotionActivity(n) = \begin{cases} 0, & \text{if } \ MV_{16\times16}(n) = (0,0) \\ 1, & \text{otherwise} \end{cases} \tag{15}$$

where $MV_{16\times16}(n)$ is the motion vector of *nth* MB, and it is selected from the forward temporal reference frame after implementing motion estimation of Inter16×16 mode. Since most of Inter8×8 mode concentrate in the high textural region and RD costs are most sensitive to the high textural region, the estimation of Inter8×8 mode in this region is not optimized in order to maintain the quality of pictures. For each MB, the enabling of Inter8×8 mode estimation ($EN_{\text{INTER}8\times8}$) is decided in (16).

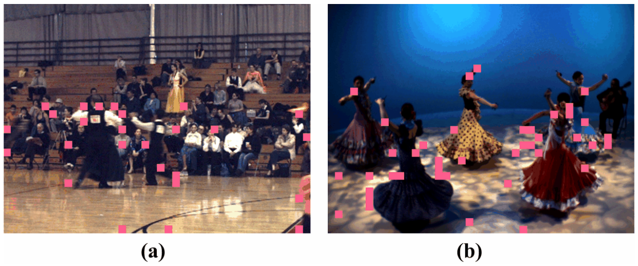

Fig. 11. Distribution of Inter8×8 mode. (a) "Ballroom" sequence. (b) "Flamenco2" sequence.

TABLE I
THE PROPORTION OF INTER8×8 MODE IN DIFFERENT TEXTURAL REGIONS

Sequences	Region type		
	LOW	MEDIUM	HIGH
Exit	0.00%	12.84%	87.16%
Ballroom	0.05%	15.58%	84.37%
Race1	0.07%	16.90%	83.03%
Flamenco2	0.00%	22.18%	77.82%
Rena	0.21%	36.44%	63.35%

$$EN_{\text{INTER}8\times8}(n)=\begin{cases}0, & \text{if } Region(n)=\text{LOW}\\ 0, & \text{if } Region(n)=\text{MEDIUM and } MotionActivity(n)=0\\ 1, & \text{others}\end{cases} \quad (16)$$

IV. Experimental Results

The proposed algorithm has been implemented in MVC reference software JMVC 4.0, and the testing configuration is based on the common test conditions in [20]. In our experiment, JMVC used the fast search mode with search range 96, and its basis QP was with 22, 27, 32, and 37. Five sequences were employed: "Exit", "Ballroom", "Race1", "Flamenco2", and "Rena". Two or three views of these sequences were chosen for experiment. The view S1 as shown in Fig.1, was used for the implementation of the proposed algorithm. The view S0 was chosen as the reference view for "Flamenco2" sequence because of the 2D-cross camera arrangement, and views S0 and S2 were chosen as reference views for other sequences. The saving of the encoding time (ΔTime) was used to evaluate the performance of complexity. The Bjontegaard delta peak signal-to-noise ratio (BDPSNR) and Bjontegaard delta bit rate (BDBR) [21] were used to evaluate the RD performance difference between algorithms.

As compared to the full mode decision in JMVC, the performance of the early decision of Skip mode, the selection of disparity estimation, and the reduction of Inter8×8 mode estimation are illustrated respectively in Table II. A negative BDPSNR or a positive BDBR indicates a coding loss to the full mode decision and is not preferred.

For the early decision of Skip mode, the value of parameter α was determined by experimental analysis: As shown in Fig.12 and Fig.13, with the increase of parameter α, the saving

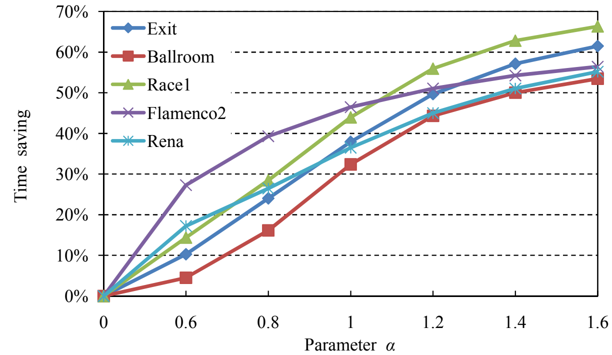

Fig. 12. The relationship between the parameter α and the saving of encoding time.

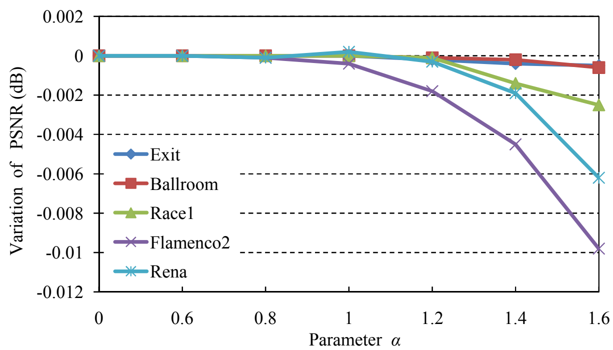

Fig. 13. The relationship between the parameter α and the variation of PSNR.

of encoding time becomes larger, whereas the peak signal-to-noise ration (PSNR) drops. Thus, 1.2 was selected for the parameter α in our experiment to have a large time saving and maintain a reasonable quality. Experimental results for different sequences are shown in Table II. It shows that an average 46.4% time saving is achieved with no drop of BDPSNR and 0.01% decrease of bit rate. From Table II, it can also be seen that the "Race1" sequence which has the largest proportion of Skip mode in Fig.2 saves more encoding time than other sequences, and the "Ballroom" sequence which has the least proportion of Skip mode has the least encoding time reduction as compared to other sequences.

For the selection of disparity estimation, values of 1.0, 1.1, and 1.2 were selected for parameter $\beta 1$, $\beta 2$, and $\beta 3$ respectively with the same methodology as selecting the value of parameter α based on the saving of encoding time and the variation of PSNR. Experimental results are also shown in Table II, and it can be seen that an average 46.7% time saving is achieved with only 0.004 dB BDPSNR drop and 0.11% bit rate increase. The "Rena" sequence saves less encoding time than other sequences. This is because that the camera spacing of "Rena" sequence is only 5cm [20] which is smaller than other sequences, and there is more disparity estimation employed for exploiting the inter-view redundancy.

The experimental reduction of Inter8×8 mode estimation is illustrated again in Table II, and it indicates that an average 25.4% reduction of encoding time is achieved with slight RD

TABLE II
PERFORMANCE OF THE PROPOSED EARLY DECISION OF SKIP MODE, SELECTION OF DISPARITY ESTIMATION, AND REDUCTION OF INTER8×8 MODE ESTIMATION AS COMPARED TO THE FULL MODE DECISION.

Sequences	Early decision of Skip mode			Selection of disparity estimation			Reduction of Inter8x8 mode estimation		
	ΔTime (%)	BDPSNR (dB)	BDBR (%)	ΔTime (%)	BDPSNR (dB)	BDBR (%)	ΔTime (%)	BDPSNR (dB)	BDBR (%)
Exit	45.2	0.000	0.00	53.9	-0.002	0.09	31.8	0.001	-0.01
Ballroom	40.8	0.000	0.00	50.9	-0.013	0.40	20.7	0.000	0.01
Race1	53.1	0.003	-0.07	56.0	0.001	-0.06	22.2	0.002	-0.07
Flamenco2	47.8	-0.002	0.04	40.7	-0.001	0.03	24.9	-0.001	0.01
Rena	45.1	0.000	0.00	31.8	-0.004	0.10	27.2	0.001	-0.03
Average	46.4	0.000	-0.01	46.7	-0.004	0.11	25.4	0.001	-0.02

TABLE III
PERFORMANCE OF OUR PROPOSED ALGORITHM AND SHEN'S ALGORITHM [11] AS COMPARED TO THE FULL MODE DECISION.

Sequences	Proposed Algorithm			Shen's Algorithm in [11]		
	ΔTime (%)	BDPSNR (dB)	BDBR (%)	ΔTime (%)	BDPSNR (dB)	BDBR (%)
Exit	66.3	-0.002	0.12	61.5	-0.024	0.99
Ballroom	60.3	-0.014	0.41	56.9	-0.034	1.00
Race1	70.3	0.001	-0.06	52.5	-0.024	0.58
Flamenco2	59.9	-0.004	0.09	47.5	-0.006	0.14
Rena	56.2	-0.006	0.12	50.1	-0.016	0.33
Average	62.6	-0.005	0.13	53.7	-0.021	0.61

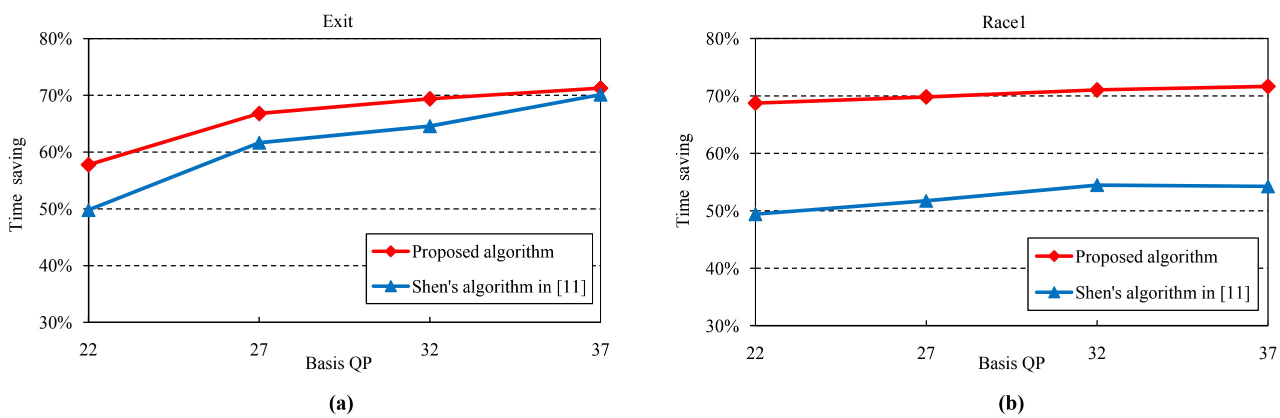

Fig. 14. Time saving curves of our proposed algorithm and Shen's algorithm in [11] under different basis QPs (22, 27, 32 and 37). (a) Time saving curves for "Exit" sequence. (b) Time saving curves for "Race1" sequence.

performance increase. Since the "Exit" sequence has a large proportion of homogenous regions with no motion activity, it achieves larger reduction of encoding time than others.

The performance of our proposed fast Inter mode decision which includes all three parts of the optimization is shown in Table III. It demonstrates that our proposed algorithm achieves about 55~70% reduction of encoding time for different sequences, and the average time saving is 62.6% with only 0.005 dB BDPSNR drop and 0.13% bit rate increase. The fast mode decision algorithm proposed by Shen *et al* in [11] was also implemented, and the performance is also shown in Table III for comparison; it demonstrates an average time saving of 53.7% with 0.021dB BDPSNR drop and 0.61% bit rate increase. As illustrated in Table III, our proposed algorithm obtains better RD performance, and reduces 8.9% more in average encoding time as compared to Shen's algorithm. Fig.14 illustrates the comparison of the time saving between our proposed algorithm and Shen's algorithm under different QPs. It can be seen that our algorithm can achieve larger time saving over different QPs. For "Exit" sequence which has less motion activities and smooth motion vectors, the time saving of Shen's algorithm is comparable to our proposed algorithm. For "Race1" sequence which has large motion activities and some noisy motion vectors, the time saving of Shen's algorithm doesn't compare as favorable to our proposed algorithm.

V. CONCLUSIONS

In this paper, a fast Inter mode decision is proposed to reduce the computational complexity of MVC. Experimental observations and analysis are carried out to study the modes' features, and the main focuses of optimization are demonstrated. The correlation between textural complexities and RD costs is also studied, and it is found that RD costs are sensitive to textural complexities. Consequently, each frame is segmented into three textural regions by utilizing Intra mode RD costs and their correlation between views. Then, based on textural region types and correlations, Skip mode is early decided, the disparity estimation is selected, and the Inter8×8 mode estimation is reduced. As compared to the full mode decision, experimental results show that the proposed algorithm achieves about 55~70% saving of the total encoding time with negligible loss of coding efficiency. In addition, the proposed algorithm could also be combined with fast motion search and disparity search algorithms to further save the computation time.

ACKNOWLEDGMENT

The authors would like to thank the journal reviewers for their valuable suggestions, and with special appreciation to Dr. Shen for help in discussing the implementation of his fast algorithm.

REFERENCES

[1] A. Smolic and P. Kauff, "Interactive 3-D video representation and coding technologies," *Proc. IEEE*, vol. 93, no. 1, pp. 98-110, Jan. 2005.

[2] M. Tanimoto, "Overview of free viewpoint television," *Signal Process.: Image Commun.*, vol. 21, no. 6, pp 454-461, July 2006.

[3] A. Vetro, P. Pandit, H. Kimata, A. Smolic and Y. K. Wang, ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16. Joint Multiview Video Model (JMVM) 8.0, Doc. JVT-AA207, Geneva, Switzerland, Apr. 2008.

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

[5] M. Flierl, A. Mavlankar, and B. Girod, "Motion and disparity compensated coding for multiview coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 11, pp. 1474-1484, Nov. 2007.

[6] H. Schwarz, D. Marpe and T. Wiegand, "Hierarchical B pictures," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-P014, Poznan, Poland, July 2005.

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

[8] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. Wu, and S. Wu, "Fast mode decision for intraprediction in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 813–822, July 2005.

[9] D. Wu, F. Pan, K. P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahardja, and C. C. Ko, "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 953–958, July 2005.

[10] J. You, W, Kim and J. Jeong, "16×16 macroblock partition size prediction for H.264 P slices," *IEEE Trans. Consumer Electron.*, vol. 52, no. 4, pp. 1377-1383, Nov. 2006.

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

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

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

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

[15] L. F. Ding, S. Y. Chien, and L. G. Chen, "Joint prediction algorithm and architecture for stereo video hybrid coding systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 11, pp.1324–1337, Nov. 2006.

[16] X. M. Li, D. B. Zhao, X. Y. Ji, Q. Wang and W. Gao, "A fast inter frame prediction algorithm for multi-view video coding," *IEEE Int. Conf. on Image Processing*, vol. 3, pp. 417-420, Sep. 2007.

[17] L. S. Young, S. K. Mu and C. K. Dong, "An object-based mode decision algorithm for multi-view video coding," *IEEE Int. Symposium on Multimedia*, pp. 74-81, Dec. 2008.

[18] Y. Chen, P. Pandit, and S. Yea, "WD 4 reference software for MVC," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-AD207, Jan. 2009.

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

[20] Y. P. Su, A. Vetro and A. Smolic, "Common test conditions for multiview video coding," ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-U211, Oct. 2006.

[21] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," ITU-T Q6/SG16, Doc. VCEG-M33, Apr. 2001.

BIOGRAPHIES

Wei Zhu was born in Zhejiang Province, China, in 1982. He received the B.Sc. degree from Zhejiang University, Hangzhou, China, in 2004. He is currently a Ph.D. candidate of Zhejiang University. His major research fields are digital video processing, multiview video coding, and compressed domain video analysis and retrieval.

Xiang Tian was born in Anhui Province, China, in 1979. He received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2001 and 2007, respectively. He is currently a lecturer in the Institute of Advanced Digital Technologies and Instrumentation, Zhejiang University. His major research field is the system design for multimedia and parallel processing.

Zhou Fan was born in Zhejiang Province, China, in 1978. He received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 2000 and 2006, respectively. He is currently an associate professor in the Institute of Advanced Digital Technologies and Instrumentation, Zhejiang University. His major research fields are embedded system and networking multimedia system.

Yaowu Chen was born in Liaoning Province, in China, in 1963. He received the Ph.D. degree from Zhejiang University, Hangzhou, China, in 1998. He is currently a professor and the director of the Institute of Advanced Digital Technologies and Instrumentation, Zhejiang University. His major research fields are embedded system, networking multimedia system, and electronic instrumentation system.