Fast inter-view prediction and mode selection for multiview video coding

Wei Zhu, Peng Chen, Yayu Zheng
College of Information Engineering
Zhejiang University of Technology
Hang Zhou, China

Jie Feng
School of Informatics and Electronics & Electronics
Zhejiang Sci-Tech University
Hang Zhou, China

***Abstract*—Multiview video coding (MVC) compresses the multiview video for efficiency storage and transmission, and it has been standardized as a recent extension of the H.264/AVC. In this paper, a fast inter-view prediction and mode selection algorithm is proposed to reduce the heavy computational complexity of MVC. First, partition activity of Inter modes is calculated by using the optimal Intra mode. Then, inter-view prediction is determined by employing the partition activity and the textural complexity of depth. Finally, Inter modes are selected by utilizing the partition activity and the textural complexity of pixels. As compared to the full mode decision, experimental results show that the proposed algorithm achieves 68% time saving on average with negligible loss of rate-distortion performance. As compared to the state-of-the-art fast algorithm, the proposed algorithm obtains higher and more stable time saving over a wide range of quantization parameters for different test sequences.**

Keywords-multiview video coding; mode decision; inter-view prediction; textural complexity; depth

I. INTRODUCTION

Multiview video which is captured from a set of view points can provide different visual perception as compared to the single view video, and it can be used to many new applications, such as 3D TV and free viewpoint TV. Multiview video coding (MVC) has been developed for the storage and transmission of the huge multiview video data, and it has been standardized as a recent extension of the H.264/AVC [1],[2]. Because MVC not only employs variable size motion estimation for traditional temporal prediction but also employs variable size disparity estimation for inter-view prediction [3], its computational complexity is every heavy.

Currently, some fast mode decision algorithms have been proposed to reduce the complexity of MVC. Ding et al. [4] presented a mode decision to choose the most probable coding mode by exploiting the correlation of coding information between views. Zatt *et al.* [5] proposed an adaptive early Skip mode decision scheme. But these two algorithms haven't reduced the complexity of inter-view prediction. Shen *et al*. [6] proposed a fast decision algorithm for both inter-view prediction and mode based on the motion homogeneity. Zeng *et al.* [7] proposed a motion activity-based mode decision algorithm. These two algorithms reduce complexity effectively with negligible loss of rate-distortion (RD) performance. However, they can't be applied to anchor frames, which can't provide the motion information. A fast Inter mode decision based on textural segmentation and correlations has been proposed in our previous work [8]. Although it can reduce the complexity of anchor frames, it can't be applied to the base view due to need the mode information of neighboring views.

In addition, above mentioned algorithms haven't exploited additional data of multiview video to further reduce the complexity. Depth information which is assumed to be obtained during the acquisition process indicates distances between camera and object, and it can be combined with original multiview video to synthesis virtual views. So multiview video plus depth (MVD) is useful to 3D applications, and joint multiview video plus depth coding (JMVDC) can reduce the total bit rate of 3D data [9]-[11]. Moreover, depth can be used to reduce the computational burden of MVC. In [12], a fast mode decision based on the analysis of the homogeneity of depth is proposed. And in [13], a fast mode decision based on depth information is presented. But these two algorithms haven't employed depth to reduce the complexity of inter-view prediction. The homogeneity of depth indicates the disparity consistency, so depth information can be used to determine inter-view prediction of small size Inter modes for the reduction of small size disparity estimation.

In this paper, a fast inter-view prediction and mode selection algorithm is proposed. First, partition activity of Inter modes is calculated by using the optimal Intra mode. Then, the partition activity and the textural complexity of depth are employed to determine inter-view prediction of small size Inter modes. Finally, the partition activity and the textural complexity of pixel are used to select small size Inter modes for non-anchor frames. The proposed algorithm can be applied to all Inter frames of all views for reducing the whole complexity of MVC.

The rest of paper is organized as follows: Section II presents our algorithm. Section III gives the experimental results. Finally, conclusions are given in Section IV.

978-1-4244-9306-7/11/$26.00 ©2011 IEEE

II. PROPOSED ALGORITHM

A. Partition Actitvity of Inter Modes

Like H.264/AVC, MVC also employs rate-distortion (RD) optimization technique [14] to select the optimal mode by using RD cost, which includes rate part and distortion part. Inter modes and Intra modes have different feature of rate part and distortion part for diverse modes. Small size Inter modes (Inter16x8, Inter8x16, and Inter8x8) have less distortion part of RD cost than large size Inter modes (Skip, Inter16x16) because of providing more precise motion and disparity estimation. However, because small size Inter modes need more bits for encoding motion vector of each partition block, they have larger rate part of RD cost than that of large size Inter modes. As similar with Inter modes, small size Intra modes (Intra4x4, Intra8x8) can also obtain less distortion part and more rate part than that of large size Intra mode (Intra16x16). Furthermore, the importance of rate part and distortion part are variable for different QPs. Fig. 1 illustrates proportion curves of Inter modes and Intra modes selected in Inter mode decision and Intra mode decision, respectively. It can be observed that proportion of small size Inter modes decreases as the increasing of quantization parameters (QPs), while the proportion of large size Inter modes increases. It demonstrates that partition activity of Inter modes is sensitive to QP, and the partition activity decreases as the increasing of QP. From Fig. 1, it can also be found that curves of Intra modes have the similar trends with curves of Inter modes. Hence, Inter modes and Intra modes have similar partition activities, which are sensitive to QP. And Intra modes can be used to predict the partition activity of Inter modes adaptively.

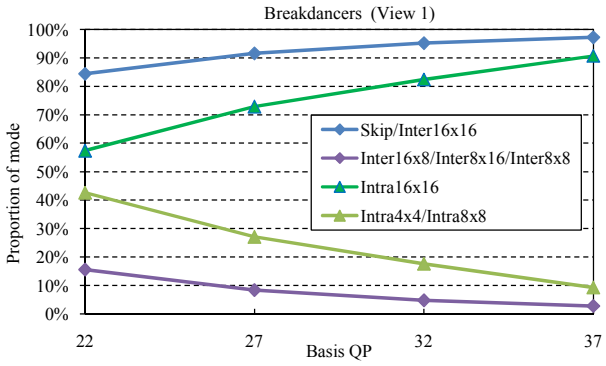

Figure 1. Proportion curves of Inter modes selected in Inter mode decision and Intra modes selected in Intra mode decision on view 1 for "Breakdancers" sequence.

Based on above analysis, the optimal Intra mode ($Mode_{\text{Intra}}$) selected in Intra mode decision is used to predict the partition activity of Inter modes, and it is calculated as follows:

$$PartitionActivity(n) = \begin{cases} 1, \text{ if } Mode_{\text{Intra}} == \text{Intra4x4} \,\|\, \text{Intra8x8} \\ 0, \text{otherwise} \end{cases} \quad (1)$$

where n is the index of macroblock (MB), 1 indicates that current MB has high partition activity, and 0 indicates that current MB has low partition activity. If the optimal Intra mode ($Mode_{\text{Intra}}$) is Intra4x4 or Intra8x8, current MB has high partition activity. In our algorithm, partition activity will be employed as a requirement condition for the selection of inter-view prediction and mode.

B. Selection of Inter-view Prediction Based on Partition Activity and Textural Complexity of Depth

Inter-view prediction is employed to reduce the redundancy between views, and it is especially useful for anchor frames, which have no temporal prediction. Because variable size disparity estimation is adopted to improve the efficiency of inter-view prediction, the complexity of MVC increases as compared to single view video coding. As the disparity of one object between two views is related with the distance between camera and the object, background objects which are far from camera have small disparities, and foreground objects which are near to camera have large disparities. Because depth information indicates the distance between camera and objects, it can be used to predict disparity homogenization of MB. Fig. 2 gives the pixel map and depth map of first frame on view 1 for "Ballet" sequence. If the depth texture with MB is homogeneous, the MB has disparity homogenization, and disparity estimation of large size Inter modes can achieve good prediction. If depth texture within MB is diversity, small size disparity estimation can achieve better prediction efficiency. So depth texture can be used to determine the selection of inter-view prediction.

(a) (b)

Figure 2. Pixel map and depth map of the first frame on view 1 for "Ballet" sequence. (a) Pixel map. (b) Depth map.

As the analysis in subsection A and the above analysis, the selection of inter-view prediction for small size Inter modes is decided by using both the partition activity and the textural complexity of depth. The textural complexity of depth for each Inter mode is calculated as follows:

$$DepthDev_{\text{WxH}}(n) = \sum_{j=0}^{B} \sum_{i=0}^{\text{WxH}} \left| Depth(i,j) - Depth_{\text{AVG}}(j) \right| \quad (2)$$

where WxH is the partition size of Inter mode, $Depth(i, j)$ is the value of depth i in block j, and $Depth_{\text{AVG}}(j)$ is the average depth value of block j. The selection steps of inter-view prediction for small size Inter modes are given as follows:

First, the textural complexity of depth for large size Inter modes ($DepthDev_{16x16}$) and the textural complexity of depth for small size Inter modes ($DepthDev_{16x8}$, $DepthDev_{8x16}$, and $DepthDev_{8x8}$) are calculated in (2), respectively. Then, the selection of inter-view prediction for small size Inter modes are determined based on the partition activity and by comparing the textural complexity of small size Inter modes with $DepthDev_{16x16}$.

$$InterViewEn_{16x8}(n)=\begin{cases}1, & \text{if } DepthDev_{16x8}<\alpha\times DepthDev_{16x16} \\ & \&\&PartitionActivity(n)==1 \\ 0, & \text{otherwise}\end{cases} \quad (3)$$

$$InterViewEn_{8x16}(n)=\begin{cases}1, & \text{if } DepthDev_{8x16}<\alpha\times DepthDev_{16x16} \\ & \&\&PartitionActivity(n)==1 \\ 0, & \text{otherwise}\end{cases} \quad (4)$$

$$InterViewEn_{8x8}(n)=\begin{cases}1, & \text{if } DepthDev_{8x8}<\beta\times DepthDev_{16x16} \\ & \&\&PartitionActivity(n)==1 \\ 0, & \text{otherwise}\end{cases} \quad (5)$$

where $InterViewEn_{16x8}$, $InterViewEn_{8x16}$, and $InterViewEn_{8x8}$ are enable of inter-view prediction for Inter16x8 mode, Iner8x16 mode, and Inter8x8 mode, respectively. α and β are inserted as a tradeoff between accuracy and complexity. The experimental analysis of α and β values is performed by utilizing MVC reference software with "Breakdancers" and "Ballet" sequences, and their typical values belong to [0.5, 1.0]. Because MBs with dramatic disparity differences among 8x8 blocks have larger $DepthDev_{16x16}$ than that of MBs with only dramatic disparity differences among 16x8 or 8x16 blocks, β should be less than α.

For anchor frames which only employ inter-view prediction, the selection of small size Inter modes are directly decided by (3), (4), and (5). For non-anchor frames, the selection of Inter-view prediction for small size Inter modes are determined by (3), (4), and (5).

C. *Selection of Inter Modes Based on Partition Activity and Textural Complexity of Pixel*

Because temporal prediction has better compression efficiency than inter-view prediction [3], the selection of Inter modes for non-anchor frames is decided according to the feature of temporal prediction. If a MB has textural homogeneity of pixel, it is very likely to have motion homogenization due to may well be in a same object, and the selection of small size Inter modes is not needed. On the other hand, if a MB select a small Inter mode as the optimal Inter mode, it is very likely to have diverse texture due to include different parts of motion objects. Therefore, the textural complexity of MB can be used to decide the selection of small size Inter modes for non-anchor frames. In our algorithm, the pixel deviation of MB is employed to quantify the textural complexity, and it is defined as follows:

$$MbPixelDev(n)=\frac{1}{256}\times\sum_{i=0}^{256}\left|Pixel(i)-Pixel_{\mathrm{AVG}}\right| \quad (6)$$

where $Pixel(i)$ is the value of pixel i in current MB, 256 is the pixel number of MB, and $Pixel_{\mathrm{AVG}}$ is the average pixel value. The correlation between textural complexity and small size Inter modes is analyzed by using experimental methodology. Fig. 3 shows proportion curves of small size Inter modes with *MbPixelDev* larger than different thresholds for non-anchor frames of "Breakdancers" sequence on view1. It can be observed that above 80% of Inter16x8 and Inter8x16 modes' *MbPixelDev* are larger than 3, and above 80% of Inter8x8 mode's *MbPixelDev* are larger than 8. These indicate that MB with small size Inter modes has large textural complexity, and MBs with Inter8x8 mode has larger textural complexity as compared to MBs with Inter16x8 mode and Inter8x16 mode. In addition, it can be found that MBs with Inter16x8 mode and Inter8x16 mode have similar textural complexity.

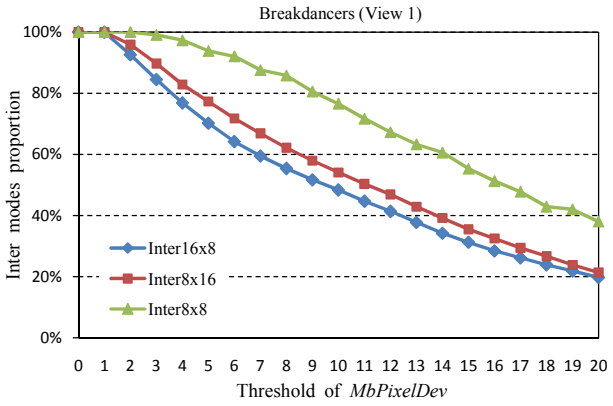

Figure 3. Proportion curves of Inter modes with MbPixelDev larger than different thresholds on view 1 for "Breakdancers" sequence. (Basis QP=32)

Based on above analysis and the analysis of partition activity in subsection A, the selection of small size Inter modes for non-anchor frames is determined based on both the partition activity and the textural complexity of pixel. The selection of Inter16x8 mode and Inter8x16 mode are determined in (7), and the selection of Inter8x8 mode is determined in (8):

$$InterModeEn_{16x8\_8x16}(n)=\begin{cases}1, & \text{if } MbPixelDev(n)>TH_1 \\ & \&\&PartitionActivity(n)==1 \\ 0, & \text{otherwise}\end{cases} \quad (7)$$

$$InterModeEn_{8x8}(n)=\begin{cases}1, & \text{if } MbPixelDev(n)>TH_2 \\ & \&\&PartitionActivity(n)==1 \\ 0, & \text{otherwise}\end{cases} \quad (8)$$

where TH_1 and TH_2 are thresholds of *MbPixelDev*. Because MB with Inter8x8 mode has larger average textural complexity, TH_2 is larger than TH_1. If textural complexity of MB is larger than TH_1 and MB has high partition activity, the estimation of Inter16x8 mode and Inter8x16 mode are performed. If pixel deviation of MB is larger than TH_2 and MB has high partition activity, the estimation of Inter8x8 is performed. The experimental analysis shown in Fig. 3 is also used to select values of TH_1 and TH_2.

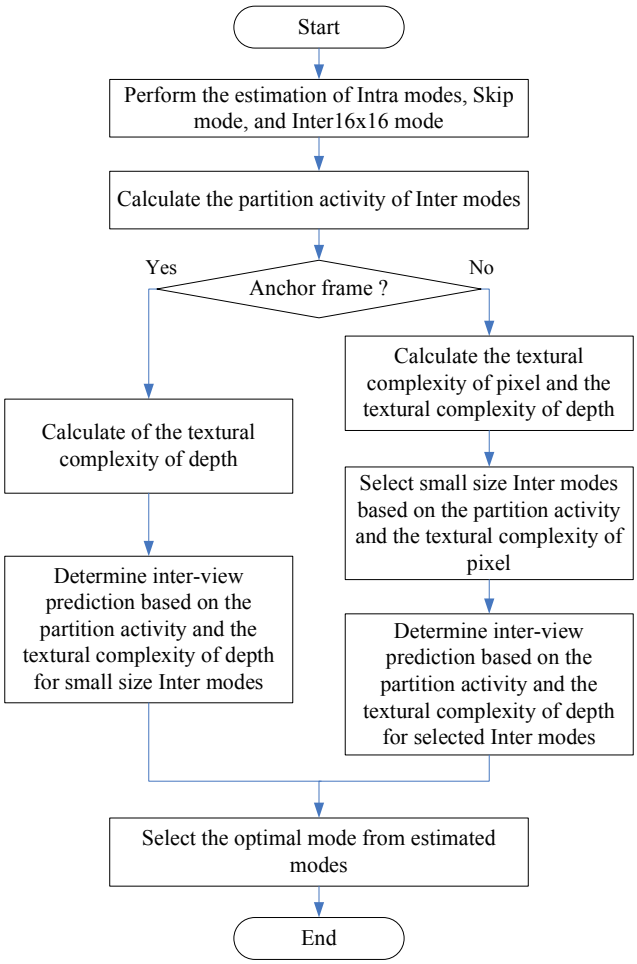

Figure 4. Flowchart of proposed algorithm.

For each small size Inter modes selected in (7) and (8), the selection of inter-view prediction is decided by (3), (4), and (5) in subsection B.

D. Overall Algorithm

A flowchart of proposed algorithm is illustrated in Fig. 4. The overall algorithm includes the two methods presented in subsection B and C, and the detail steps for each MB are given as follows:

1) Perform the estimation of Intra modes, Skip mode, and Inter16x16 mode.
2) Calculate the partition activity of Inter modes in (1) by using the optimal Intra mode.
3) If current frame is an anchor frame, go to step 4, else go to step 6.
4) Calculate the textural complexity of depth in (2).
5) Determine the inter-view prediction for each small size Inter modes in (3), (4), and (5). Then go to step 9.
6) Calculate textural complexities of pixel and depth in (2) and (6), respectively.
7) Select small size Inter modes in (7) and (8).
8) For each small size Inter modes selected in step 7, its inter-view prediction is determined in (3), (4), and (5).
9) Select the optimal mode from the estimated modes.

III. EXPERIMENTAL RESULTS

The proposed algorithm has been implemented on MVC reference software JMVC 4.0, which adopts the fast search algorithm with search range 96. Two test sequences, "Breakdancers" and "Ballet" with depth maps, were chosen for experiment. "Breakdancers" sequence has large motion activities with many picture noises, and "Ballet" sequence has small motion activities with little picture noises. Three views (view 0, view 1 and view 2) of these two sequences were chosen, and our algorithm was implemented on all three views. Based on the common test condition in [15], the GOP length of each sequence was set to 15 and four basis QPs (22, 27, 32, and 37) were performed to generate different bit rates. The saving of entire encoding time (ΔTime) was calculated to evaluate the computational performance. The Bjontegaard delta peak signal-to-noise ratio (BDPSNR) and Bjontegaard delta bit rate (BDBR) [16] were used to evaluate the RD performance under four QPs. A negative BDPSNR or a positive BDBR indicates a coding loss and is not preferred. α=1.0, β=0.6, TH_1=3, and TH_2=8 are selected for a favorable experimental results.

As compared to the full mode decision in JMVC, the performance of proposed algorithm for different views is illustrated in Table I. For "Breakdancers" sequence, the proposed algorithm achieves an average 64%, 68%, and 64% time saving for view 0, view 1, and view 2, respectively. For "Ballet" sequence, proposed algorithm achieves 70%, 73%, and 68% for view 0, view 1, and view 2, respectively. The average of time saving for two sequences with three views is 68%, while the average BDPSNR drop is only 0.020 dB and the average BDBR increase is only 0.76%. It demonstrates that our algorithm achieves significant time saving with maintaining negligible RD performance.

To compare proposed algorithm with a state-of-the-art fast mode decision algorithm, the selective disparity estimation and variable size motion estimation (SDEME) algorithm proposed by Shen *et al.* in [6] was implemented. Because SDEME needs the motion vectors from reference views, it doesn't be applied to view 0. As compared to the full mode decision, the performance of SDEME is given in Table I. It can be seen that SDEME achieved 45% time saving on average with 0.016 dB BDPSNR drop and 0.57% BDBR increase. As compared to SDEME, the proposed algorithm achieves 22% more time saving on average with similar RD performance. Fig. 5 shows the comparison of the time saving between proposed algorithm and SDEME on view 1 and view 2 under different QPs. It can be observed that proposed algorithm achieves larger time saving over all view under all QPs. For "Breakdancers" sequence, which has large motion region and many noisy motion vectors, the time saving of SDEME doesn't compare as favorable to proposed algorithm on both view 1 and view 2. For "Ballet" sequence, which has less motion region and smooth motion, the time saving of SDEME is comparable to proposed algorithm at high QP setting on view 1, while also achieves less time saving than proposed algorithm on other cases. From Fig. 5, it can be also found that the gap of time saving between

TABLE I
PERFORMANCE OF PROPOSED ALGORITHM AND SDEME AS COMPARED TO THE FULL MODE DECISION

Sequences	View Index	Proposed Algorithm			SDEME		
		ΔTime (%)	BDPSNR (dB)	BDBR (%)	ΔTime (%)	BDPSNR (dB)	BDBR (%)
Breakdancers	View 0	64.4	-0.015	0.73	N/A	N/A	N/A
	View 1	67.5	-0.021	0.92	36.6	-0.020	0.81
	View 2	63.5	-0.018	0.77	25.9	-0.003	0.20
Ballet	View 0	69.8	-0.009	0.27	N/A	N/A	N/A
	View 1	73.0	-0.039	1.19	68.5	-0.035	1.04
	View 2	68.1	-0.018	0.61	50.6	-0.007	0.22
Average		67.7	-0.020	0.75	45.4	-0.016	0.57

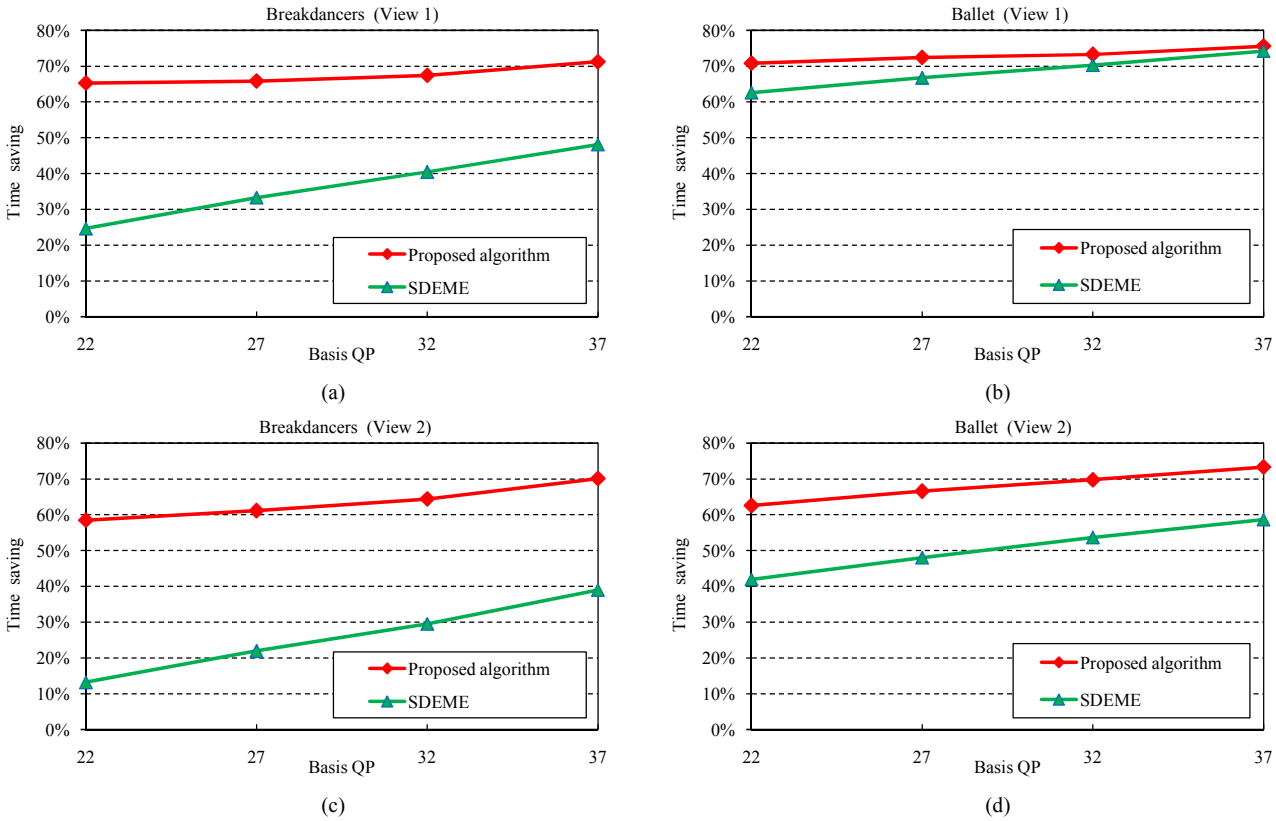

Figure 5. Time saving curves of our proposed algorithm and SDEME under different basis QPs (22, 27, 32 and 37). (a) Time saving curves on view 1 for "Breakdancers" sequence. (b) Time saving curves on view 1 for "Ballet" sequence. (c) Time saving curves on view 2 for "Breakdancers" sequence. (d) Time saving curves on view 2 for "Ballet" sequence.

proposed algorithm and SDEME on view2 is larger than that on view 1, and proposed algorithm achieves stable time saving for two sequences on all views under different basis QPs.

IV. CONCLUSION

To reduce the computational complexity of MVC, a fast inter-view prediction and mode selection algorithm is proposed in this paper. First, the partition activity of Inter modes is predicted by using the partition size of Intra mode. Then, inter-view prediction is determined by using the partition activity and the textural complexity of depth for small size Inter modes. Finally, small size Inter modes are selected based on the partition activity and textural complexity of pixel for non-anchor frames. Experimental results illustrates that our proposed algorithm achieved an average 68% encoding time saving, while the loss of RD performance is negligible. Moreover, our algorithm could be combined with fast multi-reference frame selection algorithm to further reduce the encoding time, and it will be in our next work.

ACKNOWLEDGMENT

This work was supported in part by the Natural Science Foundation of Zhejiang University of Technology (2011), and in part sponsored by the Natural Science Foundation of

Zhejiang Province under Grants No. Y1110532, Y1110175, and Y1100632. In addition, the authors would thank Microsoft Research for providing the test sequences.

REFERENCES

[1] ITU-T and ISO/IEC JTC 1, Advanced video coding for generic audiovisual services, ITU-T Recommendation H.264 and ISO/IEC 14496 (MPEG-4 AVC), 2010.

[2] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/AVC standard," Proc. IEEE, vol. 99, pp. 626-642, 2011.

[3] 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.

[4] 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.

[5] B. Zatt, M. Shafique, S. Bampi, and J. Henkel, "An adaptive early skip mode decision scheme for multiview video coding," Picture Coding Symposium, pp.42-45, Dec. 2010.

[6] 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.

[7] H. Zeng, K. K. Ma, and C. Cai, "Motion activity-based block size decision for multi-view video coding," Picture Coding Symposium, pp. 166-169, Dec. 2010.

[8] W. Zhu, X. Tian, and Y. W. Chen, "Fast Inter mode decision based on textural segmentation and correlations for multiview video coding," IEEE Trans. Consumer Electron., vol. 56, No. 3, pp. 1696–1704, Aug. 2010,

[9] K. Muller, P. Merkle, and T. Wiegnad, "3-D video representation using depth maps," Proc. IEEE, vol. 99, pp. 643-656, Apr. 2011.

[10] J. Zhang, M. Hannuksela, and H. Li, "Joint multiview video plus depth coding," IEEE Int. Conf. Image Process., pp. 2865-2868, Sep. 2010.

[11] M. K. Kang, J. Lee, I. Lim, and Y. S. Ho, "Object-adaptive depth compensated Inter prediction for depth video coding in 3D video system," SPIE Proc. Vis. Info. Process. Commun. II, vol. 7882, pp. 78820N-1-14, Jan. 2011.

[12] G. Cernigliaro, F. Jaureguizar, A. Ortega, and J. Cabrera. "Fast mode decision for multiview video coding based on depth maps," SPIE Proc. Vis. Commun. Image Process., vol. 7257, pp. 72570N-1-10, Jan. 2009.

[13] Y. H. Lin, J. Ling, "A depth information based fast mode decision algorithm for color plus depth-map 3D videos," IEEE Trans. Broadcasting, vol. 57, no. 2, June 2011.

[14] 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.

[15] 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.

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