2011 IEEE International Symposium on Multimedia

An adaptive Inter mode decision for multiview video coding

Wei Zhu, Peng Chen, Yayu Zheng
College of Information Engineering
Zhejiang University of Technology
Hang Zhou, China
weizhu@zjut.edu.cn, chenpeng@zjut.edu.cn, yayuzheng@zjut.edu.cn

Jie Feng
School of Informatics and Electronics & Electronics
Zhejiang Sci-Tech University
Hang Zhou, China
fengjie_zju@yahoo.cn

***Abstract*—Multiview video coding (MVC) plays an important role in 3D video system, while the huge computational complexity blocks its applications. This paper proposes an adaptive Inter mode decision algorithm to reduce the complexity of MVC. First, the selection of Inter modes is determined by using the textural region type of macroblock (MB). Then, the estimation of small size Inter modes (Inter16x8, Inter8x16, and Inter8x8) is decided based on the motion homogenization of MB, which is predicted by utilizing the motion estimation results of Inter16x16 mode. Finally, the complexity of Inter8x8 mode estimation is progressively reduced by employing rate-distortion (RD) costs of estimated modes. As compared to the full mode decision in MVC reference software, the proposed algorithm achieved 71% encoding time saving on average with 0.026 dB peak signal-to-noise ratio loss and 0.74% bit rate increase.**

Keywords-multiview video coding; mode decision; motion homogenization; rate-distortion cost

I. INTRODUCTION

Multiview video is captured by different camera from different view angle at the same time instance. It can be applied to many applications, such as 3D TV and free viewpoint TV. Multiview video coding (MVC) compresses multiview video data for efficient storage and transmission. The MVC standard has been finalized by the joint video team (JVT) as the extension of H.264/AVC [1]. MVC encodes different views, and adopts variable size Inter prediction like H.264/AVC [2]. In addition, MVC not only employs the temporal prediction in traditional video compression but also adopts inter-view prediction to reduce the redundancy between views [3],[4]. These result in the heavy computational complexity of MVC. Thus, fast algorithms should be developed for MVC.

Recently, some fast mode decision algorithms have been proposed to reduce the complexity of MVC. Zatt *et al.* [5] proposed an early skip mode decision algorithm, while the computational performance could be further improved. Shen *et al.* [6] presented a low complexity mode decision algorithm by using inter-view correlation in the coding information. However, because of employing the global disparity vector to obtain the local coding information in neighbor views, the algorithm hasn't performed very well for the video sequences with widely different disparities. Yang *et al.* [7] proposed a fast reference frame and mode selection algorithm by utilizing coded block patterns, spatial mode correlation, and rate-distortion (RD) costs. The algorithm can reduce the encoding time significantly, but it has significant bit rate increase for inter-view frames under high quantization parameter (QP). Chiang *et al.* [8] proposed a hierarchical two-stage neural classifier for fast mode decision. This algorithm can also reduce the encoding time significantly. However, its RD performance for video scenes with high motion and complicated background is not very well. A fast Inter mode decision based on textural segmentation and correlations has been proposed in our previous work [9]. It can reduce the complexity effectively with maintaining high RD performance. But the algorithm hasn't fully studied the selection of Inter mode sizes, and the complexity can be further reduced.

In this paper, an adaptive Inter mode decision algorithm is proposed. It employs the textural region type, motion homogenization, and RD costs to reduce Inter modes estimation. First, the selection of Inter modes is determined by using the textural region type, which is calculated by employing the textural segmentation technique in [9]. Second, the estimation of small size Inter modes is decided based on the motion homogenization of MB, which is predicted by utilizing motion estimation results of Inter16x16 mode. Finally, the complexity of Inter8x8 mode estimation is progressively reduced by employing RD costs of estimated modes.

The rest of this paper is organized as follows: Section II describes the proposed algorithm. Section III introduces the experimental results. Section IV gives the conclusions.

II. ADAPTIVE INTER MODE DECISION ALGORITHM

A. Select Inter Modes by Using Textural Region Type

The mode decision algorithm in [9] segments each picture into low, medium, high textural regions by using Intra RD cost and its correlation between views, then the region type of MB is used for the early decision of Skip mode, the selection of disparity estimation, and the reduction of Inter8x8 mode estimation. However, the correlation between textural region type and Inter mode sizes haven't been fully studied. As shown in Table I, we studied the statistics of the proportion of different Inter mode sizes within the three textural regions. In the low textural region, it can be observed that 16x16 occupies an average 98.6% of Inter mode sizes, while proportions of 16x8, 8x16, and 8x8 are only 0.7%, 0.7%, and 0.0%, respectively. Thus, the

978-0-7695-4589-9/11 $26.00 © 2011 IEEE
DOI 10.1109/ISM.2011.54

IEEE computer society

TABLE I
INTER MODE SIZES PROPORTION IN DIFFERENT TEXTURAL REGIONS ON VIEW 1 (BASIS QP = 22).

Sequence	Proportions in The Low Textural Region (%)				Proportions in The Medium Textural Region (%)				Proportions in The High Textural Region (%)			
	16x16	16x8	8x16	8x8	16x16	16x8	8x16	8x8	16x16	16x8	8x16	8x8
Exit	98.7	0.6	0.6	0.0	90.3	4.2	4.7	0.8	72.2	9.0	11.8	7.1
Ballroom	96.8	1.4	1.6	0.1	89.7	4.4	4.3	1.6	74.0	8.0	9.1	8.9
Race1	99.1	0.6	0.3	0.0	95.0	2.9	1.6	0.5	75.1	10.8	6.4	7.6
Flamenco2	99.8	0.1	0.1	0.0	87.4	5.6	4.7	2.3	60.5	12.2	14.4	12.8
Rena	98.5	0.5	0.9	0.1	90.3	2.6	5.8	1.2	79.5	6.2	10.6	3.6
Average	98.6	0.7	0.7	0.0	90.5	3.9	4.2	1.3	72.3	9.2	10.5	8.0

estimation of Inter16x8, Inter8x16, and Inter8x8 modes in the low textural region is not necessary. In the medium textural region, proportions of 16x8 and 8x16 are 3.9% and 4.2%, which are not negligible. The proportion of 8x8 in the medium textural region is still very small, so the estimation of Inter8x8 mode is not needed. In the high textural region, the average proportions of 16x8, 8x16, and 8x8 are 9.2%, 10.5%, and 8.0%, respectively, which indicate considerable quantities of their Inter modes.

Based on above analysis, the selection of Inter modes is determined in Table II according to textural region type. If the current MB belongs to the low textural region, Skip and Inter16x16 modes are selected as mode candidates for Inter mode decision. If the current MB belongs to the medium textural region, Skip, Inter16x16, Inter16x8, and Inter8x16 modes are selected. If the current MB belongs to the high textural region, all Inter modes are selected.

TABLE II
SELECTION OF INTER MODES ACCORDING TO TEXTURAL REGION TYPE

Textural region type	Mode candidates
Low	Skip, Inter16x16
Medium	Skip, Inter16x16, Inter16x8, Inter8x16
High	All Inter modes

B. Decide Small Size Inter Modes Estimation Based on Motion Homogenization

If there are consistent motions in a MB, large size Inter modes (Skip, Inter16x16) may well be selected as the optimal Inter mode. So if we can predict the motion homogenization of MB, the estimation of small size Inter modes (Inter16x8, Inter8x16, and Inter8x8) is not needed. Considering the selection of Inter modes estimation in the last subsection, Inter16x16 mode estimation is always performed, so the motion estimation result of Inter16x16 mode can be used to predict the motion homogenization of MB. In this paper, the forward motion vector of Inter16x16 mode (MV_{16x16}) and its predictor of motion vector (PMV) are employed to predict the motion homogenization. If MV_{16x16} is equal to PMV, it indicates that the current MB has consistent motion trend with its spatial neighbor MBs, and it is deemed to have motion homogenization. Therefore, the motion homogeneity of MB is calculated as follows:

$$MotionHomogenity(n) = \begin{cases} 1, \text{ if } \quad MV_{16\times16}(n) == PMV(n) \\ 0, \text{otherwise} \end{cases} \quad (1)$$

where n is the index of MB, 1 indicates that MB has motion homogenization with its spatial neighboring MBs, and 0 indicates that MB has no motion homogenization. Furthermore, Table III gives the statistical analysis of the correlation between the motion homogenization and Inter mode sizes for five test sequences with different video scenes. It can be found that small mode sizes occupy negligible proportion for MBs with motion homogenization.

TABLE III
PROPORTION OF INTER MODE SIZES WITH MOTION HOMOGENIZATION ON VIEW 1 (BASIS QP = 22).

Sequences	Proportion of Inter Mode Sizes With Motion Homogenization			
	16x16	16x8	8x16	8x8
Exit	97.7	0.9	1.0	0.5
Ballroom	96.2	1.3	1.4	1.1
Race1	97.5	1.0	0.7	0.8
Flamenco2	98.4	0.6	0.5	0.4
Rena	98.3	0.5	1.0	0.2
Average	97.6	0.9	0.9	0.6

According to above analysis, the estimation of small size Inter modes is decided based on motion homogenization of MB. If the current MB has motion homogenization, the estimation of small size Inter modes is not performed. Because of needing motion information, the proposed method in this subsection is only used for non-anchor frames.

C. Reduce Inter8x8 Mode Estimation by Employing RD costs

Inter8x8 mode includes 8x8, 8x4, 4x8, and 4x4 partition sizes, and its optimal partition size for each 8x8 block is selected after performing estimation of these four sizes. The estimation of 8x8 includes motion estimation, disparity estimation, and direct prediction. And the estimation of other sizes includes motion estimation and disparity estimation. Due to the motion and disparity estimation for four partition sizes, Inter8x8 mode consumes more than 50% encoding time, while it occupies the least proportion among all Inter modes. Therefore, the complexity of Inter8x8 mode should be reduced. RD cost is employed to select the optimal mode, and the numerical relation between large size modes RD costs and small size modes RD costs can reflect the trend of selecting the optimal mode. Based on this idea, our algorithm employs RD costs of estimated modes to progressively reduce the complexity of Inter8x8 mode estimation.

First, Inter8x8 mode estimation is decided by using RD costs of Inter16x16, Inter16x8, and Inter8x16 modes. If Inter16x16 RD cost is less than Inter16x8 RD cost and Inter8x16 RD cost, it indicates that the large size (16x16) estimation can achieve better compression than that of smaller sizes (16x8, 8x16). So it is not necessary to perform the estimation of Inter8x8 mode, which has the smallest size.

Second, the estimation of 8x8 is performed, and its RD cost is combined with the minimum RD cost of estimated modes to decide the estimation of 8x4, 4x8, and 4x4. If the RD cost of 8x8 is less than quarter of the minimum RD cost, it indicates that the estimation of 8x8 has high efficiency, and the estimation of 8x4 and 4x8 is not needed. Otherwise, perform the estimation of 8x4 and 4x8.

Finally, the estimation of 4x4 is determined by using the RD costs of 8x8, 8x4, and 4x8. If the RD cost of 8x8 is less than RD costs of 8x4 and 4x8, the estimation of the smaller size (4x4) is also not necessary.

D. Overall Algorithm

Overall algorithm is the combination of above mentioned methods, and the detail steps are introduced as follows:

1) Perform mode estimation of Intra modes, Skip mode, and Inter16x16 mode.
2) Calculate the textural region type of MB based on the textural segmentation technique in [9].
3) If current MB belongs to the low textural region, go to step 12, else go to step 4.
4) If current frame is an anchor frame, go to step 7, else go to step 5.
5) Predict the motion homogenization of MB by using the motion vector and *PMV* of Inter16x16 mode.
6) If current MB has motion homogenization, go to step 12, else go to step 7.
7) If current MB belongs to medium region, perform the estimation of Inter16x8 and Inter8x16 modes, and go to step 12, else go to step 8.
8) If Inter16x16 RD cost is less than Inter16x8 and Inter8x16 RD costs, go to step 12, else go to step 9.
9) Get the current minimum RD cost among the estimated modes.
10) For each 8x8 block of Inter8x8 mode, the estimation of 8x8 is performed first. If RD cost of 8x8 is less than quarter of the current minimum RD cost, go to step 11, else perform the estimation of 8x4 and 4x8. If the RD cost of 8x8 is less than RD costs of 8x4 and 4x8, go to step 11, else perform the estimation of 4x4.
11) Select the optimal partition size for the current 8x8 block. If the current block is the last 8x8 block of MB, go to step 12, else go to step 10.
12) Select the optimal mode with minimum RD cost.

III. EXPERIMENTAL RESULTS

The proposed algorithm was implemented on the MVC reference software JMVC 4.0, which employs the fast search mode with search range 96. Five test sequences: “Exit”, “Ballroom”, “Race1”, “Flamenco2”, and “Rena” were selected for our experiment. “Exit” and “Rena” have small motion with moderate texture, “Flamenco2” has moderate motion and simple texture, and “Ballroom” and “Race1” have large motion with complicated texture. Two views (view 0 and view 1) were chosen for Flamenco2 sequence due to its 2D-cross camera arrangement, and three views (view 0, view 1, and view 2) were employed for other sequences. Our algorithm was implemented on view 1, and view 0 and view 2 were only used as reference views. Based on the common test condition in [10], the GOP lengths of these sequences was set to 12 or 15 for keeping the temporal access ability of 0.5 s, and four basis QPs (22, 27, 32, and 37) were adopted to generate different bit rates. The saving of entire encoding time (ΔTime) was used to evaluate the computational performance. The Bjontegaard delta peak signal-to-noise ratio (BDPSNR) and Bjontegaard delta bit rate (BDBR) [11] were calculated to evaluate the overall RD performance with all basis QPs. A negative BDPSNR or a positive BDBR indicates an overall coding gain loss.

As compared to the full mode decision, the performance of proposed algorithm is shown in Table IV. It can be seen that our algorithm achieved 71% time saving on average with 0.026 dB BDPSNR drop and 0.74% BDBR increase. For “Flamenco2” sequence, our algorithm achieves 69% time saving, which is the least among five sequences. For “Exit” sequence, our algorithm achieves 73% time saving, which is the largest among five sequences. These indicate that the proposed algorithm has a stable reduction of computational complexity for different video scenes.

The fast Inter mode decision (FIMD) proposed in [9] was also implemented. As compared to the full mode decision, the performance of FIMD is illustrated in Table IV. It can be observed that FIMD achieved 63% time saving on average with 0.005 dB BDPSNR drop and 0.13% BDBR increase. As compared to FIMD, the proposed algorithm achieves 8% more time savings on average with a slight loss of RD performance. For “Rena” sequence, the proposed algorithm achieves up to 13% more time saving.

The proposed algorithm can be combined with other fast mode decision algorithm to further reduce the complexity, and we combined it with the pre-decision of Skip mode technique and the selection of disparity estimation technique in FIMD. The performance of the proposed algorithm plus FIMD is also illustrated in Table IV, and the average time saving reaches to 77% with 0.030 dB BDPSNR loss and 0.87% BDBR increase.

To evaluate the complexity performance under different basis QPs, Figure 1 shows time saving curves of different algorithms for “Race1” and “Rena” sequences. It can be found that the proposed algorithm has higher time saving than FIMD. For “Race1” sequence which has high motion and complicated texture, the proposed algorithm obtains slightly more time saving than FIMD, while the proposed algorithm plus FIMD achieves higher gain of time saving than other sequences. For “Rena” sequence which has small motion and simple texture, the proposed algorithm achieves significantly more time saving than FIMD, and it obtains less

TABLE IV
PERFORMANCE OF PROPOSED ALGORITHM, FIMD, AND PROPOSED ALGORITHM PLUS FIMD.

Sequences	Proposed Algorithm			FIMD			Proposed Algorithm + FIMD		
	ΔTime (%)	BDPSNR (dB)	BDBR (%)	ΔTime (%)	BDPSNR (dB)	BDBR (%)	ΔTime (%)	BDPSNR (dB)	BDBR (%)
Exit	73.0	-0.025	1.14	66.3	-0.002	0.12	77.9	-0.028	1.25
Ballroom	69.9	-0.034	1.00	60.3	-0.014	0.41	76.0	-0.045	1.32
Race1	72.6	-0.025	0.63	70.3	0.001	-0.06	80.6	-0.026	0.66
Flamenco2	68.6	-0.026	0.57	59.9	-0.004	0.09	75.9	-0.027	0.59
Rena	69.2	-0.018	0.37	56.2	-0.006	0.12	74.1	-0.025	0.54
Average	70.7	-0.026	0.74	62.6	-0.005	0.13	76.9	-0.030	0.87

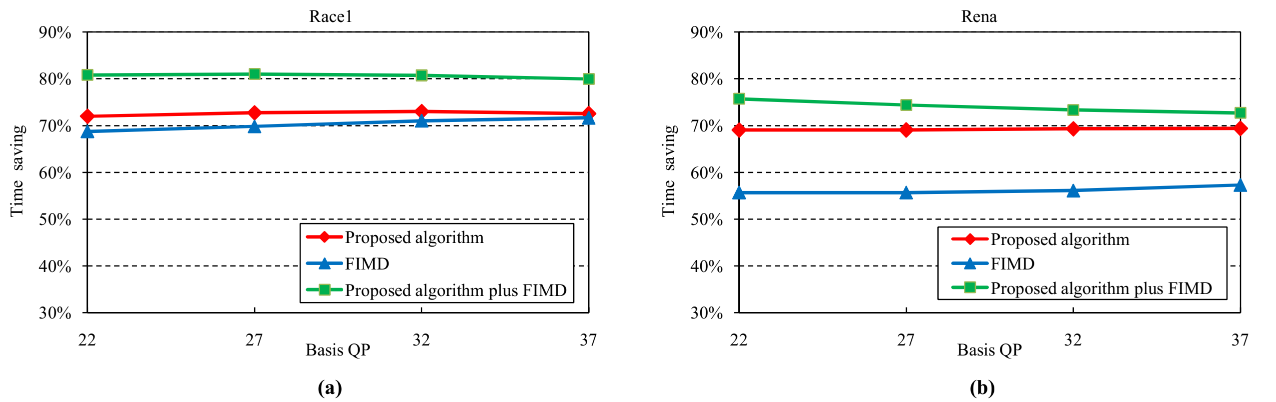

Figure 1. Time saving curves of proposed algorithm, FIMD, and proposed algorithm plus FIMD under different basis QPs. (a) Time saving curves for "Race1" sequence. (b) Time saving curves for "Rena" sequence.

gain of time saving than that of other sequences after be combined with FIMD.

IV. CONCLUSION

This paper proposes an adaptive Inter mode decision algorithm to reduce the computational complexity of MVC. For each MB, the decision process of Inter modes is determined by using textural region type, motion homogenization, and RD costs. As compared to the full mode decision in JMVC, the proposed algorithm achieves a significant encoding time saving for different sequences, and maintains a negligible loss of RD performance. To further reduce the computational complexity, our next work will focus on fast Intra mode decision and combine fast Inter mode decision with fast search algorithm.

ACKNOWLEDGMENT

This work was sponsored in part by the Natural Science Foundation of Zhejiang University of Technology (2011), supported in part by the Research Foundation of Education Bureau of Zhejiang Province under Grant Y201018355, and the Natural Science Foundation of Zhejiang Province under Grants Y1110175, Y1100803, and Y1100632.

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] 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.

[3] 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.

[4] 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.

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

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

[7] S. H. Yang and Y. S. Liou, "Fast reference frame and mode selection for multiview video coding based on coded block patterns," IEEE Int. Conf. on Multimedia and Expo, pp. 1102-1107, July, 2010.

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

[9] 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.

[10] 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.

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