A Fast Inter Mode Decision for Multiview Video Coding

Wei Zhu, Wei Jiang, and Yaowu Chen
Institute of Advanced Digital Technology and Instrumentation
Zhejiang University
Hangzhou, P.R. China
zhuwei_zju@zju.edu.cn, jason_sl006@gmail.com, and cyw@mail.bme.zju.edu.cn

Abstract—**Multiview video coding (MVC) plays a critical role in reducing the ultra high data bandwidth of 3-D video, and it has attracted great attention from industries and research institutes. However, with the increasing number of views, the complexity of MVC increases greatly which affects its realistic applications. In this paper, a fast Inter mode decision scheme for the Skip mode and the Inter sub modes is proposed. Based on the Rate-Distortion (RD) cost correlation between neighboring views, and the RD cost of different textural segmentation regions, a pre-decision of the Skip mode is introduced to reduce other modes' estimation. In addition, the estimated direction of Inter sub modes is predicted based on the optimal direction of the Inter16×16 mode. For the Peak Signal to Noise Ratio, experimental results show that an average 55% reduction of the total computation time with degradation of less than 0.01 dB is achieved as compared to the MVC reference software.**

Keywords-MVC; 3-D video; mode decision; RD cost; textural segmentation regions

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 results in 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 standardization of the MVC is developing in the joint video team of the ITU-T video coding experts group (VCEG) and ISO/IEC moving picture experts group (MPEG). The team has developed a Joint Multiview Video Model (JMVM) [3] which is based on H.264/AVC [4]. Since views are captured from 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. Therefore, MVC computational complexity is much higher than the H.264/AVC. The basic reference coding scheme [5] of MVC is shown in Fig. 1, and it uses the hierarchical B prediction structure [6] for each view. All of the views can be classified into two categories; main view (such as S0, S2, S4, and S6) which mainly using temporal prediction, and auxiliary view (such as S1, S3, S5, and S7) which can be referred to the neighboring main view. In Fig. 1, the frame S1/T2 which means the frame of auxiliary view1 at time T2 has 4 reference frames, while the frame S2/T2 in main view2 has only 2 reference frames. Thus, auxiliary views are much more complex than main views. The Group of Picture (GOP) length in Fig. 1 is 8 for coding scheme explanation.

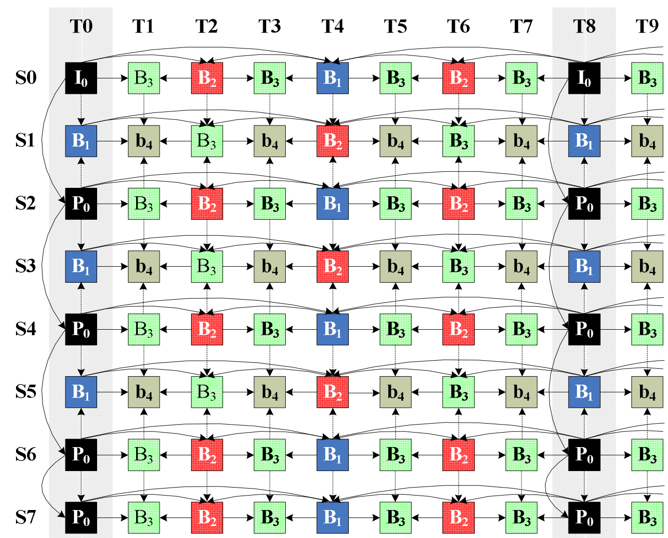


Figure 1. Multiview video coding prediction scheme in JMVM.

MVC uses Rate-Distortion Optimization (RDO) technique to select the optimal coding mode [7], and the Rate-Distortion cost (RD cost) is calculated as follows;

$$J = D + \lambda \cdot R \tag{1}$$

where J represents the RD cost of the current mode, D reflects the distortion between the original and the reconstructed Macroblock (MB), R represents the total bits of MB header, motion vectors, and DCT coefficients together, and λ is the Lagrange multiplier. There are many MB modes in MVC reference software JMVC 4.0 [8], including Skip, Direct, Inter16×16, Inter16×8, Inter8×16, Inter8×8, Inter8x8Frext, Inter8×4, Inter4×8, Inter4×4, Intra16×16, Intra8×8, Intra4×4, and Intra PCM. For each MB, the MVC encoder calculates the RD cost of all modes and selects the best mode having the minimum RD cost. Because the process of RDO is repeatedly carried out for each MB, the computational complexity of MVC is enormous.

Recently, some fast mode decision algorithms of MVC have been proposed. A mode pre-decision algorithm has been proposed for reduction of disparity estimation in [9]. A content-aware prediction algorithm with inter-view mode decision has been proposed in [10]. Yu et al. have proposed a fast MB mode decision algorithm based on inter-view global disparity in [11].

In this paper, a fast Inter mode decision is proposed for reduction of MVC computational complexity. The rest of paper is organized as follows: Section II analyses the feature of mode decision in MVC. In Section III, the detailed algorithm of the proposed approach is described. In this proposed approach, based on average RD cost of Skip mode in different textural regions of neighboring views, RD cost thresholds of Skip mode are calculated first. Then these thresholds are used to terminate mode decision early. In addition, the inter-view prediction for Inter sub modes is reduced based on the Inter16×16 mode estimation results. Section IV provides the experimental results. The conclusions are given in the section V.

II. Analysis of Mode Decision in MVC

The MVC reference software JMVC 4.0 is used to analyze the feature of mode decision, and the Intra mode decision is processed before Inter mode decision. The modes of MVC are classified into three categories; Skip mode, Inter mode, and Intra mode. Skip mode utilizes the motion vector predictor to predict current MB without performing estimation. It not only saves the motion vectors coding bits, but also reduces the computational complexity. Inter mode which includes 7 block size of inter frame prediction can refer to either temporal direction reference frame or inter-view direction reference frame. Intra mode which can also be classified into 3 block size predicts the current MB by utilizing boundary pixels of the neighboring MBs. The proportion of optimal modes after the RDO process for Exit sequence is shown in Table 1. From the analysis of modes proportion in Table I, it is found that most areas of video content are predicted by Skip and Inter modes, and the Intra mode only occupies a small proportion. The background of video which occupies a large proportion is more likely to be encoded as the Skip mode, while the encoding time for calculation of the Skip mode RD cost is small in comparison. Thus, if the Skip mode could be determined to be the optimal MB mode as earlier as possible, it would reduce a lot of estimation operation for other modes which consume most of the encoding time.

Table I indicates that there is almost the same proportion of the Skip mode at the same time around neighboring views. Because the textural feature of each region in video contents is different, the RD cost of Skip mode should be analyzed further based on the best Intra sub modes which indicate textural feature of video contents [12]. The average RD costs for Skip mode MBs with Intra16x16 as the best Intra mode, and with other Intra sub modes as the best Intra mode are both shown in Table II. From Table II, it can be seen that different Skip mode MB classifications have different average RD costs, and the average RD costs of the same classification in neighboring views are similar at the same time.

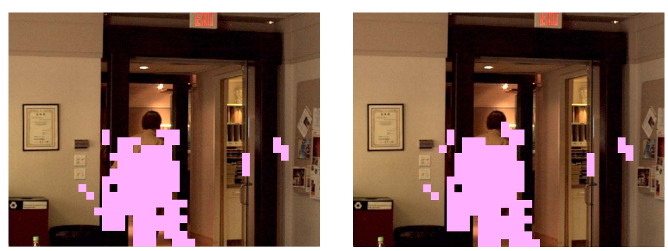
TABLE I. Proportion of Optimal MB Modes for Exit Sequence (QP32)

Frame Position	Skip Mode (%)	Inter Mode (%)	Intra Mode (%)
S0/T2	91.42	8.41	0.17
S1/T2	92.58	7.42	0.00
S2/T2	91.00	8.92	0.08
S0/T6	82.25	16.33	1.42
S1/T6	85.67	13.42	0.92
S2/T6	84.25	13.08	2.67

TABLE II. Average RD Cost of The Skip Mode According to Best Intra Sub Modes for Exit Sequence (QP32)

Frame Position	Skip Mode MB with Intra16×16	Skip Mode MB with other Intra modes
S0/T2	3057.21	7478.09
S1/T2	3327.86	6891.14
S2/T2	3047.26	6869.59
S0/T6	2697.80	6538.11
S1/T6	2949.30	5802.30
S2/T6	2692.52	5639.12

For auxiliary views, the estimation of Inter mode includes motion estimation for temporal reference frame and disparity estimation for inter-view reference frame. From our statistics, the Inter mode which selects the inter-view reference frame occupies only about a small proportion of all the MB counts, but the encoder consumes more than 50% of the encoding time to calculate the RD cost because of the disparity estimation. The distribution of inter-view direction selected by the Inter16×16 mode estimation is shown in Fig 2a, and by all Inter modes estimation is shown in Fig. 2b. From Fig. 2, it can be seen that distributions of these two cases are consistent.



(a) (b)

Figure 2. Distributions of inter-view direction for frame S1/T6 in Exit sequence. (a) Inter-view direction from Inter16×16 mode (b) Inter-view direction from all Inter modes

III. Proposed Fast Inter Mode Decision Algorithm

According to the analysis in the previous section, a Skip mode pre-decision is proposed to save unnecessary calculation by exploiting the correlation between neighboring views as well as the correlation between the textural regions and the RD cost. In addition, the estimated direction of Inter sub modes is predicted based on the optimal direction of the Inter16x16 mode.

A. Prediction of RD Cost Based on Textural Region for Skip Mode

In this proposed scheme, the Intra mode decision is processed before other modes, and the result of Intra mode decision is used to select different region types. From the point of H.264/AVC codec, the area with the high textural feature often selects Intra4×4 mode, and the homogeneous area which contains the low textural feature often selects Intra16×16 mode [12]. Thus, in order to distinguish the low and the high textural region, the result of Intra mode decision is used to segment textural regions first. Table III shows the proposed segmentation of low and high textural regions based on the type of best Intra sub mode. The distribution of different textural segmentation regions for Race1 sequence is shown in Fig.3. Fig.3 (a) is the original frame, and Fig.3 (b) is the distribution map of high textural region and low textural region. The black area of Fig.3 (b) is a high textural region, and the white area is a low textural region. For every MB, the textural region type is obtained after the Intra mode decision.

TABLE III. SEGMENTATION OF LOW AND HIGH TEXTURAL REGION

Intra mode	Intra16×16	Intra4×4, Intra8×8, Intra PCM
Region type	Low textural region (LT)	High textural region (HT)

(a) (b)

Figure 3. (a) Orginal frame of Race1 sequence (b) Low textural region in white and high textural region in black

Then, for the frame S_kT_t, the average RD cost of the Skip mode MBs in the high textural region is calculated as follows;

$$S_kT_t\_RD_{HT} = \frac{\sum_{i\in HT} J(i,m)\times B(i,m)}{\sum_{i\in HT} B(i,m)} \quad (2)$$

where *J(i, m)* denotes the RD cost for MB, and *B* is calculated in (3);

$$B(i,m)=\begin{cases}1 & m{=}{=}Skip\\ 0 & otherwise\end{cases} \quad (3)$$

where i is the index of MB in the frame S_kT_t and m represents the optimal MB mode.

Finally, RD cost threshold of the Skip mode in the high textural region for current frame is calculated according to $S_kT_t\_RD_{HT}$ in the neighboring views frames. For the frames in auxiliary views, the calculation of the threshold for current frame is as follows;

$$S_kT_t\_TH_{HT}=\alpha\times\left(S_{k-1}T_t\_RD_{HT}+S_{k+1}T_t\_RD_{HT}\right)/2 \quad (4)$$

Parameter α is inserted as a tradeoff between the performance and the efficiency of the approach, and 1.2 is selected from experimental results to keep the accuracy while reducing the complexity greatly. The low textural threshold can be obtained identically. Then the thresholds are used to terminate the mode decision early.

B. Reduce Inter-view Prediction Based on Inter16×16 Estimate Result

In this proposed scheme, the Inter16×16 mode decision which includes both temporal prediction and inter-view prediction is processed before other Inter sub modes. From the analysis in section II, there is similar inter-view predict direction selected by Inter16×16 mode and all Inter modes. If Inter16×16 mode employs inter-view direction, the Inter sub modes with variable block size are also likely to be encoded in the same direction. Thus, the estimated direction of Inter sub modes is predicted based on the optimal direction of Inter16×16 mode, and it is shown in Table IV. The Inter sub modes only employ inter-view prediction when the optimal prediction of Inter16×16 mode is the inter-view direction or a mixture of the inter-view direction and the temporal direction.

TABLE IV. THE PREDICT DIRECTION FOR INTER SUB MODES

Optimal predict direction of Inter16×16	**Predict direction of Inter sub modes**
Mix of temporal and inter-view prediction or only inter-view prediction	Both temporal and inter-view prediction is estimated
Only temporal prediction is selected	Only temporal prediction is estimated

C. Framework of the Proposed Fast Mode Decision

The framework of the proposed scheme is illustrated in Fig.4. First, based on the average Skip mode RD cost of the same textural segmentation region in neighboring views, thresholds of the Skip mode RD cost in different textural regions for the current frame are calculated as in (4). Then, the optimal mode is selected from Intra, Skip, or Inter16x16 mode based on their RD costs. The textural region type of the current MB is obtained from the best Intra sub mode in Table III. If the Skip mode were the optimal mode among the three estimated modes and the RD cost of Skip mode were smaller than the threshold, an early decision of the Skip mode could be made. When an auxiliary view is selected, the inter-view prediction of Inter sub modes is reduced based on the optimal direction of Inter16×16. To prevent prediction error propagation, the first encoded frame (anchor frame) of each view in a GOP is not included in our fast Inter mode decision scheme.

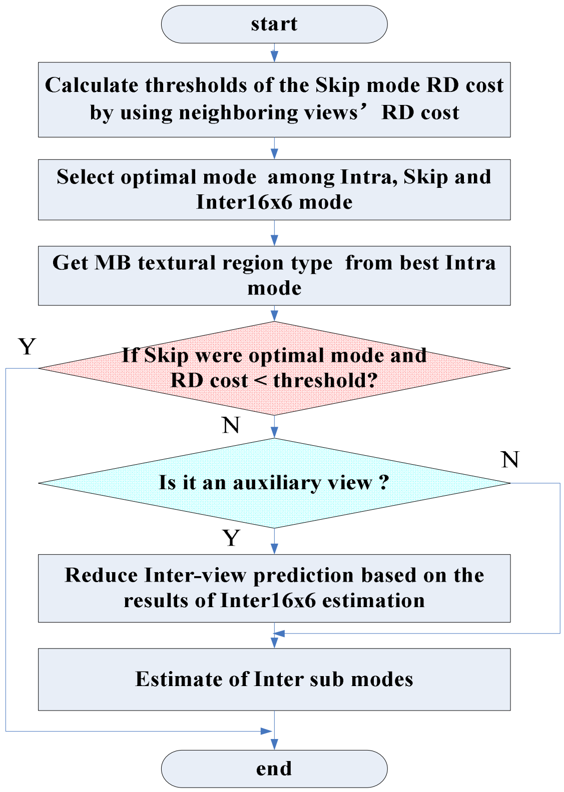


Figure 4. Flowchart of the proposed fast Inter mode decision for MVC

IV. EXPERIMENTAL RESULTS

The performance of our proposed algorithm is analyzed in this section. The proposed algorithm is implemented in MVC reference software JMVC 4.0 only for auxiliary views. The employed sequences are Ballroom, Exit, and Race1; 3 views with 4 GOPs are adopted for each of them. The sequences are tested according to the common test conditions for MVC [13]. Comparing to the full MB mode decision in JMVC 4.0, the average deviation of encoding time, Peak Signal to Noise Ratio (PSNR), and bits in the auxiliary view 1 using inter-view prediction (e.g. S1 in Fig.1) are listed. As shown in Table V, the proposed method reduces the average encoding time about 55% with the average PSNR degradation of less than 0.01dB.

V. CONCLUSIONS

In this paper, a fast Inter mode decision algorithm for the Skip mode and Inter sub modes is proposed in order to reduce the computational complexity of the encoding process. Based on the RD cost correlation between neighboring views, and the RD cost of different textural segmentation regions, an early Skip mode decision is proposed. In addition, the inter-view prediction for Inter sub modes is reduced after Inter16x16 mode estimation is processed. The prediction result of Inter16x16 mode estimation is used to predict the Inter sub modes' estimation direction. For auxiliary views, the experimental results show that the proposed scheme achieves about 55% reduction of the encoding time with negligible quality degradation.

TABLE V. PERFORMANCE COMPARISON OF PROPOSED ALGORITHM

QP	Performance	Exit	Ballroom	Race1
22	ΔPSNR(dB)	-0.0028	-0.0074	-0.0002
	ΔBits(%)	0.03	0.29	0.14
	ΔTime(%)	-67.21	-59.02	-53.80
27	ΔPSNR(dB)	-0.0019	-0.0061	-0.0025
	ΔBits(%)	-0.03	0.41	0.01
	ΔTime(%)	-70.68	-60.31	-55.43
32	ΔPSNR(dB)	-0.0043	-0.0006	-0.0003
	ΔBits(%)	-0.28	0.86	-0.20
	ΔTime(%)	-71.73	-59.84	-55.94
37	ΔPSNR(dB)	0.0041	-0.0068	-0.008
	ΔBits(%)	0.67	0.25	0.13
	ΔTime(%)	-71.89	-59.07	-55.08
Average	ΔPSNR(dB)	-0.0012	-0.0052	-0.0028
	ΔBits(%)	0.10	0.45	0.02
	ΔTime(%)	-70.38	-59.56	-55.06

REFERENCES

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

[2] M. Tanimoto, "Free viewpoint television – FTV," in Proceedings of 2004 Picture Coding Symposium, Dec. 2004.

[3] K. Mueller, P. Merkle, A. Smolic, and T. Wiegand, "Multiview coding using AVC," ISO/IEC JTC1/SC29/WG11, Bangkok, Thailand, Doc. M12945, January 2006.

[4] G. Sullivan, T. Wiegand, and A. Luthra, Draft of Version 4 of H.264/AVC (ITU-T Recommendations H.264 and ISO/IEC 14496-10 (MPEG-4 Part 10) Advanced Video Coding), ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-N050d1, 2005.

[5] 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, April 2008.

[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. 11, pp. 74–90, November 1998.

[8] 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, Januray 2009.

[9] 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, November 2006.

[10] 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, December 2008.

[11] M. Yu, Z. J. Peng, W. Y. Liu, F. Shao, G. Y. Jiang and Y. D. Kim "Fast macroblock selection algorithm for multiview video coding based on inter-view global disparity," Congress on Image and Signal Processing, vol. 1, pp. 575-578, May 2008.

[12] 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.

[13] 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, October 2006.