A Low Complexity Multi-view Video Encoder Exploiting B-frame Characteristics

Yuan-Hsiang Miao[1], Guo-An Jian[1], Li-Ching Wang[2], Jui-Sheng Lee[1], and Jiun-In Guo[1]
[1]Department of Electronics Engineering, National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
E-mail: {yuanhsiang, guoan, jslee.ee00g, jiguo}@nctu.edu.tw
[2]Department of Computer Science and Information Engineering, National Chung Cheng University, Chia-Yi, Taiwan, R.O.C.
E-mail: wwllccrr@yahoo.com.tw

Abstract— **This paper proposes a low complexity multi-view video encoder which includes mode decision and early termination based on B-frame characteristics. According to the statistics of coding mode distribution in different B-frame types, we classify all the coding modes into several classes and propose an early terminated mode decision algorithm that can largely reduce the computing complexity. On the other hand, MVD-based adaptive search range scheme is also included in the proposed encoding strategy. In our experimental results, the encoding time is saved up to 91% - 93% but the quality loss is controlled within 0.1 dB PSNR drop.**

I. INTRODUCTION

With the development of the brand-new video technology like three-dimensional television (3DTV) and free-viewpoint television (FTV), multi-view videos become more and more popular. Multi-view videos consist of many views, captured by different cameras at the same time. However, such enormous data brought by multi-view videos will be a disaster when storing or transferring it. Therefore, multi-view video coding (MVC) [1] has been widely investigated in recent years. We classified these works in several categories: (1) adaptive search range algorithm [2-4]; (2) early termination algorithm [5-6]; (3) fast mode decision algorithm [7-13]; and (4) ME/DE selection algorithm [14-16].

Fig. 1 illustrates an example of MVC structure that has three views. Within each view, hierarchical B-pictures are used to achieve high-density compression in temporal domain. For each group of pictures (GOP) under hierarchical B-pictures, the first frame is named anchor frame while the rest are named non-anchor frames. Moreover, the views in MVC structure can also be classified into three types, including I-view, P-view, and B-view, based on the type of the anchor frames. Among different views, inter-view prediction is employed to exploit spatial correlation.

For non-anchor frames in B-view, their complexity is much higher than that of non-anchor frames in other views because they have both the temporal and spatial references and it is why most of MVC researches focus on the complexity reduction of B-view encoding. From the referencing mechanism defined in MVC, we discovered that the determination of referencing for each non-anchor frame in B-view is highly related to the temporal position of its reference frames. In other words, using respective encoding strategies for distinct B-frames could be much more efficient than using a shared strategy. Therefore, we analyze the characteristics of B-frames and propose a low complexity multi-view video encoder including early terminated mode decision and MVD-based adaptive search range. The experimental results show that the encoding time is saved up to 91% - 93% but the quality loss is controlled within 0.1 dB PSNR drop.

This paper makes the following contribution:

(1) An early terminated mode decision for MVC encoding based on B-frame characteristics.
(2) An adaptive search range scheme based on MVD.

The remainder of this paper is organized as follows: Section II illustrates how B-frames work in multi-view video and analyzes the characteristics of them. Section III describe the proposed low complexity multi-view encoding strategy including early terminated mode decision and MVD-based adaptive search range. Section IV reports the simulation results of the proposed low complexity multi-view encoder and compares them with the ones of some existing researches. Finally, this paper is concluded in Section V.

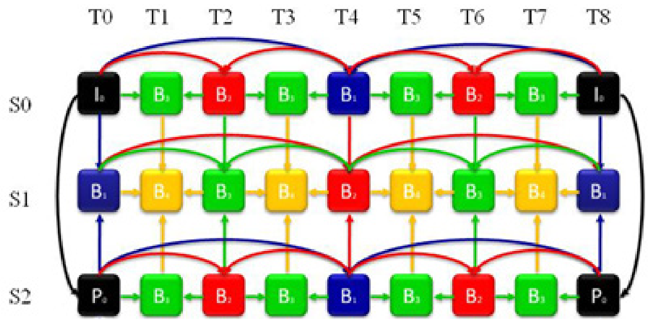

Fig. 1. Example of MVC structure (three views, GOP size = 8).

II. ANALYSIS OF B-FRAMES IN MULTI-VIEW VIDEO

According to the MVC structure, B-frames can be classified into four types including B1, B2, B3, and B4. B1 is the anchor frame encoded without temporal reference in B-view. B2, B3, and B4 are the non-anchor frames encoded with both temporal and spatial references. For non-anchor frames, the distance of temporal reference frames has a great effect upon the predicting result. If the GOP size is 8, the distance of temporal reference frames is 4 in B2, 2 in B3, and 1 in B4. In the ideal situation, the percentage of temporal prediction would increase from B2 to B4 due to the decrease in reference distance.

To verify this thought, we analyze the distribution of temporal and spatial references for different B-frame types. Here we use the conventional JMVC to perform encoding based on full search. As shown in Fig. 2, we mark MV (Motion Vector), DV (Disparity Vector), Bi-directional MV, and INTRA (Intra Prediction) for each macroblock. Clearly, the percentage of temporal prediction truly increases from B2 to B4. At the same time, we also observe that most of macroblocks are encoded with bi-directional prediction. Table I proves this observation with quantitative data. So we make a further investigation on bi-directional prediction. Table II and Table III tell us that most of macroblocks encoded with bi-directional prediction select SKIP mode in temporal reference. This means we should put SKIP mode of temporal bi-directional prediction at the first place of early termination scheme. On the other hand, we also analyze the distribution of each coding mode under uni-directional prediction. As summarized in Table IV, we can classify these coding modes into several classes and add early termination scheme to them for improving the efficiency of mode decision.

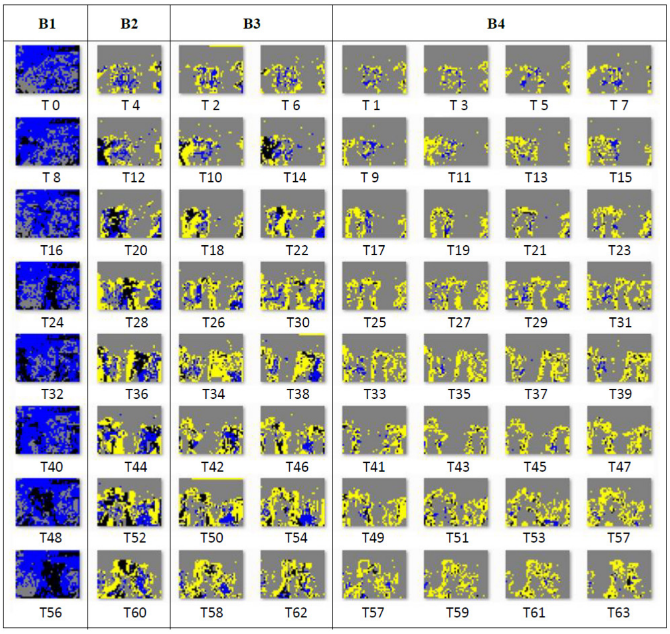

Fig. 2. Distribution of temporal and spatial references under different B-frame types. (Yellow: MV, Blue: DV, Grey: Bi-directional MV, and Black: INTRA)

TABLE I. DISTRIBUTION OF BI-DIRECTIONAL AND UNI-DIRECTIONAL PREDICTION.

	B2		B3		B4	
	(MBs)	(%)	(MBs)	(%)	(MBs)	(%)
Bi-directional prediction	26479	74.38	57497	78.86	126629	85.64
Uni-directional prediction	9121	25.62	15413	21.14	21233	14.36

TABLE II. DISTRIBUTION OF TEMPROAL AND SPATIAL REFERENCE IN BI-DIRECTIONAL PREDICTION.

	B2		B3		B4	
	(MBs)	(%)	(MBs)	(%)	(MBs)	(%)
MV+MV	23996	90.64	54760	95.25	124412	98.25
MV+DV	1030	3.88	1358	2.36	1453	1.15
DV+DV	1453	5.48	1379	2.39	764	0.60

TABLE III. DISTRIBUTION OF CODING MODES IN TEMPORAL PREDICTION.

	B2		B3		B4	
	(MBs)	(%)	(MBs)	(%)	(MBs)	(%)
MODE_SKIP	24683	93.22	54494	94.78	121583	96.02
MODE_16×16	1080	4.08	1777	3.09	3064	2.42
MODE_16×8	200	0.76	299	0.52	489	0.39
MODE_8×16	255	0.96	502	0.87	798	0.63
MODE_8×8	261	0.99	425	0.74	695	0.55

TABLE IV. DISTRIBUTION OF GROUPED CODING MODES.

	B2		B3		B4	
	(MBs)	(%)	(MBs)	(%)	(MBs)	(%)
MV_SKIP MV_16×16	4200	46.05	7800	50.61	10860	51.15
MV_16×8 MV_8×16	1187	13.02	2671	17.33	5296	24.94
MV_8×8	176	1.93	150	2.92	685	3.23
DV_SKIP DV_16×16	2276	24.95	3026	19.63	3159	14.88
DV_16×8 DV_8×16	1120	12.28	1297	8.41	1120	5.28
DV_8×8	162	1.78	169	1.10	113	0.53

III. PROPOSED LOW COMPLEXITY MULTI-VIEW ENCODER

A. *Early Terminated Mode Decision*

Based on the analysis in Section II, we classify all the coding modes into six mode classes as shown in Table V. Then we arrange them according to the possibility of being chosen as the best mode. Bi-directional prediction is only used in class A but other classes perform uni-directional prediction. With these mode classes, we propose an early terminated mode decision as shown in Fig. 3. The proposed scheme is only applied to non-anchor frames under B-view. In other words, B1 frames will be encoded just as they are encoded in conventional MVC encoder. For B2 and B3 frames, the proposed mode decision will go through these six mode classes in order from A to F. We put temporal bi-directional prediction at the first stage since 74% - 85% of macroblocks select it as the best mode. However, the order between the mode class C and D will be exchanged for B4 frames because the possibility for class C is smaller than class D even though it is larger inside B2 and B3 frames. After the execution order of each mode class is arranged, we add early termination into the proposed mode decision so as to avoid unnecessary computation. But thresholds are distinct under different mode classes and B-frame types. Therefore, we refresh the thresholds after the anchor frame is encoded and apply them to the non-anchor frames in the same GOP. The calculation of these thresholds will be explained in detail in the following section.

TABLE V. MODE CLASSES DEFINED IN THE PROPOSED MODE DECISION.

A	B	C	D	E	F
Bi-DIR MV_SKIP	MV_SKIP MV_16×16	DV_SKIP DV_16×16	MV_16×16 MV_8×16	DV_16×16 DV_8×16	Others

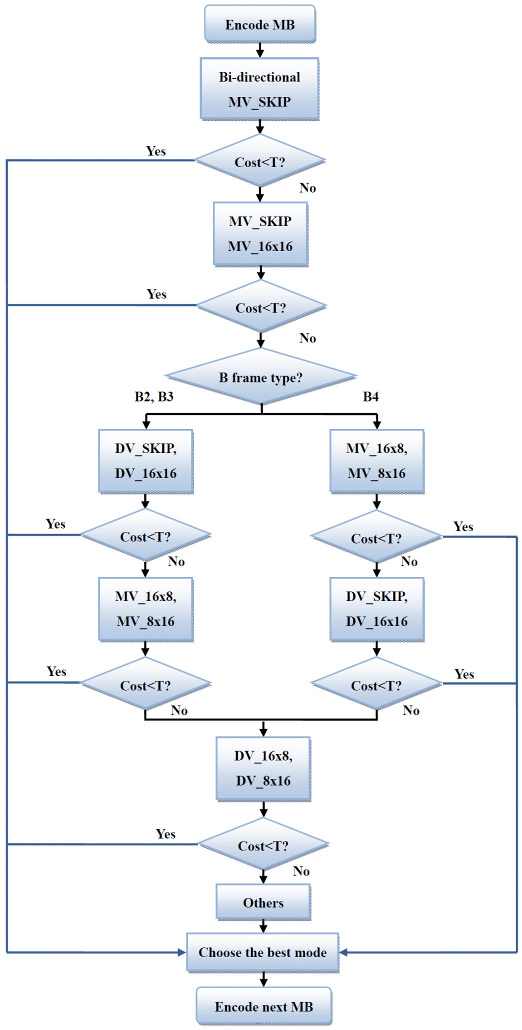

Fig. 3. Flowchart of the proposed early terminated mode decsion.

B. *Thresholds for Early Termination*

In the proposed early terminated mode decision, thresholds play important roles because they dominate the balance between the performance improvement and the quality loss. To get appropriate thresholds, we analyze the average RDCost (Rate-Distortion Cost) for each mode class under different B-frame types. For each mode class, we take the average RDCost of anchor frame (B1) as a basis because the proposed early termination is only applied to the non-anchor frames (B2, B3, and B4). Table VI shows the final thresholds we choose for each mode class under different B-frame types. Therefore, all the thresholds depend on the average RDCost of anchor frame and have to be refreshed in the beginning of every GOP.

TABLE VI. THRESHOLDS FOR THE PROPOSED EARLY TERMINATION SCHEME. (T IS THE AVERAGE RDCOST OF ANCHOR FRAME.)

Mode Class	A	B	C	D	E	F
B1	—	—	—	—	—	—
B2	0.60*T	0.88*T	1.61*T	1.49*T	1.99*T	—
B3	0.65*T	1.01*T	1.89*T	1.69*T	2.17*T	—
B4	0.69*T	1.24*T	2.11*T	1.76*T	2.40*T	—

C. *MVD-based Adaptive Search Range*

In order to reduce the redundant computation of motion estimation, we also propose a simple search range adjusting method using motion vector differences (MVD). For each macroblock, its MVD means the difference between the motion vector and the motion vector predictor and it can be used as a basis to check if the search range needs to be adjusted. To relieve the overhead brought by dynamic adjustment of search range, we adjust the search range once for each frame rather than adjust it for each macroblock. Therefore, MVDs of all the macroblocks within the same frame are recorded at first. After encoding the entire frame, the search range is updated with a new value that must be a multiple of 16 and larger than the average of the non-zero MVDs. Finally, the next frame will be encoded according to the updated search range.

IV. Performance Evaluation

In this section, we evaluate the complexity reduction and quality measurement of the proposed low complexity MVC encoder. Table VII lists the hardware environment and software settings. Tested videos, including Ballroom, Ballet, Race1, and Exit, are in VGA resolution and widely used in related research area. We encode these videos with the conventional JMVC and the proposed low complexity encoder, respectively. Table VIII summarizes overall performance and quality measurement of the proposed low complexity MVC encoder. Moreover, we also compare the results with the ones of the conventional JMVC encoder under the same environment. Table IX compares the complexity reduction and quality loss between the proposed encoder and other works. Clearly, the proposed low complexity MVC encoder can save up to 91% - 93% of encoding time but the quality loss is controlled within 0.1 dB.

TABLE VII. CONFIGURATION OF THE EXPERIMENTAL ENVIRONMENT.

Hardware Environment	**CPU Type**	Intel Xeon E5420 @ 2.50GHz
	Memory	32 GB
Software Settings	**Sequence**	Ballroom, Race1, Exit
	Resolution	VGA
	GOP Size	8
	Search Mode	4 (Fast Search)
	Search Range	Adaptive
	BasisQP	32

TABLE VIII. OVERALL PERFORMANCE AND QUALITY MEASUREMENT.

Sequence	Method	Time (sec.)	PSNR (dB)
Ballroom	JMVC	7281	38.6224
	Proposed	600	38.5751
	Comparison	-91.76%	-0.0473
Ballet	JMVC	2312	43.5326
	Proposed	183	43.4726
	Comparison	-92.08%	-0.0600
Race1	JMVC	9551	38.2252
	Proposed	615	38.1498
	Comparison	-93.56%	-0.0754
Exit	JMVC	7046	40.6916
	Proposed	499	40.6084
	Comparison	-92.92%	-0.0832

TABLE IX. COMPARISON OF REDUCED ENCODING TIME AND PSNR DROP.

Sequence	Method	Encoding time reduction (%)	PSNR drop (dB)
Ballroom	Proposed	91.76	0.0473
	[4]	75.81	0.08
	[6]	72.57	0.19
	[10]	66.20	0.10
	[15]	59.84	0.0006
	[16]	74.05	0.06
	[17]	78.88	0.09
	[18]	49.63	0.14
Race1	Proposed	93.56	0.0754
	[4]	89.22	0.01
	[6]	66.02	0.13
	[10]	73.77	0.11
	[13]	67.01	0.07
	[15]	55.94	0.0003
	[16]	76.36	0.06
	[17]	88.68	0.04
	[18]	35.88	0.18
Exit	Proposed	92.92	0.0832
	[4]	85.51	0.08
	[6]	71.03	0.12
	[10]	78.50	0.10
	[15]	71.73	0.0043
	[16]	83.04	0.09
	[17]	83.68	0.08
	[18]	36.01	0.14

V. CONCLUSION

In this paper, we have proposed a low complexity multi-view video encoder which includes early terminated mode decision and MVD-based adaptive search range. Based on the statistical analysis of B-frame characteristics, the proposed early terminated mode decision can largely reduce the computing complexity. Moreover, the proposed MVD-based adaptive search range scheme also helps us reduce the redundant computation within motion estimation. The experimental results show that the encoding time is saved up to 91% - 93% but the quality loss is controlled within 0.1 dB PSNR drop.

ACKNOWLEDGMENT

This work was sponsored by the National Science Council of Taiwan under grants NSC99-2221-E-009-194-MY3 and NSC-100-2221-E-194-034-MY2.

REFERENCES

[1] Y. S. Ho and K. J. Oh, "Overview of Multi-view Video Coding", *Proc. International Workshop on Systems, Signals and Image Processing*, pp. 5-12, Jun. 2007.

[2] J. Kim and K. Sohn, "Fast Disparity and Motion Estimation for Multi-view Video Coding", *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 712-719, May 2007.

[3] X. Xu and Y. He, "Fast Disparity Motion Estimation in MVC Based on Range Prediction", *Proc. IEEE International Conference on Image Processing*, pp. 2000-2003, Oct. 2008.

[4] P. He, M. Yu, Z. Peng, and G. Jiang, "Fast Mode Selection and Disparity Estimation for Multiview Video Coding", *Proc. International Symposium on Intelligent Information Technology Application Workshops*, pp. 209-212, Nov. 2009.

[5] X. Li, D. Zhao, S. Ma and W. Gao, "Fast Disparity and Motion Estimation Based on Correlations for Multiview Video Coding", *IEEE Transactions on Consumer Electronics*, vol. 54, no. 4, pp. 2037-2044, Nov. 2008.

[6] M. Yu, Z. Peng, W. Liu, F. Shao, G. Jiang, and Y. D. Kim, "Fast Macroblock Selection Algorithm for Multiview Video Coding Based on Inter-view Global Disparity", *Proc. International Congress on Image and Signal Processing*, vol. 1, pp.575-578, May 2008.

[7] D. H. Han and Y. L. Lee, "Fast Mode Decision Using Global Disparity Vector for Multiview Video Coding", *Proc. International Conference on Future Generation Communication and Networking Symposia*, vol. 3, pp. 209-213, Dec. 2008.

[8] Z. Peng, G. Jiang, and M. Yu, "A Fast Multiview Video Coding algorithm based on Dynamic Multi-threshold", *Proc. IEEE International Conference on Multimedia and Expo*, pp. 113-116, Jul. 2009.

[9] L. Shen, T. Yan, Z. Liu, Z. Zhang, P. An, and L. Yang, "Fast Mode Decision for Multiview Video Coding", *Proc. IEEE International Conference on Image Processing*, pp. 2953 -2956, Nov. 2009.

[10] T. Y. Kuo, Y. Y. Lai, and Y. C. Lo, "Fast Mode Decision for Non-Anchor Picture in Multiview Video Coding", *Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pp. 1-5, Mar. 2010.

[11] M. Ai and J. Wang, "A Fast Mode Decision Algorithm for Multiview Video Coding", *Proc. International Congress on Image and Signal Processing*, vol. 7, pp. 3252-3257, Oct. 2010.

[12] P. Gao, Q. Peng, Q. Wang, X. Liu, and C. Zhang, "Adaptive Disparity and Motion Estimation for Multiview Video Coding", *Proc. International Congress on Image and Signal Processing*, vol. 1, pp. 66-71, Oct. 2011.

[13] P. J. Lee, H. J. Lin, S. H. Huang, and W. J. Wang, "A Fast Mode Determination Algorithm for Multi-view Video Coding", *Proc. IEEE International Conference on Consumer Electronics*, pp. 689-690, Jan. 2011.

[14] J. P. Lin and A. C. W. Tang, "A Fast Direction Predictor of Inter Frame Prediction for Multi-View Video Coding", *Proc. IEEE International Symposium on Circuits and Systems*, pp. 2589-2592, May 2009.

[15] W. Zhu, W. Jiang, and Y. Chen, "A Fast Inter Mode Decision for Multiview Video Coding", *Proc. International Conference on Information Engineering and Computer Science*, pp. 1-4, Dec. 2009.

[16] H. Zeng, K. K. Ma, and C. Cai, "Fast Mode Decision for Multiview Video Coding Using Mode Correlation", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 11, pp. 1659-1666, Nov. 2011.

[17] Z. Peng, G. Jiang, M. Yu, W. Liu, and Q. Dai, "Fast Macroblock Mode Selection Algorithm for Multiview Video Coding", *EURASIP Journal on Image and Video Processing*, vol. 2008, no. 393727, pp. 1-14, Oct. 2008.

[18] A. S. Akbari, N. Canagarajah, D. Redmill, D. Bull, and D. Agrafiotis, "A Novel H.264/AVC Based Multi-View Video Coding Scheme", *Proc. 3DTV Conference*, pp. 1-4, May 2007.