

# New Multi-reference Frame Selection for Multiview Video Coding

Yuehou Si<sup>1</sup>, Mei Yu<sup>1</sup>, Zongju Peng<sup>1</sup>, and Gangyi Jiang<sup>1,2</sup>

<sup>1</sup> Faculty of Information Science and Technology, Ningbo University, Ningbo, 315211, China  
Email: siyuehou444@163.com, yumei2@126.com, pengzongju@nbu.edu.cn

<sup>2</sup> National Key Lab of Software New Technology, Nanjing University, Nanjing, 210093, China  
Email: jianggangyi@126.com

**Abstract**—With the development of multimedia technologies and the demand for three-dimensional video systems, multiview video coding (MVC) has attracted great attention from industries and research institutes. A joint multiview video model (JMVM) with multi-reference frame technology had been developed, in which motion and disparity estimation for eliminating temporal and inter-view redundancies are employed in multi-reference frame technology to enhance the coding efficiency. However, the process of searching blocks with variable sizes for motion and disparity estimation in multi-reference frames significantly requires much more memory bandwidth and computational complexity relative to mono-view video coding systems. On the study of the statistical features of multi-reference frames in MVC, in this paper we propose a new fast multi-reference frame selection algorithm based on an adaptive threshold technology. The proposed algorithm can terminate the process of searching multi-reference frames early. An adaptive threshold technique is given and correlated with the types of frames. Experimental results show that the proposed algorithm can promote the encoding speed by 2.01~3.60 times without noticeable quality degradation compared with the JMVM 7.0.

**Index Terms**—multiview video coding; multi-reference frame; motion and disparity estimation; adaptive threshold

## I. INTRODUCTION

With the development of multimedia technologies and the demand for realistic visual systems [1], three dimensional (3D) video technologies have been widely researched and gradually used. With the technology of free-viewpoint television (FTV) [2], 3D television [3] and surveillance systems [4], multiview video coding (MVC) draws more and more attention. Multiview video are captured simultaneously at different positions and angles by multiview imaging system[5]-[6]. The straightforward solution for MVC is to encode all the videos independently by using the existing state-of-the-art video codecs such as H.264/AVC[7]-[9]. Multiview video signals contain a large amount of inter-view redundancies and temporal redundancy, since all cameras capture the

same scene from different viewpoints simultaneously [10]. The joint multiview video model (JMVM) is developed by the Joint Video Team (JVT) as the reference software and the research platform, and it uses Hierarchical B Pictures (HBP) prediction structure to exploit both temporal and inter-view correlations [11]. JMVM provides various sophisticated coding techniques for MVC, in which multi-reference frame search technology is exploited to encode and decide the current frame with the found optimal reference frame. However, these technologies will increase the computational complexity significantly. Hence, it is urgent to propose a fast algorithm to improve the encoding efficiency.

There had been many fast algorithms for mono-view proposed about mode selection and multi-reference frame selection recently[12]-[14]. However, these fast algorithms for mono-view video coding can not be used directly for MVC. In addition, some fast algorithms for MVC had been proposed [15]-[16]. A fast macroblock mode selection algorithm for multiview video coding which had been proposed in [15] used the correlations of the macroblock modes during the neighboring views to simplify the process of searching all kinds of macroblock modes. In [16], a fast prediction algorithm, content-aware prediction algorithm with inter-view mode decision, is proposed.

The remaining of this paper is organized as follow. In section II, we first describe principle of multi-reference frame selection algorithm in JMVM and investigate why searching all reference frames is unnecessary. Then we propose a new fast multiple reference frame selection algorithm based on dynamic threshold. Experimental results will be given in Section III and the work is concluded in Section IV.

## II. FAST MULTIPLE REFERENCE FRAME SELECTION ALGORITHM

### A. Multi-reference frame selection in JMVM

Motion estimation (ME) and disparity estimation (DE) involve searching an area in a reference frame encoded, which closely matches the current block. The reference frame with the minimal sum-absolute-difference (SAD) value is then selected as the best reference frame for inter frame coding. SAD value is calculated by

Manuscript received July 30, 2009; revised Sept. 7, 2009; accepted Oct., 2009. This work was supported by Natural Science Foundation of China (Grant No:60672073, 60872094, 60832003), and the National 863 High Technology Research and Development Program of China (Grant No:2009AA01Z327). Mei YU is corresponding author, yumei2@126.com.

$$SAD(s, c) = \sum_{i=1}^{B_1, B_2} |s[i, j] - c[i - m_x, j - m_y]|^2 \quad (1)$$

where  $s$  and  $c$  denote the source and reconstructed signals, respectively,  $B_1$  and  $B_2$  denote the numbers of horizontal and vertical pixels.

In JMVM, HBP is used as a prediction structure for MVC, as shown in Fig. 1, where the arrows indicate the prediction directions of ME or DE. Here,  $S_n$  denotes the individual view and  $T_n$  is the consecutive time instant. For instance,  $S_{I0}$ ,  $S_{I8}$ ,  $S_{O4}$  and  $S_{2T4}$  are the reference frames for  $S_{I4}$ . In implementation, the referring relations in HBP structure can be divided into two categories, that is, the inter-view referring relation and the temporal referring relation. The proportion of B-frames is very high in HBP structure, since B-frames can take advantage of ME and DE to reduce the temporal and spatial redundancy. Although MVC is an emerging technology, huge amount of video data and ultra high computational complexity make it difficult to be realized. Hence, it is necessary to develop a fast algorithm to reduce computational complexity of encoding B-frames.

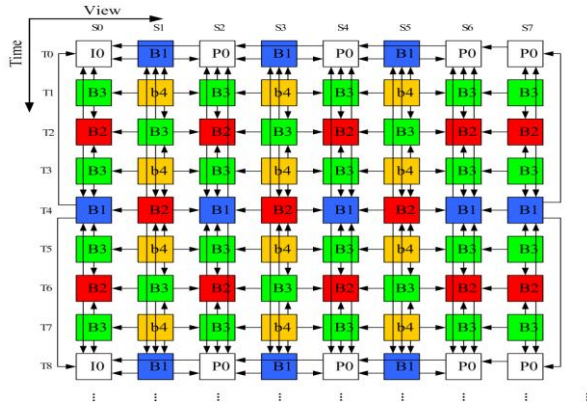


Figure 1. Illustration of an MVC structure

Decoded Picture Buffer (DPB) in JMVM contains the previously decoded  $N$  frames in order,  $Ref_i$ , ( $i = 1, 2, \dots, N$ ). Generally, the probability of getting the best macroblock match in  $Ref_1$  is higher than other reference frames. Therefore, searching multiple reference frames to find the best reference frame is inefficient.

### B. Analyses the statistical feature of multi-reference frames in MVC

As we know, the distribution of the reference blocks is not uniform across reference frames. To verify this phenomenon, we carry out the following experiments. Three test sequences are used, including Ballroom sequence which has great disparity and violent motion, Exit sequence which has great disparity, Race1 sequence which has violent motion. According to the analyses in last section that the majority of frames in HBP prediction structure are B-frames, we have done motion search of all the candidate blocks with two forward reference frames and two backward reference frames. Table I shows the average proportion of the optimal macroblocks found in

each reference frame. Here, list0 means the list of forward reference frames and list1 means the list of backward reference frames.

TABLE I.  
TYPE SIZES FOR CAMERA-READY PAPERS

	Race1		Exit		Ballroom	
	$Ref_1$	$Ref_2$	$Ref_1$	$Ref_2$	$Ref_1$	$Ref_2$
list0	98.73%	1.27%	94.11%	5.89%	90.72%	9.28%
list1	98.80%	1.20%	94.61%	5.39%	91.84%	8.16%

Obviously, the probability of getting the best matched block in  $Ref_1$  is high and decreases very fast to  $Ref_2$ . This suggests that we have high probability to find the optimal matched block early in the process of searching the first frame in DPB, and thus may terminate the process of ME/DE early without searching all  $Ref_i$ , ( $i = 1, 2, \dots, N$ ).

When a B-frame is encoded, three motion prediction methods are used in JMVM, that is, forward prediction, backward prediction and bi-directional prediction. We select the frame  $S_{O4}$  of Exit sequence to investigate the motion prediction methods of the JMVM. As shown in Fig. 2, the area with black rim denotes the macroblocks using the forward or backward prediction method, and the blocks with red border use bi-directional prediction.

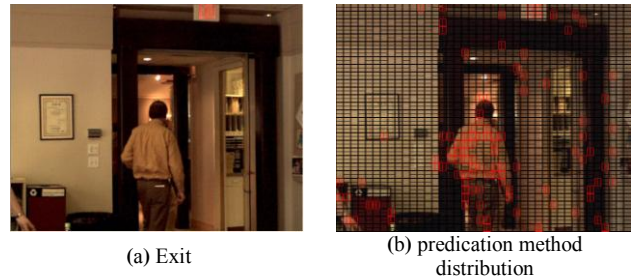


Figure 2. The distribution of macroblocks from prediction method is Bi-direction in red area of (b)

Obviously, the macroblocks with red border mainly concentrate in the area with violent motion, such as the man who is walking and the doorframe caused by the changes of man's shadow. It suggests that bi-directional prediction is used when macroblocks with violent motion are encoded. So it is possible to find the best matched block early in the process of forward prediction and backward prediction, and there is unnecessary to implement bi-directional prediction in flat areas.

Based on the analyses of the reference frames selection process of the JMVM, a fast multi-reference frames selection algorithm is proposed to reduce the computational complexity in MVC.

### C. The proposed fast multi-reference frames selection algorithm

For HBP prediction structure, eight views are categorized into three types, the basic view ( $S_0$ ), the first layer views ( $S_2, S_4, S_6$ ), the second layer views ( $S_1, S_3, S_5, S_7$ ). So we can divide all the frames in HBP structure into three categories, that is,  $C_1$ ,  $C_2$  and  $C_3$ .  $C_1$  denotes the anchor frame without any reference frames in the eight views,  $C_2$  is non-anchor frame in the base view and

the first layer view, and  $C3$  is the non-anchor frame in other views.

We can make use of the correlations of SAD values of blocks previously encoded to design the dynamic threshold recorded as  $T$ . The dynamic threshold is calculated by

$$T = \begin{cases} \text{median}(Rd_A, Rd_B, Rd_C) & C(n) \in \{C2\} \\ \min(Rd_v, Rd_t) & C(n) \in \{C3\} \end{cases} \quad (2)$$

where  $Rd_A$ ,  $Rd_B$  and  $Rd_C$  denote SAD values of the neighboring left-top block, top block and top-right (or top-left) block of the current block, and  $\text{median}(Rd_A, Rd_B, Rd_C)$  denotes the median of  $Rd_A$ ,  $Rd_B$  and  $Rd_C$ . And  $\min(Rd_v, Rd_t)$  is the minimal SAD value between the corresponding block at the same time and the block in neighboring views.  $C(n)$  denote the type of the current frame  $n$ . To analyze possibility of having bi-directional predication,  $S(V)$  is defined to describe the motion intensity of current encoded block. It is calculated by

$$S(V) = \{V \parallel |V| < 10\} \quad (3)$$

where  $V$  denotes motion vector of the current block, and  $S(V)$  represents a set of motion vectors whose mode less than 10. Here,  $V_A$ ,  $V_B$  and  $V_C$  denote the motion vector on the neighboring left block, top block and top-right (or top-left) block, respectively.

The above analysis motivates to develop a fast multi-reference frame selection algorithm based on dynamic threshold. Fig. 3 shows the flow chart of the proposed algorithm.

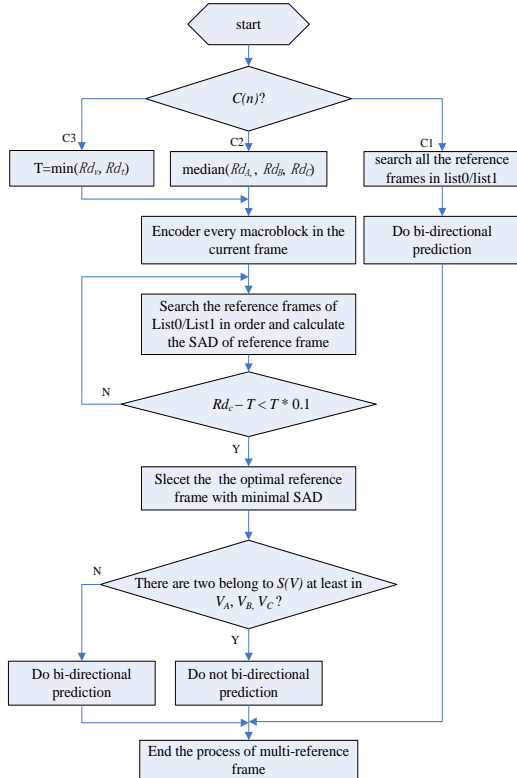


Figure 3. Flow chart of the proposed fast algorithm

### III. EXPERIMENTAL RESULTS AND ANALYSES

In experiments, the proposed fast multi-reference frame selection algorithm is evaluated compared with JMVM7.0. The experiments have been carried out on PC with the Intel Core 3.0GHz CPU and 3.25GB RAM. Test sequences are Door Flowers (1024×768), Alt Moabit (1024×768), Exit (640×480), Ballet (1024×768), Race1 (320×240) and Breakdancers (1024×768), respectively. Test configurations are listed in Table II. Table III shows the comparison experimental results of JMVM7.0 implementation and the proposed algorithm. In Table III,  $\Delta PSNR$ ,  $\Delta BR$  and  $\Delta T$  are defined by

$$\begin{aligned} \Delta PSNR &= PSNR_{proposed} - PSNR_{JMVM} \\ \Delta BR &= \frac{BR_{proposed} - BR_{JMVM}}{BR_{JMVM}} \times 100\% \\ \Delta T &= \frac{T_{JMVM} - T_{proposed}}{T_{JMVM}} \times 100\% \end{aligned} \quad (4)$$

where  $\Delta PSNR$  means the difference of JMVM and the proposed algorithm,  $\Delta BR$  reflects the bit rates increase or decrease with our algorithm.  $\Delta T$  reflects the encoding time saved by our algorithm.

TABLE II. TEST CONDITIONS

Encoded frames	41						
Basis QP	22, 27, 32, 37						
Delta Layer XQuant	0	1	2	3	4	5	
Delta QP Values	0	3	4	5	6	7	
GOP length	8						
Search range	±96						
SearchMode	Diamond search						

The proposed fast multi-reference frames selection algorithm reduces the useless candidate reference frames by filtering out those reference frames that are less likely to contain the best matched results. From Table III, it is clear that the proposed algorithm can reduce the encoding time ranging from 50.13% to 72.19% encoding time, while it hardly influences the RD performance. The real speedup performance of the proposed fast method may be better because the data listed in Table III include the encoding time of the anchor frames.

TABLE III. COMPARISON OF THE PROPOSED ALGORITHM AND JMVM7.0

QP	Compared terms	Door Flowers	Alt Moabit	Exit
22	$\Delta PSNR$ (dB)	-0.02	-0.02	0.00
	$\Delta BR$ (%)	0.20	0.30	0.14
	$\Delta T$ (%)	70.28	65.95	61.60
27	$\Delta PSNR$ (dB)	-0.30	-0.03	-0.03
	$\Delta BR$ (%)	0.12	0.84	1.03
	$\Delta T$ (%)	70.66	64.77	65.90
32	$\Delta PSNR$ (dB)	-0.03	-0.03	-0.04
	$\Delta BR$ (%)	-0.15	1.00	0.26
	$\Delta T$ (%)	70.34	64.22	67.18
37	$\Delta PSNR$ (dB)	-0.03	-0.02	-0.04
	$\Delta BR$ (%)	-0.03	0.48	0.17
	$\Delta T$ (%)	69.30	63.80	67.63

TABLE III. COMPARISON OF THE PROPOSED ALGORITHM AND JMVM7.0 (CONT.)

QP	Compared terms	Ballet	Race1	Breakdancers
22	$\Delta PSNR$ (dB)	-0.03	0.05	-0.02
	$\Delta BR$ (%)	1.36	3.40	2.43
	$\Delta T$ (%)	65.20	72.19	50.13
27	$\Delta PSNR$ (dB)	-0.03	0.04	-0.02
	$\Delta BR$ (%)	1.42	3.63	2.87
	$\Delta T$ (%)	66.59	71.22	53.96
32	$\Delta PSNR$ (dB)	-0.03	0.03	-0.03
	$\Delta BR$ (%)	0.99	4.75	3.15
	$\Delta T$ (%)	66.54	69.24	55.58
37	$\Delta PSNR$ (dB)	-0.03	0.01	-0.05
	$\Delta BR$ (%)	0.55	4.19	2.34
	$\Delta T$ (%)	66.02	66.62	56.73

#### IV. CONCLUSIONS

Multiview video coding (MVC) has attracted great attention from industries and research institutes. It is essential to design a fast multi-reference frames selection algorithm to reduce computational complexity of MVC. This paper presents an efficient fast algorithm for the prediction part in MVC. Based on high inter-view correlations between views and the distribution features of the optimal matched blocks, unnecessary computation of searching multi-reference frames can be early terminated. The features of the best reference frame distribution, SAD values of various modes are exploited as the basis of the proposed algorithm.

In this paper, we propose a fast multi-reference frame selection algorithm for MVC, which use a dynamic threshold to terminate the process of searching multi-reference frames early. Firstly, we have analyzed the statistical features of multi-reference frames in MVC and defined a dynamic threshold according to the type of frames. Secondly, motion intensity of the current encoded block is used to make sure if we can do bi-directional predication. Experimental results show that the proposed algorithm can reduce 62.63% coding time on average with the similar same coding quality.

#### REFERENCES

[1] A. Vetro, S. Yea, M. Zwicker, et al., "Overview of multiview video coding and anti-aliasing for 3D displays", *Proceedings of International Conference on Image Processing*, Texas, 2007, pp. 117-120.  
 [2] A. Smolic, K. Mueller, N. Stefanoski, et al., "Coding algorithms for 3DTV-A survey", *IEEE Transactions on*

*Circuits and Systems for Video Technology*, vol. 17, no. 11, 2007, pp. 1606-1621.

[3] Y. Sugihara, M. Tanimoto, T. Fujii, et al., "Requirements for FTV and 3DTV to Multi-view video coding (MVC)", *ISO/IEC JTC1/SC29/WG11*, M13169, Montreux, Switzerland, April 2006.  
 [4] H. Kalva, L. Christodoulou, L. Mayron, et al., "Challenges and opportunities in video coding for 3DTV", *Proceedings of IEEE International Conference on Multimedia and Expo*, Toronto, 2006, pp.1689-1692.  
 [5] B. S. Wilburn, M. Smulski, H.-H. K. Lee, and M. A. Horowitz, "Light field video camera," in *Proc. Media Processors, SPIE Electronic Imaging*, 2002, vol. 4674, pp. 29-36.  
 TABLE IV. C. Zhang and T. Chen, "A self-reconfigurable camera array," in *Proceedings of the 15th Eurographics Workshop on Rendering Techniques*, Sweden, June 2004, pp. 243-254.  
 [6] ITU-T Rec. H.264, "Advanced video coding for generic audiovisual services", *ITU-T Rec. H.264 - ISO/IEC 14496-10 AVC*, March 2005.  
 [7] T. Wiegand, G. J. Sullivan and G. Bjøntegaard, et al. "Overview of the H.264/AVC video coding standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, 2003, pp. 560-560.  
 [8] G. Sullivan and T. Wiegand, "Video compression—From concepts to the H.264/AVC standard", *Proceeding of The IEEE, Special Issue on Advances in Video Coding and Delivery*, vol. 93, no. 1, 2005, p. 18-31.  
 [9] Y. Zhang, G.Y. Jiang, M. Yu, et al., "An approach to multi-modal multi-view video coding", *Proceedings of Int. conf. on Signal Processing*, Guilin, China, Nov. 2006, pp.1401-1404.  
 [10] ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, "Joint Multiview Video Model (JMVM) 4.0", *JVT-W207*, San Jose, USA, April 2007.  
 [11] Z. Yang, L. Mo, J. Bu, C.Chen, "Fast multi-frame based motion estimation algorithm for MPEG-4 to H.264 video transcoder", *Journal of Zhejiang University (Engineering Science)* 42 (12), pp. 2055-2061.  
 [12] A. K. Mahajan, S. Kodayya, X. Su, "Exploiting Reference Frame History in H.264/AVC Motion Estimation", *IEEE Asia-Pacific Conference on Circuits and Systems, Proceedings*, artical no. 4145418, Singapore, Dec. 2006, pp. 410-413.  
 [13] Y. Ismail, M. Elgamel, M., Bayoumi, "A fast block-based motion estimation using early stop search techniques for H.264/AVC standard", *2007 IEEE North-East Workshop on Circuits and Systems*, artical no. 4487964, Montreal, Aug. 2007, pp. 397-400.  
 [14] G. Jiang, Z. Peng, M. Yu, Q. Dai, "Fast Macroblock Mode Selection Algorithm for Multiview Video Coding", *Eurasip Journal on Image and Video Processing*, vol. 2008, Article ID 393727, 2008, 14 pages.  
 [15] L. Ding, P. Tsung, S. Chien, W. Chen, and L. Chen, "Content-Aware Prediction Algorithm With Inter-View Mode Decision for Multiview Video Coding", *IEEE Transactions on Circuits and Multimedia*, vol.10, no.8, 2008, pp. 1553-1564.