Multi- view Video Coding Scheme based upon enhanced Random Access capacity

Xiaoxing Lv[1] Lini Ma [2] Jingjing Guo [3]

[1] School of Automation,Beijing Institute of Technology, Beijing, China 100081

[2] Computer Science School ,Beijing Information Science and Technology University,Beijing, China 100192

[3] Computer Science School ,Beijing Information Science and Technology University,Beijing, China 100192

Abstract

Due to the multi-view video coding scheme using inter-view prediction stucture, increased coding complexity, and reduced multi-view video random access performance, so one proposed multi-view video coding prediction scheme is proposed on the basis of analysis and study of several typical multi-view video coding schemes in this paper. This coding prediction scheme calculates the location of base-view by global disparity, and introduces a rational inter-view prediction structure, as well as, according the relationship between the length of GOP and random access performance to select the number of frames in a GOP. Experimental results show that the proposed multi-view video coding scheme can significantly improve the random access performance, while maintaining high coding efficiency.

Keywords: *Multi-view Video Coding; Global Disparity; Random Access; Hierarchical B Frames; Temporal Layer Identification*

1. Introduction

Multi-view video consists of several video sequences captured by multiple cameras which are aligned in a parallel, which includes the depth information of images and provides users with three-dimensional and interactive features to meet them watch video images from different angles. The MVC will be applied to a number of emerging multimedia services including free-viewpoint video (FVV), free-viewpoint television (FVT), three-dimensional video (3DV) and three-dimensional television (3DTV) [1].

Compared with the traditional single-view video coding, the multi-view video technology as an important area of research, which to be more comprehensive consideration to the dynamic scene and gives the immersive feel. Therefore, the amount of datas that need to be processed is also multiplied and which reduces the efficiency of video coding. However, there is a considerable temporal and spatial correlation among the various views, which offers the possibility for the coding efficiency. So, how to remove the correlation of intra-view and inter-view has become one of the hot issues in the field of multi-view video technology.

Presently, there are a variety of multi-view video coding schemes have been proposed, In which the use of hierarchical B frame prediction structure coding scheme can significantly improve the coding efficiency, but most of the current coding scheme mainly for compression efficiency, less considerations for coding complexity and random access performance. This paper studies a number of typical multi-view video coding schemes, then selects the location of base-view, and changes inter-view prediction structure, while decreaces the GOPlength in a Hierarchical B frame to improve the random access performance while maintaining high coding efficiency.

2.Typical MVC Coding Scheme

2.1 Hierarchial B Frames Coding Scheme

Hierarchical B frames can be better to remove the temporal redundancy, which can improve the image quality, while maintaining low bit rate. Experiments show that hierarchical B frames coding structure can achieve more coding efficiency than traditional IPPP structure [2]. Coding structure shown in Fig. 1: I denotes intra code frame; P denotes single direction of inter prediction farame; B1 and B2 are double direction inter prediction frame, can use them as reference frames; B3 is also double direction inter prediction frame, but can't use as reference frame.

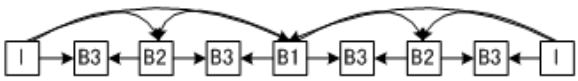

Fig. 1 Hierarchical B frames coding structure

2.2 MVC Coding Scheme

The 3DAV special group did unified subjective and objective tests for MVC schemes which all companies and research institutes proposed, and then got the test results [3]. The test results show that Fraunhofer-HHI proposed the coding prediction structure with both spatial and temporal references based on the hierarchical B frames that obtains higher coding efficiency [4]. Compared with the simulcast scheme, this coding scheme achieves up to1.4~1.6dB in coding efficiency [5], so this scheme was chosen as a reference prediction structure for MVC by the Joint Video Team (JVT). This scheme removes temporal redundancy with hierarchical B frames prediction coding structure in intra-view direction and removes intra-view redundancy with IBPBP prediction structure in inter-view direction.

The MVC encoding prediction scheme includes 8-channel views, and the length of each view's GOP is 8, the last picture is called anchor picture, which helps improve the random access performance and synchronization, the other pictures are called no-anchor picture. Where, View V0 only uses temporal references, but don't use inter-view references. So V0 is called base view, and the other views are called no-base views. In this scheme, non-base view is classified P-view (V2, V4, V6 and V7) and B-view (V1, V3, V5).Fig. 2 shows that Compared with simulcast scheme, this coding scheme greatly improved the coding efficiency and data transmission capacity, but it included prediction relation in inter-view direction, which reduced random access performance and increased decoding delay. The prediction structure is illustrated in Fig. 2.

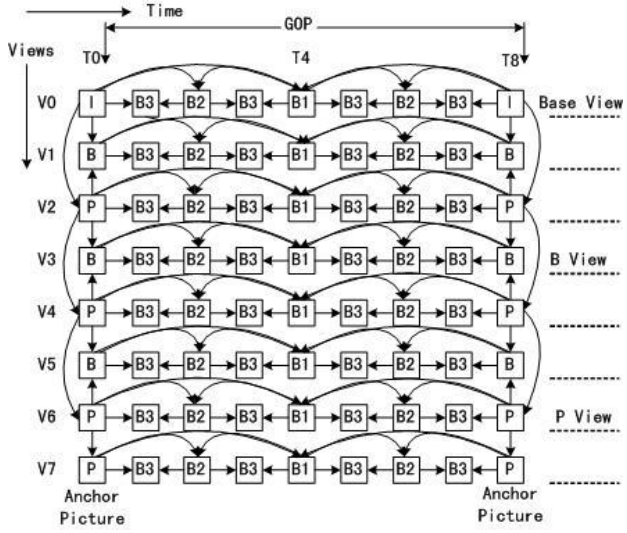

Fig. 2 MVC coding scheme

3. Proposed MVC Coding Scheme

3.1Coding Scheme based on Selection of Base-view

The base-view is reference of other views, so the selection of base-view can improve coding efficiency and random access performance. This paper calculates the location of base-view by global disparity. Park defines mean absolute global disparity (MAGD) as follows:

$$MAGD(v) = \frac{1}{N}\sum_{w=0}^{N-1} |\, g(v,w) \,| \qquad (1)$$

Where N is the number of views and g (v, w) is the global disparity between views v and w. MAGD (v) means the average of absolute global disparity between view number v and all the other views. TABLEI lists MAGD values of test sequences and the views location with minimum MAGD values are marked with grey boxes.

TABLE 1:MAGD VALUES FOR VARIOUS SEQUENCES

View	0	1	2	3	4	5	6	7
Ballroom	38.9	27.4	24.0	21.5	23.8	33.9	28.9	38.8
Exit	101.5	79.4	64.9	58.0	57.0	64.4	80.9	104.8
Race1	54.9	27.6	34.4	20.8	19.9	16.9	35.5	40.4

However, for N-view sequences require $(N-1)^2/2$ times of global disparity calculations, this process has significant complexity [6]. So Park proposes a simplified method to select the location of base-view, S_l as follows:

$$S_l = \lfloor (N-1)/2 \rfloor [6] \qquad (2)$$

The MAGD values show that, base-view mainly locates in the middle of views in the 8-channel multi-view video sequences. Compared with the MVC coding prediction scheme, this scheme significantly improves the random access capability.

In this proposed multi-view encoding scheme, V4-view be chosen as base-view, and others are no base-view, while using the hierarchical B pictures prediction structure in the temporal direction, e.g.Fig. 3.

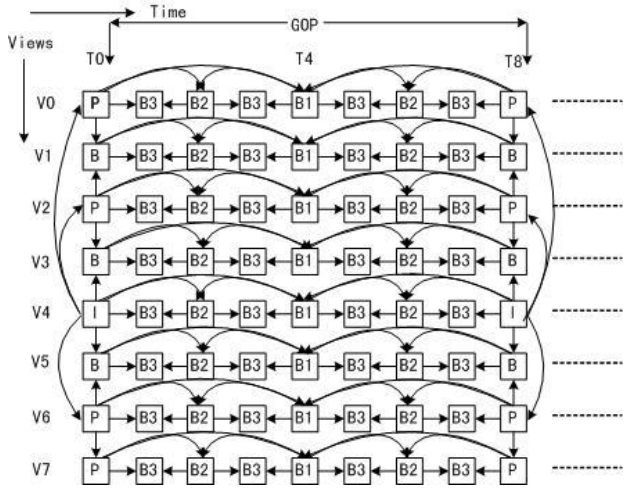

Fig. 3 Proposed coding prediction scheme MVC-1 (GOP=8)

3.2Coding Scheme based on Decreasing the GOPlength

According to the relation difference of the hierarchical B frames prediction structure in the temporal direction, images can be divided into different temporal layers，the temporal layer can be marked by temporal layer identification (TLID) [7]. In one GOP of MVC reference prediction structure, when TLID is equal to i, the number of images Ni is:

$$N_i = \begin{cases} 1 & i = 0 \\ 2^{i-1} & i = 1 \sim TL_{\max} - 1 \\ GOPlength - (1 + \sum_{j=1}^{TL_{\max}-1} 2^{j-1}) & i = TL_{\max} \end{cases} \quad (3)$$

Where, GOPlength is the length of a GOP, and is the maximum value of TLID, TLmax is given by:

$$TL_{\max} = \lceil \log_2 GOPlength \rceil \quad (4)$$

Where, $\lceil x \rceil$ denotes the minimum integer that is more than or equal to x. If the GOPlength is equal to 8, TLmax is equal to 3. According to the formula (3) and (4), the number of images is 1, 1, 2 and 4, when TLID is equal to 0, 1, 2 and 3 respectively. If getting frame of TLID is equal to 0, 1, 2 and 3, the frame number that to be decoded is 0, 2, 3 and 4 respectively, as shown in Fig. 5. If GOPlength is equal to 4, TLmaxis equal to 2. According to the formula (3) and (4), the number of images is 1, 1 and 2, when TLID is equal to 0, 1 and 2 respectively. If getting frame of TLID is equal to 0, 1 and 2, the frame number that to be decoded is 0, 2 and 3 respectively, as shown in Fig. 4 and Fig. 5.

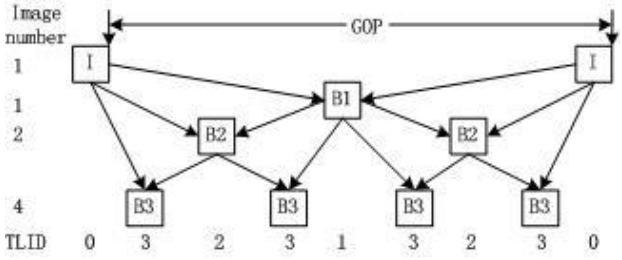

Fig. 4 Image number of the different temporal layer (GOP=8)

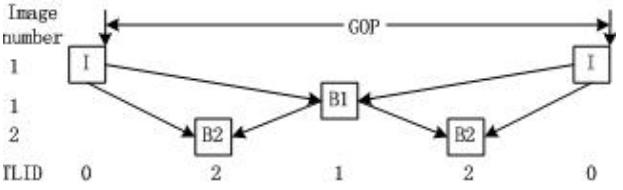

Fig. 5Image number of the different temporal layer (GOP=4)

This figure can be seen the length of GOP has direct influence on maximum number of frames which need to be decoded to access a frame, the random access performance of hierarchical B frame prediction structure decreases with the increase of GOP length, therefor, the improved random access performance can be got by decreasing the GOP length of hierarchical B frame. Based on the above mentiond points, this paper puts forward another coding prediction scheme that V4-view be chosen as base-view and GOPlength is equal to 4, which as shown in Fig. 6.

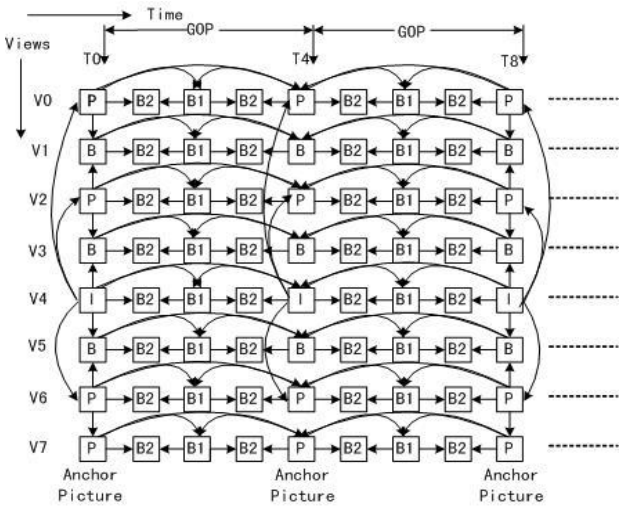

Fig. 6 Proposed coding prediction scheme MVC-2 (GOP=4)

3.3Random Access Performance

Random access performance is cost for accessing any frame in one video sequence, which is important indicator of

evaluating prediction structure. In this paper random access performance measured by the number of frames which need to be decoded to access a frame in one GOP.

When access a frame, the number of frames which need to be decoded are marked as n_{ij} .Where, i denotes the serial number of the frame accessed, j denotes the relative serial number of the frame in the GOP.

For evaluating synthetically random access performance of the prediction structure, two parameters are defined [8]: average number of frames marke as N_{avg} ; maximum number of frames to access a frame are marked as $N_{\max}$. Where N_{avg} denotes the average random access performance, and $N_{\max}$ denotes the worst random access performance, the smaller the value of $N_{\max}$ and N_{avg} , the better random access performance of prediction structure. The formuals are given by:

$$N_{avg} = \frac{1}{ViewNum \times GOPlength} \sum_{i=0}^{ViewNum} \sum_{j=0}^{GOPlength} n_{ij} \quad (5)$$

$$N_{\max} = \max(n_{ij}) \quad i = 0 \sim \text{ViewNum} - 1, \text{j} = 0 \sim \text{GOPlength} - 1 \quad (6)$$

Where, ViewNum denotes the number of views, GOPlength denotes the length of one GOP.

4.Experiment Results

To test the coding efficiency of MVC scheme and proposed scheme, some experiments has been done, which bases on H.264/AVC MVC codec JMVC 7.0 with the sequences “Flamenco1” and “Race1”, which consists of 8 views captured by KDDI Corp. and Nagoya University. Because scene change slowly in sequence “Flamenco1”, and scene change quickly in sequence “Race1”, so we select the two typical sequences for test. The spatial resolution is 320×240 for Flamenco1, and 640×480 for Race1. The frame rate is 30 fps for both Flamenco1 and Race1. BasisQP of encoder JMVC is 32, 27 and 22 respectively. The GOPSize is 4 and 8 respectively. The SearchMode is FastSearch. The rate-distortion performance about different coding schemes in two MVC sequences as shown in Fig. 7. The MVC-1 (GOP=8) scheme changes inter-view prediction structure, while the MVC-2 (GOP=4) scheme decreases the GOPlength and changes inter-view prediction structure, experiment rusults show that The PSNR of MVC-1 (GOP=8) scheme was slightly higher than MVC (GOP=8) scheme, which less than 0.05 dB. The PSNR of MVC-2 (GOP=4) scheme was slightly lower than MVC (GOP=8) Scheme, which less than 0.15 dB, but The PSNR of MVC-2 (GOP=4) scheme was slightly higher than MVC (GOP=4) Scheme, which less than 0.03 dB.

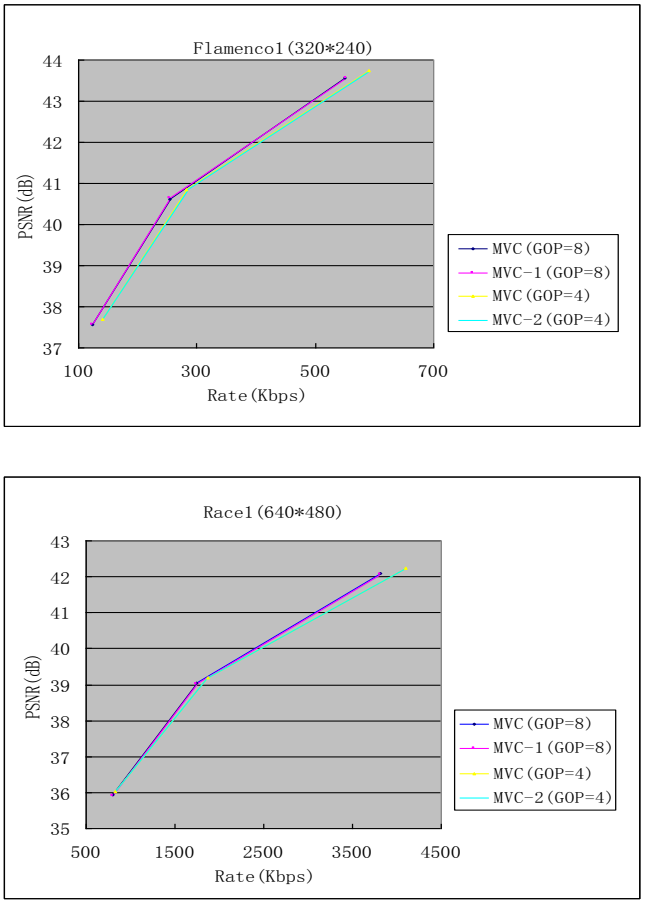

Fig. 7 Comparing the proposed scheme with MVC coding prediction structure

Random accsee performance as shown in TABLE II, TABLE III, TABLE IV, and TABLE V.

TABLE 2:NUMBER OF REFERENCE FRAMES NEED TO BE DECODED IN MVC (GOP=8) SCHEME

V/T	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
V_0	4	3	4	2	4	3	4	0
V_1	8	7	8	6	8	7	8	2
V_2	6	5	6	4	6	5	6	1
V_3	10	9	10	8	10	9	10	3
V_4	8	7	8	6	8	7	8	2
V_5	12	11	12	10	12	11	12	4
V_6	10	9	10	8	10	9	10	3
V_7	12	11	12	10	12	11	12	4

N_{avg} =7.45 $N_{\max}$ =12

TABLE 3:NUMBER OF REFERENCE FRAMES NEED TO BE DECODED IN MVC-1 (GOP=8) SCHEME

V/T	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
V_0	6	5	6	4	6	5	6	1
V_1	8	7	8	6	8	7	8	2
V_2	6	5	6	4	6	5	6	1
V_3	8	7	8	6	8	7	8	2
V_4	4	3	4	2	4	3	4	0
V_5	8	7	8	6	8	7	8	2
V_6	6	5	6	4	6	5	6	1
V_7	8	7	8	6	8	7	8	2

N_{avg} = 5.57 $N_{\max}$ =8

TABLE 4:NUMBER OF REFERENCE FRAMES NEED TO BE DECODED IN MVC (GOP=4) SCHEME

V/T	T_1	T_2	T_3	T_4
V_0	3	2	3	0
V_1	7	6	7	2
V_2	5	4	5	1
V_3	9	8	9	3
V_4	7	6	7	2
V_5	11	10	11	4
V_6	9	8	9	3
V_7	11	10	11	4

N_{avg} = 6.15 $N_{\max}$ =11

TABLE 5:NUMBER OF REFERENCE FRAMES NEED TO BE DECODED IN MVC-2 (GOP=4) SCHEME

V/T	T_1	T_2	T_3	T_4
V_0	5	4	5	1
V_1	7	6	7	2
V_2	5	4	5	1
V_3	7	6	7	2
V_4	3	2	3	0
V_5	7	6	7	2
V_6	5	4	5	1
V_7	7	6	7	2

N_{avg} = 4.40 $N_{\max}$ =7

It can be seen that random access performance of MVC-1 (GOP=8) scheme were higher than MVC (GOP=8)one, average number of frame(N_{avg}) to be referred decreases 1.88, more than 25%, and maximum number of frame($N_{\max}$)to be referred lower 4; Compare the coding scheme of MVC-2 (GOP=4) with the MVC(GOP=8) , the average number of frame(N_{avg}) to be referred decreases 1.17, more than 21%, and the maximum number of frame($N_{\max}$)to be referred decreases 1; Compare the encoding scheme of MVC-2 (GOP=4) with the MVC (GOP=4) , the average number of frame(N_{avg}) to be referred decreases 1.75,more than 28%, and the maximum number of frame($N_{\max}$) to be referred decreases 4.

5. Conclusion

On the basis of the analysis of the several typical Multi-view Video Coding schemes, this paper puts forward a proposed coding scheme, which selects the location of base-view and the prediction structure of inter-view by calculating global disparity; while improving the random access performance by decrease the length of GOP. Experiment results show that the proposed scheme with more performance of view random access while maintaining high coding efficiency.

References

[1] Ho Yo-Sung, Oh Kwan-Jung, "Overview of Multi-view Video Coding" ,Systems, Signals and Image Processing, 2007 14th Intl Conf On Held With Speech and Image Processing, Multimedia Comm&Svcs, 2007 6th Eurasip Conference Focused On, Maribor, Slovenia. Maribor, Slovenia: IWSSIP, 2007, 5-12.

[2] Lini Ma and Feng Pan，Efficient Compression of Multi-View Video Using Hierarchical B Pictures, Proceedings of MUE 2008, 118-121, Busan, Korea.

[3] Jens-Rainer Ohm,"Submissions received in Cfp on Multiview Video Coding",75th MPEG meeting,M12969, 2006.

[4] MPEG Test and Video subgroup, "Subjective test results for the Cfp on Multi-view Video Coding",75th MPEG meeting, W7779, 2006.

[5] PhiliPPMerkle,AljisehaSmolie,KarstenMuller,etal,"Efficient Predictio Structures for Multiview Video Coding".IEEE Transaetions on Circuits and Systems for Video Teehnology, 2007, Vol.17, No.ll，pp.1461-1471.

[6] Park P K, Oh K J, and Ho Y S, "Efficient view-temporal prediction structures for multi-view video coding", Electronics Letters, Vol.44, No.2,2008, pp.102-103.

[7] P.Merkle, A.Smolic, K.Muller,and T.Wiegand,"Efficient Prediction Structures for Multiview Video Coding", IEEE Trans. On Circuits and Systems for Video Technology,Vol.17, No.11,2007, pp.1461-1473.

[8] Heiko Schwarz, Detlev Marpe, Thomas Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard" IEEE Trans, On Circuits and Systmes for Video Technology, Vol.17, No.9,2007,pp.1103-1120.