International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 5 (2015), pp. 1-8
http://dx.doi.org/10.14257/ijsip.2015.8.5.01

Research and Optimization of Prediction Structure of Multi-view Coding

[1]Meileng Yuan and [2]Shuping Yang

[1,2]Shenzhen Polytechnic, Shenzhen 518055
mlyuan@szpt.edu.cn, ysping@szpt.edu.cn

Abstract

This paper analysis the prediction structure based on layered B frame used by multi view video coding, introduces the common predictive structure and their evaluation index, proposes a kind of improved prediction structure, and makes comparative analysis on the performance of encoding time, PSNR and code rate by experiment which uses the typical test sequences Ball-room and Exit provided by MERL on the multi view video coding test platform. The experimental results show that: the improved prediction structure showed better coding efficiency; coding complexity reduction is up to 15%, the better to improve the real-time performance of code.

Keywords: *MVC, reference frames, prediction structure, complexity of coding*

1. Introduction

Now, the main video coding often use predictive coding, so the design of prediction structure is very important in video coding. Prediction structural design will not only affect the quality of video, the compression performance, will also affect the random read performance, fast decoding performance, scalability and parallelism, so it has been an important topic in MVC research, but also one of the focuses of this paper.

Reasonable prediction structure can effectively reduce the information redundancy in time domain, space information domain and inter viewpoint, to achieve higher compression efficiency. The most typical prediction structures are GOGOP prediction structure, sequence prediction structure, improved prediction structure, minimum cost tree structure and hierarchical B prediction structure [1-3]. The GOGOP [1] prediction structure presented by NTT laboratory in Japan, by setting multiple reference frame strategy in the Inter GOP, improve the random access performance, but because of the multiple I frame coding, the coding efficiency is not high, and the computational complexity is relatively high. Sequence structure prediction in paper [2] by using multiple reference frames, effectively reduces the time redundant information of intral view and reduce redundant space information between adjacent inter viewpoint, improves the coding efficiency to some extent but the random access performance, and is easy to cause error accumulation and error transfer. SIMULCAST, KS_IPP, KS_IBP, HBP. Merkle [3] from German HHI Institute studied the correlation of multi view video in the time direction and view direction, and design 4 kinds of MVC structure: SIMULCAST, KS_IPP, KS_IBP, and HBP. The results of the study indicate that, the direction selected as the best macro-block reference frame most times are as follows: time direction, view direction, temporal and inter view mixed direction. Comparing the prediction using inter view prediction to that without using of inter view prediction, the coding rate distortion performance is improved obviously. Considering the hybrid directional prediction of rate distortion performance is not improved much, and the complexity of multiple reference frames will increase the inter frame prediction coding, the mainstream multi-view coding prediction use only the time direction and view direction. Following Merkle, *et al.*, proposed two prediction by using inter view correlation, respectively is AS_IPP and

ISSN: 2005-4254 IJSIP

AS_IBP structure, which compared with KS_IPP prediction structure and KS_IBP prediction structure, the difference lies in whether there has the reference relationship between different point of view in the non anchor time, the latter has not. The experiments show that, the HBP coding predictive structure proposed by HHI based on the combination of time-domain prediction and inter view hierarchical B frame get higher coding efficiency, so it is selected as the reference prediction structure of MVC [4] by JVT.

For example, literature [5] presented a prediction structure based on minimum spanning tree, compared with the reference prediction structure, PSNR increased about 0.1 ~ 0.2dB. A new prediction structure designed by the literature [6] reduces the coding complexity by using two-way inter view prediction for non key frame of the enhancement layer, and by selecting the greater possible prediction direction using code information of basic viewpoint to identify the Macro-block moving slowly, but two-way inter view prediction of the non key frame will increase the encoding complexity. Some research group specially study the prediction relationship of coding time and inter view of MVC, and use the dependence of multi view video effectively.

2. Multi-view Video Coding Prediction Structure

2.1 Layered Frame B Prediction Structure

The multi-view video coding prediction structure adopts layered frame B structure. This structure was put forwarded by German HHI lab, and then it was accepted by JVT by its good coding performance and taken as standard reference prediction structure of JMVC. The frame type of multi-view video coding is the same as traditional single video which includes three types frame of frame I, frame P and frame B. Frame I adopts inner frame prediction, frame P adopts prediction among single direction frames, frame B adopts prediction among bidirectional frames. Figure 1 is time layered diagram of frame B prediction structure when GOP=12, we can get that this structure include a key frame (frame I or frame P) and several frame B. The picture is divided to different temporal layer (TL) according to time interval length of present frame and time reference frame, different color frames B stand at different time level, frame B at lower time layer can taken coded layer at higher time layer as reference. Supposing key frame is at the highest time layer, TL=0, frame B within one group GOP=12 can be divided to 4 layers, then T6 is B1, time layer TL=1, horizontal reference frames are T0 and T12; when T3 and T9 are B2, time layer TL=2, horizontal reference frames are T0, T6, T12; the diagram of TL=3 and TL=4 can be analogized like this.

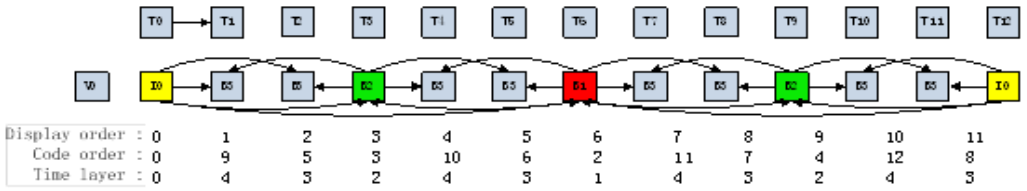
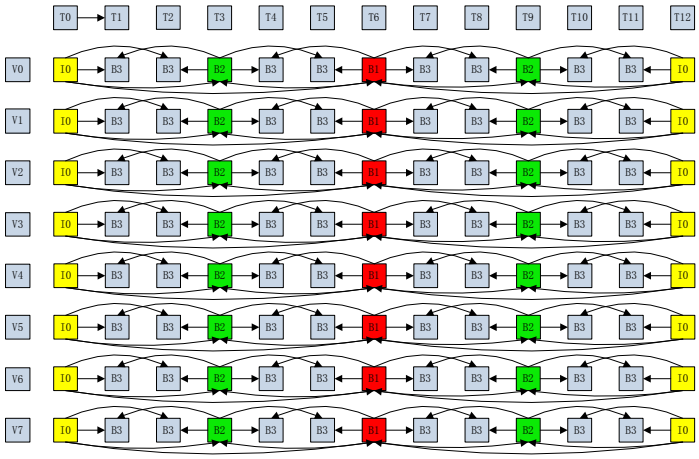


Figure 1. Frame B Prediction Structure Time Layer Diagram

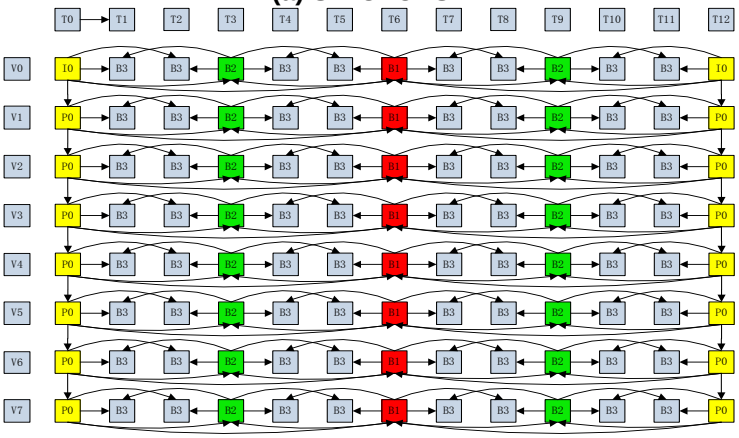
As layered coding makes picture coding and display to configure randomly, for key frame, the picture displays at front when coding comes front, for non-key frame, the picture is coded according to different time layer and displays to realistic order.

2.2 Typical Prediction Structure

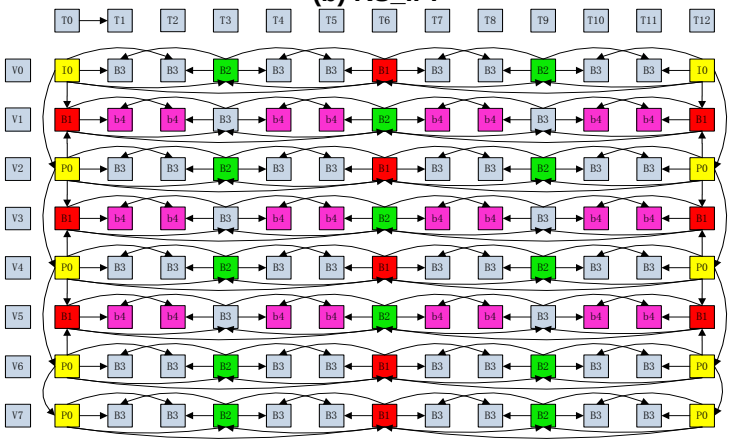
Prediction structure always plays an important role in multi-view video coding research, German HHI lab began its research at first, he designed 4 types MVC prediction structures which were SIMULCAST, KS_IPP, KS_IBP and Hierarchical IBP as shown in Figure 2 and compared each structure's compression efficiency. These 4 structures use layered frame B at horizontal and vertical are different according to different prediction structure. In picture 4.2 there are total 8 view points, No. 1, 3, 5 and 7 view points are numbered to V0, V2, V4 and V6 as even view points, No. 2, 4, 6 and 8 are numbered to V1, V3, V5 and V7 as odd view points. V0 is the basic view point; the coding way is the same as traditional single view point prediction.



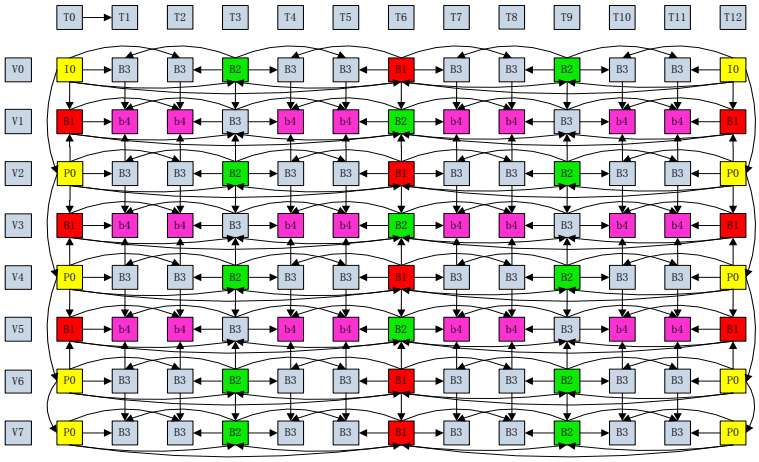
(a) SIMULCAST



(b) KS_IPP



(c) KS_IBP



(d) Hierarchical_IBP

Figure 2. Typical Prediction Structure

SIMULCAST prediction structure coding takes no reference at vertical direction, view points are totally independent among each other, each point codes separately, and it is the simplest prediction structure. Although its random access performance is the best, the redundancy among points can't be deleted, so its data volume is big and coding efficiency is low; Compared with SIMULCAST, KS_IBP and KS_IPP add key frame prediction, which KS_IPP adds IPP prediction structure and KS_IBP inserts frame B in added view points prediction (it is called IBP..BP). That means each point key frame of KS_IPP at vertical direction will refer to the previous neighboring view point picture at the same time, even view point key frame of KS_IBP refer to previous even neighboring view point view at the same time, and odd view point (except V7) refer to neighboring two even point views at the same time, key frame of point V7 refer to V6 at the same time, these two coding efficiency have some improvement, but not quite obvious; The even view points coding of Hierarchical B prediction structure is the same with KS_IBP, key frame of odd view point (except V7) refer to neighboring even point views at the same time, non key frame not only refer to layered frame B structure at horizontal but also refer to neighboring even view point at vertical, key frame of point V7 refer to V6 at the same time. This largely increases coding efficiency but also make coding more complicated.

The research indicated that PSNR of Hierarchical B is about 3dB higher than Simulcast with good coding performance. Compression efficiency and coding complexity of KS_IBP and KS_IPP are among SIMULCAST and Hierarchical B.

2.3 Prediction Structure Evaluation Indicators

The prediction structure performance are evaluated by coding efficiency, decoding picture buffer zone, random access performance and coding-decoding complexity, among them coding efficiency is shown in PSNR under certain bit rate, coding complexity is shown in coding time, below are the introduction of meaning and calculating way of decoding picture buffer zone volume and random visit performance.

(1) Decoding picture buffer zone volume

Decode Picture Buffer (DPB for short) is used to store reconstructed picture frame after decoding. As multi reference frames are introduced and multi-view video increase more frames, more spaces are needed for DPB to store reconstructed picture. In Hierarchical_IBP structure, view point numbers are 8, GOP=12, DPB volume is 32.

(2) Random visit performance

Random Access (RA for short) is the key indicator to evaluate MVC prediction structure and major factor to design prediction structure; it is cost of visiting any video frame, and it usually use needed decoding reference frame number to evaluate. Good random access performance make user to view from any point freely.

3. JMVC Prediction Structure Analysis and Optimization

3.1 JMVC Reference Prediction Structure Analysis

JMVC takes Hierarchical B as reference prediction structure which is shown in Figure 2(d). Compared with MVC coding in Simulcast structure, Hierarchical B structure coding efficiency is largely increased because of I/P frame non-basic view point coding and non-key frame B neighboring view point prediction. Although it increases coding efficiency, it makes coding more complicated and random access more slowly, thus it is needed to improve JMVC prediction structure. This article mainly researches how to decrease coding complexity as much as possible at keeping view quality not decrease largely.

From Figure 1 Frame B time layered diagram, we can get that the interval between non-key frame and reference frame at different time layer is different, TL is bigger and time interval is shorter. Also we can get from Figure 1 that picture amounts at different time layer are different. When GOP=12, TL max=4, picture amounts of TL=0,1,2,3,4 are 1,1,2,4,4. Picture amounts of TL=3,4 are covered 33% of total separately, so two add can cover 66% of total amount. For frame B layered structure, TL is bigger, time interval is shorter, and then time relativity is much stronger. For view point prediction, as camera distance is fixed at shooting, the relativity between present coding frame and reference will not change with TL. Table 1 is the best reference picture distribution at different time layer, the statistic data comes from macro block amount ratio of present time prediction coding frame or among view prediction coding.

Table 1 Best reference picture distribution at different time layer

Test data		TL=1	TL=2	TL=3	TL=4
Ball-room	Time[%]	77.1	84.5	89.8	90.9
	View[%]	22.9	15.5	10.2	9.1
xit	Time[%]	81.3	91.8	95.8	96.6
	View[%]	18.7	8.2	4.2	3.4

From Table 1 we can find that above 75% best reference picture of Ballroom and Exit comes from time reference frame, as TL increases, referred time ratio is bigger and bigger, and referred view point ratio is smaller accordingly. When TL reaches max and referred time ratio reaches over 90%, except for view change, when the macro block number of point prediction coding reaches 0, only time prediction can be done here.

3.2 Reference Prediction Structure Optimization

From above analysis we can get the relativity of frame at different time layer and reference frame are different within a GOP. When TL is bigger, time interval between reference frame and present frame is smaller, the relativity is stronger, and time prediction is more accurate; Picture amount at different time layer are different too, TL is bigger and Picture amount is bigger. If we can use this difference efficiently to improve prediction structure, then the coder performance can be improved largely. So the article proposes prediction structure in Figure 3:

For frame at TL=3 and TL=4, the time relativity is very strong, and picture amount ratio covers over 66%, if we don't tale view point prediction, but just time prediction

structure, it can decrease calculation complexity and improve random access performance.

From Hierarchical B prediction structure diagram analysis we can get that even view point at vertical direction only add key frame reference, non key frame only has time prediction at horizontal direction, relativity among view points is not fully used. In order to reduce decreased coding efficiency from simplify prediction structure, view point prediction is increased to all points P non key frame at TL=1 and TL=2, that also means refer to last I/P point picture at the same time. As picture at TL=1 and TL=2 will be taken as point B's reference, if increase picture prediction accuracy of point P at these two time layers, it will reduce cumulative error transferred to point B and largely improve coding efficiency. Also picture amounts at TL=1 and TL=2 cover only 25%, which is less than 66% of picture amount at TL=3 and TL=4, It will not increase much coding complexity.

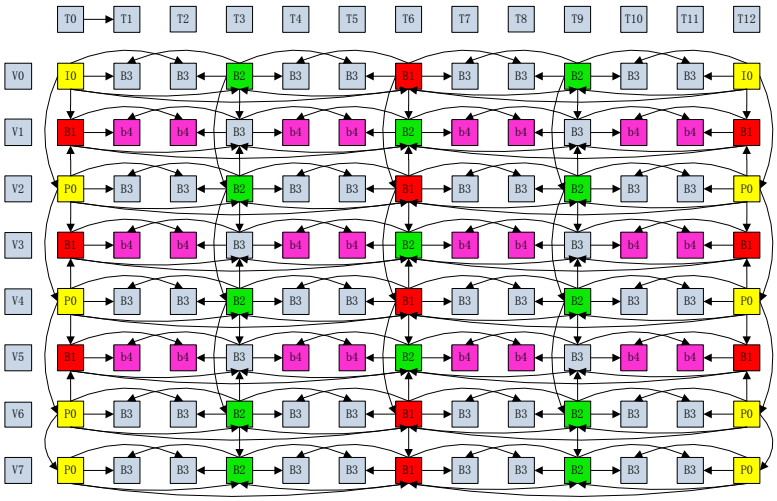


Figure 3. Improved Prediction Structure Diagram

4 The Simulation Experiment and Result Analysis

4.1 The Experimental Configuration

The experiment adopts typical test sequences Ballroom and Exit provided by MERL, the test condition configuration is as shown in Table 2. The version number of the public test platform of multi view video coding is JMVC8.5, the experimental platform is the Win8 64bit operating system, host configuration: memory 4GB (3.88GB available), Intel (R) Core (TM) i3-3227U CPU @1.90GHz.

Table 2. Test Condition Configuration

parameter name	configuration values
uantization parameter	24、28、32、36
The length of GOP	12
Coding frames	37
Unidirectional search range	64
The maximum reference frames	2
Bidirectional iteration number and scope	4 and 8

4.2 Experimental Results and Analysis

Table 3 and Table 4 gives the coding efficiency and coding complexity comparison of improved prediction structure (A) and JMVC prediction structure. The △PSNR shows

the changes in average peak signal-to-noise ratio, △B shows a change in the average bit rate, △T shows the percentage change of encoding time, "+" means increase, "-" indicates reduced.

Table 3. Comparison of the Coding Efficiency

Video sequence	Quantization parameter	JMVC		The improved prediction structure A		The difference between A and JMVC	
		Output bit rate /(kbit·s^{-1})	Peak signal of noise ratio /(dB)	Output bit rate /(kbit·s^{-1})	Peak signal of noise ratio /(dB)	ΔB/(%)	ΔPSNR/(dB)
Ball-Room	24	1142.8797	39.4969	1121.1230	39.4996	-1.90	+0.0027
	28	652.9068	37.7726	638.9892	37.7773	-2.13	+0.0047
	32	385.3594	35.9187	376.5878	35.9251	-2.28	+0.0063
	36	237.5770	34.0430	232.3845	34.0476	-2.19	+0.0046
Exit	24	636.5183	40.6250	632.7074	40.6245	-0.60	-0.0006
	28	332.3824	39.3904	329.7372	39.3886	-0.80	-0.0018
	32	194.6426	37.9945	193.4182	37.9910	-0.63	-0.0035
	36	122.6696	36.3840	121.9054	36.3797	-0.62	-0.0043

Table 4. The Encoding Complexity Comparison

Video sequence	Quantization parameter	JMVC Coding time（s）	The improved prediction structure A Coding time（s）	The difference of encoding time of the improved prediction structure A and JMVC(ΔT（%）
Ball-Room	24	38733	34396	-11.20
	28	37131	32431	-12.66
	32	31906	30237	-5.23
	36	33220	28036	-15.61
Exit	24	35257	31234	-11.41
	28	33344	29760	-10.75
	32	31615	27128	-14.19
	36	28539	25174	-11.79

The Table 3 and Table 4 show: for the Ball-Room sequence, regardless of the rate, the peak signal to noise ratio, or the coding complexity, the improved structure is obviously superior to the reference prediction structure, the bit rate is reduced by about 2%, the peak signal to noise ratio increased by 0.0027~0.0063, the encoding time is reduced by 5.23%~ 15.61%. But for the Exit sequence, compared to the reference prediction structure improved, the improved prediction structure gains better code rate and code complexity, yet the peak signal to noise ratio decreased a bit. The rate of peak signal to noise ratio reduces not more than 0.005dB, code rate reduces 0.6%~0.8%, the encoding time is reduced by 10.75%~14.19%.

5 Conclusions

This paper analysis the prediction structure based on layered B frame used by multi view video coding, introduces the common predictive structure, such as SIMULCAST,

KS_IPP, KS_IBP and Hierarchical_IBP, and gives the evaluation index for prediction structure commonly used, and proposes a kind of improved prediction structure. The paper makes comparative analysis on the performance of encoding time, PSNR and code rate by experiment which uses the typical test sequences Ball-room and Exit provided by MERL on the multi view video coding test platform. The experimental results show that: the improved prediction structure showed better coding efficiency, coding complexity reduction is up to 15%, the better to improve the real-time performance of code.

References

[1] "ISO/IEC JTC1/SC29/WG11", Subjective test results for the Cfp on multi-view video coding, N7779, Bangkok, **(2006)** January.
[2] "Survey of Algorithms Used for MVC.ISO/IEC JTC1/SC29/WG11, N6909", Hong Kong, China, IEEE, **(2005)**.
[3] P. Merkle, A. Smolic, K. Muller, *et al.*, "Efficient Prediction structures for multi-view video coding", IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, **(2007)**, pp. 1461-1473.
[4] "ISO/IEC JTC1/SC29/WG11", Requirements on multi-view video coding, v.4, N7282, Poland, **(2005)** July.
[5] D.-X. Li, W. Zheng, X.-H. Xie and M. Zhang, "Optimizing inter-view prediction structure for multi-view video coding with minimum spanning tree", Electronics Letters, vol. 43, no. 23, **(2007)**, pp. 1269-1271.
[6] J.-P. Lin and A. C.-W. Tang, "A Fast Direction Predictor of Inter Frame Prediction for Multi-view Video Coding", IEEE International Symposium on Circuits and Systems (ISCAS), **(2009)** May, pp. 2589-2592.
[7] K.-J. Oh and Y.-S. Ho, "Multi-view video coding based on the lattice-like pyramid GOP structure", In Proc. PCS 2006, Picture Coding SYMP, Beijing, China, **(2007)** April.
[8] X. Cheng, L. Sun and S. Yang, "A multi-view video coding scheme using shared key frames for high interactive application", In Proc. PCS 2006, Picture Coding SYMP. Beijing, China, **(2006)** April.
[9] Y. Yang, G. Jiang, M. Yu, F. Li and Y. Kim, "Hyper-space based multi-view video coding scheme for free viewpoint television", In Proc. PCS 2006, Picture Coding Symp., Beijing, China, **(2006)** April.
[10] H. Junjun, "The assessment of multi view video coding prediction structure and the optimization of stereo video encoder", master's degree of Information and communication engineering college, Zhejiang University, China, **(2012)**, pp. 19-20.
[11] Z. Yan, "Research of adaptive prediction structure model of multi view video coding and optimization model optimization technology", master's degree of Information and communication engineering college, Huaqiao University, Xiamen, **(2012)**, pp. 16-25.
[12] Y. Hui, "Research of fast algorithm based on multi view video coding", master's degree of computer application technology, Shenzhen University, **(2013)**, pp. 7-9.

