INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2005/M12542
October 2005, Nice

Title: Multi-view Video Coding based on Lattice-like Pyramid GOP Structure
Source: GIST and KETI
Authors: Yo-Sung Ho, Kwan-Jung Oh, and Han Oh
(Gwangju Institute of Science and Technology)
Byeong-ho Choi and Je-Woo Kim
(Korea Electronics Technology Institute)
Status: Proposal

1 Introduction

This document describes the multi-view video coding (MVC) based on the lattice-like pyramid GOP (group of pictures) structure. After we explain the lattice-like GOP structure and the pyramid GOP structure in Section 2, we present our experimental results and show the effectiveness of the proposed method in Section 3.

2 Lattice-like Pyramid GOP Structure

In order to encode the multi-view video efficiently, we propose to use a lattice-like pyramid GOP structure. The lattice-like GOP structure is an effective arrangement of I frames that have a big effect on the overall bit-rate and coding efficiency. The pyramid GOP structure is a hierarchical coding framework to improve coding efficiency within the GOP. We also propose a variable search range method for successful spatial referencing.

2.1 Lattice-like GOP Structure

In video coding, coding efficiency for a sequence highly depends on coding of I frames. I frames are usually encoded with high quality and referred by P and B pictures that influence the overall video quality. Moreover, at the decoder, I frames are used as the basic access units. Despite of this structural feature, however, I frames require a lot of bits and their frequent usage may lower the coding efficiency. The anchor coding scheme currently uses I frames in 0.5 sec. In this anchor coding scheme, we propose to use a lattice-like GOP structure that can locate I frames alternately in even and odd views and substitute RB frames for I frames. The RB frame is encoded like a B frame and referred by other frames. Figure 1 shows the proposed lattice-like GOP structure.

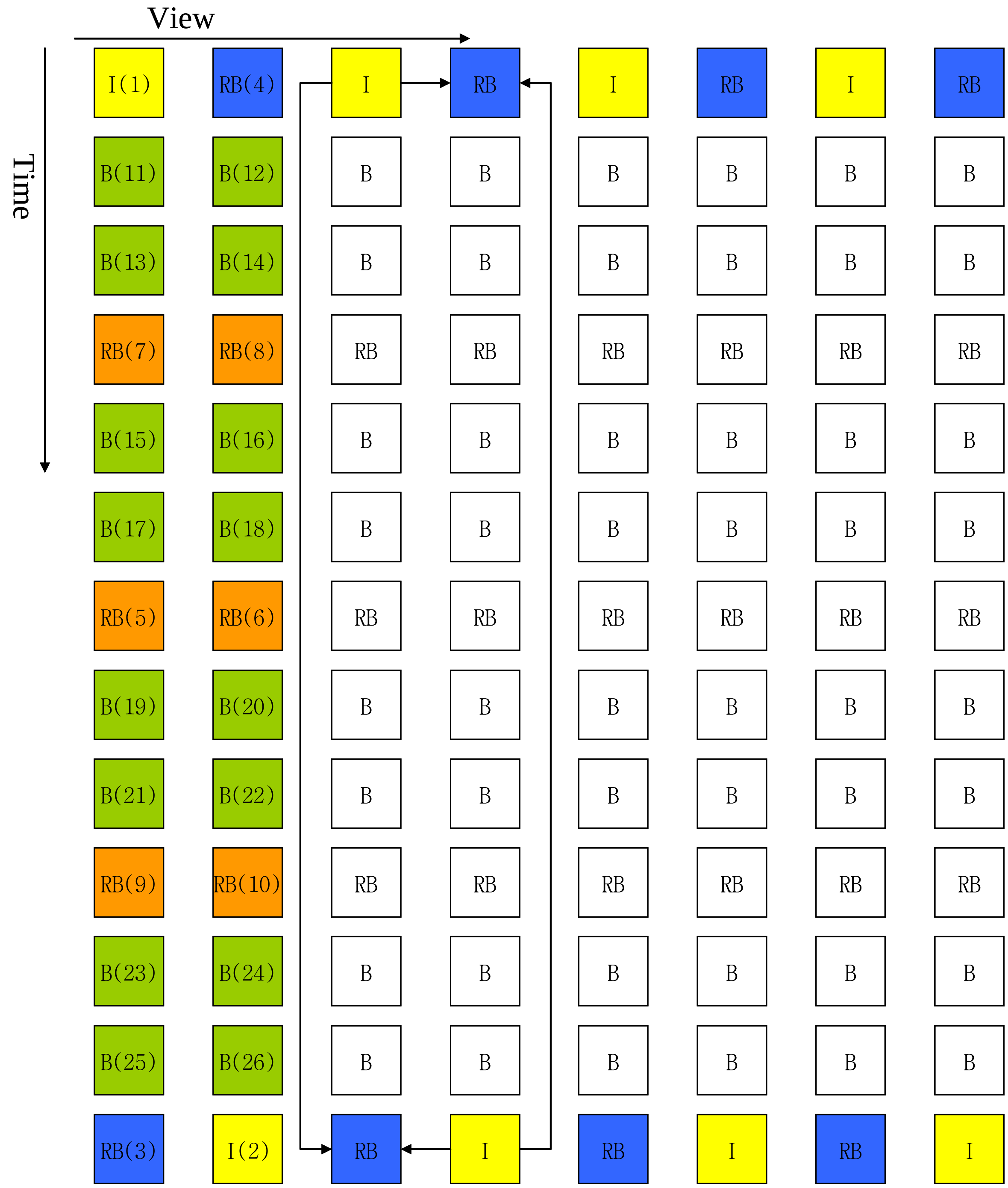

Fig. 1. Lattice-like GOP Structure

As shown in Fig. 1, certain frames of I frames in the anchor coding has been replaced by RB frames(colored blue). Each RB frame is encoded using the I frame of the adjacent view at the same time and the previous I frame at the same view. In addition, the RB frame is encoded with a smaller QP than that of the I frame to guarantee the same quality as the I frame. In Fig. 1, view 1 and view 2 are a pair, but this framework can be extended to various structures. The numbers in the parenthesis indicate the encoding order.

2.2 Pyramid GOP Structure

The coding efficiency of a GOP is influenced by the structure of the GOP. Anchor coding has the I-B-B-P structure and uses five reference frames. Even if the anchor uses multiple reference frames, there is not much difference in coding efficiency unless the reference frames have high quality, which may require significant coding time. In the proposed method, P frames in the existing anchor coding are replaced by RB frames and encoded through the hierarchical structure of the pyramid GOP to improve coding efficiency. Figure 2 shows the pyramid GOP structure [1].

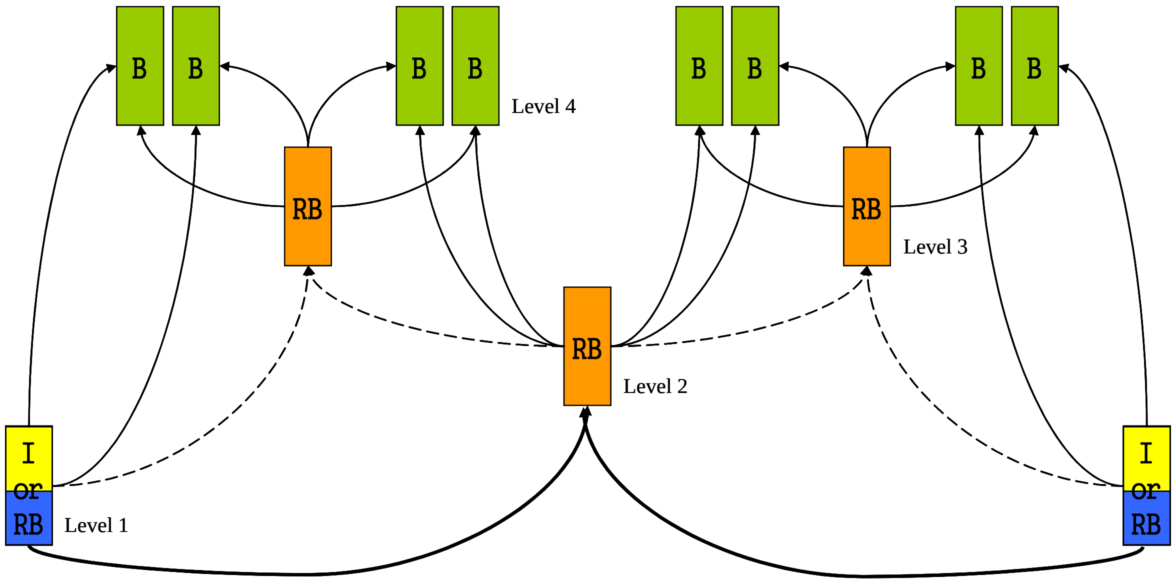

Fig. 2. Pyramid GOP Structure

In Fig 2, the RB frame in Level 2 is encoded using the I or RB frame in Level 1. Then, the RB frames in Level 3 are encoded using the RB frame in Level 2 and the I or RB frame in Level 1. Finally, B frames in Level 4 are encoded. The arrows in Fig. 2 indicate reference relationships. Besides, since the importance of each level is different, higher levels must be coded with high quality. This hierarchical structure has advantages in terms of scalability.

2.3 Variable Search Range

The main difference between MVC and conventional 2D video coding is that MVC can take a spatial reference between adjacent views. According to previous researches, the performance of spatial reference between views is inferior to that of temporal reference [2]. This is mainly due to variant characteristics of motions and illuminations depending on camera positions. The search range of the anchor coding scheme is ±32. This range is small to compensate the disparity difference between views, which is over 30 in 640×480 or 1024×768 video sequences. To address this problem, we uses ±64 as the search range temporarily for frames referring spatially.

3 Experimental Results and Analysis

In order to evaluate our proposed method, we have experimented with "Exit" and "Ballroom" sequences (640× 480, 8 views) provided by MERL and compared to the result of the anchor coding. The bit-rate has been set by adjusting QP manually. Table 1 shows our result for "Exit" and QP values we used. Figure 3 shows the rate and distortion curve for the "Exit" sequence [3].

Table 1. Simulation Results of "Exit"

View	Bit-rate(kbps)		PSNR (dB)		QP
	Anchor	Proposed	Anchor	Proposed	I(RB)-RB-RB-B
3	192.15	189.13	36.62	37.18	33(31)-32-33-33
4	192.22	191.55	36.26	36.95	33(31)-32-33-34
5	192.32	193.91	35.74	36.31	34(32)-33-34-34
6	192.33	195.11	35.66	36.26	34(32)-33-34-35
Average	192.26	192.43	36.07	36.68	
3	256.20	255.62	37.50	38.02	31(29)-30-31-31
4	256.34	253.11	37.24	37.85	31(29)-30-31-32
5	256.33	255.66	36.74	37.21	32(30)-31-32-32
6	256.42	256.43	36.71	37.18	32(30)-31-32-33
Average	256.32	255.21	37.05	37.57	
3	384.19	377.27	38.60	39.02	28(25)-27-29-30
4	384.21	379.39	38.46	38.95	28(25)-27-30-30
5	384.48	384.52	37.95	38.38	29(26)-27-30-31
6	384.62	387.70	37.98	38.32	29(26)-27-30-32
Average	384.55	382.22	37.98	38.67	

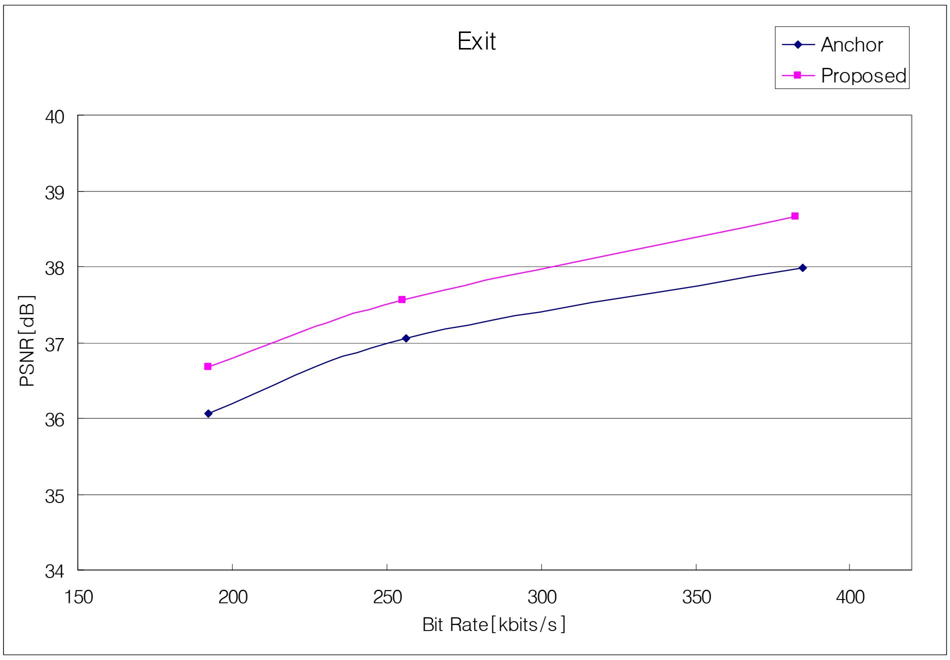

Fig. 3. RD Curve of Exit

Table 2 shows our result of "Ballroom" and QP values we used. Figure 4 shows the rate and distortion curve for the "Ballroom" sequence.

Table 2. Simulation Results of "Ballroom"

View	Bit-rate(kbps)		PSNR (dB)		QP
	Anchor	Proposed	Anchor	Proposed	I(RB)-RB-RB-B
3	256.47	256.75	31.43	32.17	38(36)-37-38-39
4	256.42	259.99	30.62	31.52	38(36)-37-39-39
5	257.04	256.50	30.49	31.28	38(37)-38-39-40
6	256.77	254.35	31.15	31.91	38(36)-38-39-40
Average	256.68	256.90	30.92	31.72	
3	384.63	386.53	33.38	34.01	35(33)-34-34-36
4	384.63	381.32	32.50	33.26	35(33)-34-36-36
5	384.83	384.61	32.41	32.85	36(34)-35-36-36
6	384.88	380.17	33.23	33.80	35(33)-34-36-37
Average	384.74	383.16	32.88	33.48	
3	512.45	513.85	34.68	35.10	33(31)-32-33-34
4	512.66	514.53	33.83	34.56	33(31)-32-33-34
5	512.84	511.04	33.71	34.05	34(32)-33-34-34
6	512.88	517.76	34.70	35.16	33(31)-32-34-34
Average	512.71	514.30	34.23	34.72	

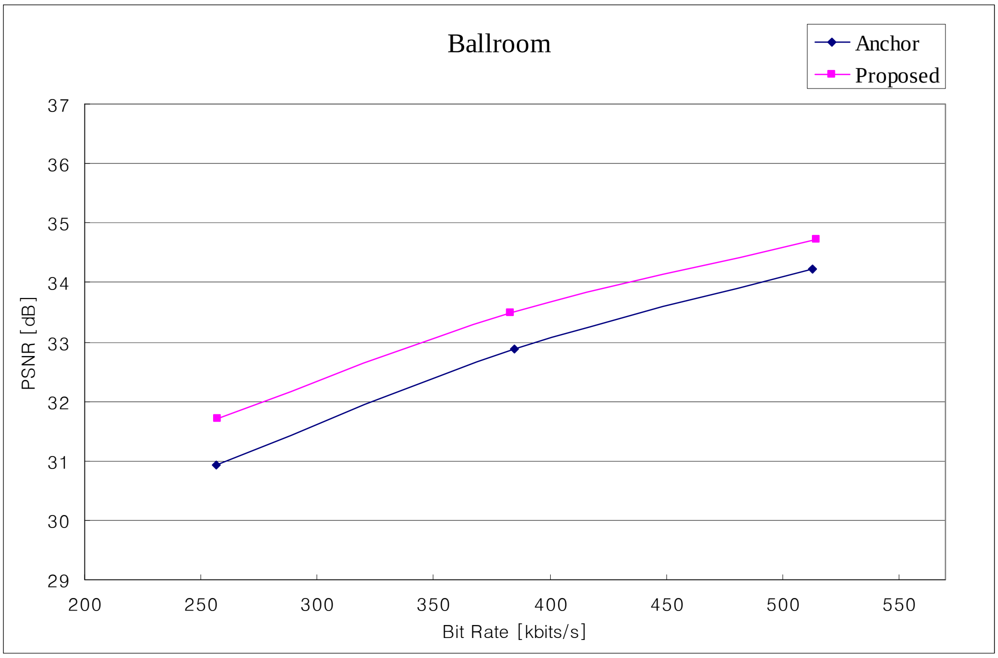

Fig. 4. RD Curve of Ballroom

As shown in Fig. 3 and Fig. 4, we have achieved quality improvement of approximately 0.5~0.8 dB when compared to the anchor coding scheme.

4 Conclusion

In this document, we have proposed a new structure for multi-view video coding using a lattice-like pyramid GOP structure and a variable search range technique. With some test sequences, we have verified quality improvement of approximately 0.5~0.8 dB when compared to the anchor coding scheme.

5 References

[1] ITU-T SG 16 Q.6 VCEG-N19, "H.264/MPEG-4 AVC Reference Software Enhancements," January 2005.

[2] ISO/IEC JTC1/SC29/WG11 M12301, "Statistical Evaluation of Spatiotemporal Prediction for MVC," July 2005.

[3] ISO/IEC JTC1/SC29/WG11 N7327, "Call for Proposals on Multi-view Video Coding," July 2005.