QUANTIZATION PARAMETER SELECTION METHOD FOR H.264-BASED MULTI-VIEW VIDEO CODING

Pil-Kyu Park, Seung-Hwan Kim, and Yo-Sung Ho

Department of Information and Communication
Gwangju Institute of Science and Technology (GIST)
1 Oryong-Dong, Buk-Gu, Gwangju, 500-712, Korea
E-mail: {ppk79, kshkim, hoyo}@gist.ac.kr

ABSTRACT

In this paper, we propose a QP (Quantization Parameter) selection method for H.264 based MVC (Multi-view Video Coding). The proposed method adaptively calculates QP for B frames in anchor (B_{ANC} frames). This method utilizes rate-distortion costs of the two frames which are referred by B_{ANC} frames. Using these costs, we estimate the correlation between the reference frames. The first reference frame is coded as I or P frame, and then the second reference frame is coded as P frame referring to the first reference frame. We assume that the rate-distortion cost of the second reference frame is smaller as the two reference frames are more highly correlated. QP for B_{ANC} frames is then calculated based on the correlation. Experimental results show that the proposed method improves coding efficiency of the reference prediction structure presented by Fraunhofer-HHI for MPEG core experiments. The proposed algorithm achieves 0.07~0.16 dB of average PSNR gain.

1. INTRODUCTION

Multi-view video is obtained by capturing a 3D (three dimensional) scene with two or more adjacent cameras. In case of traditional videos, we can watch them only from pre-determined viewpoint. However, multi-view video can offer arbitrary viewpoints of dynamic scenes. Thus viewers can choose viewpoints from which they want to watch.

Panoramic video, an example of multi-view video, is studied in many fields such as computer vision, image processing, and computer graphics. Panoramic video technology has various applications like image change detection, video compression, video indexing, and so on. Also, using computer vision technology, we can extract disparities and depth map from multi-view video. We can also implement 3D video by using two or more videos and a 3D display device. Hence, MVC (Multi-view Video Coding) is the essential technology for applications such as FTV (free viewpoint television), 3DTV, immersive teleconference, and surveillance [1].

Improvements of technologies and speed for transmitting data through internet enable diverse forms of multimedia contents. Recently, demands of interactive contents and realistic contents are growing rapidly. In this environment, multi-view video is a good alternative to satisfy those demands. However, as more cameras are used to obtain multi-view video, the amount of multi-view video data also increases. Therefore, for easier storing and transmission of data, more efficient coding technology than the previous single-view video coding technology is required.

Recently, ISO/IEC/JTC1/SC29/WG11/MPEG/adhoc group on 3-D audio and visual has started the standardization of MVC. In July 2005, CfP (Call for Proposals) and requirements of MVC algorithms were announced. CEs (Core Experiments) on view-temporal prediction structure, view interpolation prediction, and so on are currently in progress [2][3]. As reference MVC software, Fraunhofer-HHI donated a software package built based on the JSVM (Joint Scalable Video Model).

To achieve higher coding efficiency, MVC needs to exploit the spatial redundancy between sequences. Also, it’s necessary to maintain uniform qualities of sequences. The first frames of each GOP (Group Of Pictures) are called anchor frames. For additional coding efficiency, B frames are inserted to anchor and they are called B_{ANC} frames in this paper. By reference MVC software, B_{ANC} frames are coded with larger QP (Quantization Parameter) than the QP for I and P frames in anchor. Therefore, there exists non-uniformity between qualities of anchor frames and these non-uniform qualities of anchor frames affect coding efficiency of whole sequences. In this paper, we propose a QP selection method that selects QP for B_{ANC} frames not only to reduce the quality differences between sequences but also to improve the overall coding efficiency.

We used the test sequences for CE to evaluate the performance of the proposed method.

2. MULTI-VIEW VIDEO CODING (MVC)

2.1 Requirements for MVC

Algorithms proposed for MVC has some requirements. In the following, we use “shall” if a certain requirement is mandatory, and “should” if a certain requirement is desirable, but not necessarily required.

The requirements for MVC are classified into compression related requirements and system support related requirements.

Regarding compression related requirements, MVC shall provide high compression efficiency relative to independent coding of each view of the same content. View scalability shall be supported and SNR scalability, spatial scalability, and temporal scalability should be supported.

MVC should be efficient in terms of resource consumption and shall support low encoding and decoding delay modes. MVC should support robustness to error and enable flexible quality allocation over different views. Spatial resolutions from QCIF to HD and temporal random access shall be supported. MVC should also support view random access, spatial random access, efficient management of decoder resources, and parallel processing of different views or segments of the multi-view video.

For system support requirements, MVC shall support accurate temporal synchronization among the multiple views and should enable robust and efficient generation of virtual views or interpolated views. In addition, MVC should support transmission of camera parameters and efficient representation and coding methods for 3D display including IP (Integral Photography) and non-planar image display systems [4].

2.2 Prediction Structure of MVC

Since multi-view video consists of multiple views of the same scene, there is a high correlation between multiple views. Therefore, we can exploit spatial redundancy as well as temporal redundancy to achieve coding gain. Since this is usually achieved by performing spatial prediction across the different views, many inter-view-temporal prediction structures have been proposed. To define prediction structure for reference software, we can modify SequenceFormatString in its configuration file.

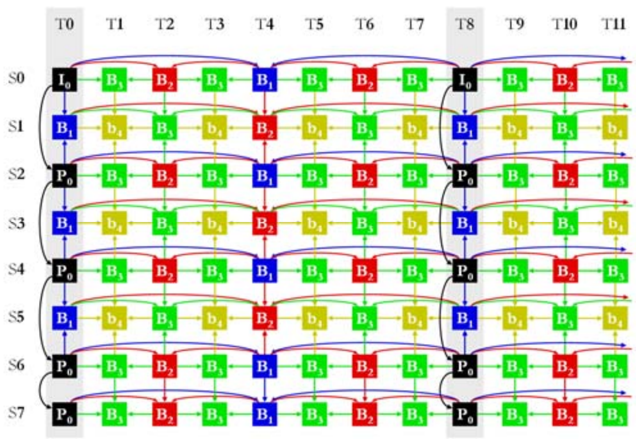


Figure 1. Prediction Structure by Fraunhofer-HHI

The inter-view-temporal prediction structure proposed by Fraunhofer-HHI is shown in Fig. 1. In this figure, Sn stands for nth-view camera, and Tn indicates the nth frame on temporal axis. This structure is used as a reference prediction structure for the standardization of MVC [5].

For temporal prediction, this structure uses hierarchical B picture structure. Anchor frames, the first frames of each GOP, are inserted every 0.5 or 1 seconds for temporal random access and prevention of error propagation. To achieve higher coding efficiency within anchor frames, S1, S3, and S5 frames at T0, T8, etc. are coded as B_{ANC} frames. The sequences whose anchor frames are coded as B_{ANC} frames, are coded using inter-view prediction as well as temporal prediction.

In Fig. 1, the length of GOP is 8, but GOP-lengths of 12 and 15 were used for experiments.

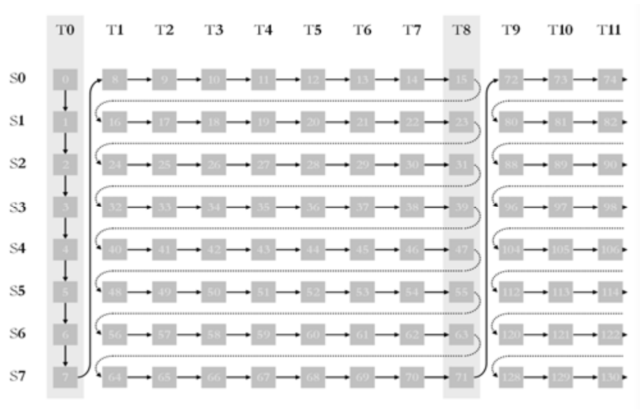


Figure 2. Frame Reordering of Input Sequences

In order to allow efficient memory management, the input video sequences are reordered as shown in Fig. 2. The first frames of all views are scanned first and the remaining frames of one GOP are zigzag scanned along the temporal axis. By reordering, multiple video sequences are combined into one sequence and coded by the reference software. Within a GOP, each layer is assigned a different QP and the QPs are defined in the configuration file of the reference software. For I and P frames in anchors, a fixed basis QP is assigned and for the remaining higher layers, the sum of basis QP and a delta QP value for each layer is used. Current settings of basis QPs for test sequences and delta QP values for CE are listed in Table 1.

Table 1. QP Settings for CE

Test Sequences	Basis QP		
Ballroom	34	31	29
Exit	31	29	26
Uli	36	30	28
Race1	28	26	24
Flamenco2	34	30	28
Breakdancers	31	26	22
Rena	33	28	23
Akko&Kayo	36	29	24

Delta QP Values	
DeltaLayer0Quant	0
DeltaLayer1Quant	3
DeltaLayer2Quant	4
DeltaLayer3Quant	5
DeltaLayer4Quant	6
DeltaLayer5Quant	7

DeltaLayer0Quant and DeltaLayer1Quant are applied for I and P frames, and B_1 frames, respectively. However, the prediction structure of Fraunhofer-HHI considers B_{ANC} frames to be in the same layer as B_1 frames, and assign the same QP. Therefore, B_{ANC} frames are coded with lower quality than I and P frames. When a frame is coded in hierarchical B picture structure, as the qualities of reference frames are lower, the coding efficiency of current frame is also lower. Hence, in case of the sequences whose anchor frames are coded as B_{ANC} frames (S1, S3, and S5 sequences in Fig. 1), most frames of these sequences refer to the B_{ANC} frames and thus the coding efficiency of these sequences are lower than the other sequences. Consequently, this causes non-uniformity between the qualities of sequences.

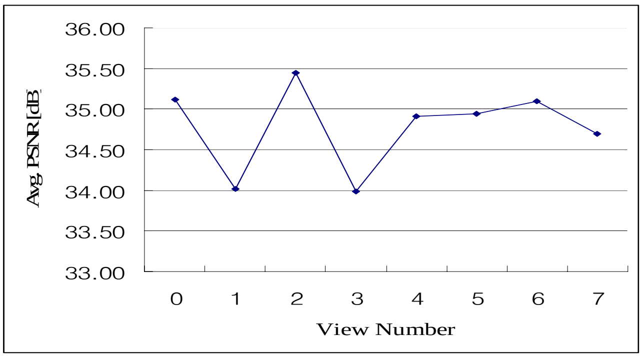


Figure 3. Coding Results of “Ballroom” sequences

Fig. 3 shows the average PSNR value of each view sequence of “Ballroom” sequences. Basis QP is 31. Anchor frames of the sequences with view number 1, 3, and 5 have been coded as B_{ANC} frames. We can see that the difference between the average PSNR values is greater than 1dB in the worst case.

3. PROPOSED QP SELECTION METHOD

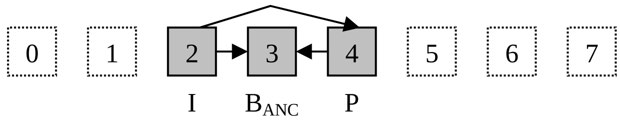
The proposed algorithm selects the QP for B_{ANC} frames independently from that of B_1 frames and thus reduces the difference of qualities between sequences and improves the overall coding efficiency.

For temporal random access, anchor frames including B_{ANC} frames are coded by using inter-view prediction only. As shown in Fig. 1, B_{ANC} frames refer to I and P frames, or two P frames of adjacent views.

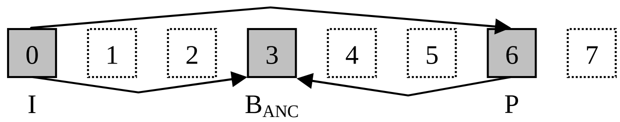
Table 2. Coding-efficiency Variations of B_{ANC} Frames

	QP for B_{ANC} = 35		QP for B_{ANC} = 36	
View Num.	PSNR [dB]	Bits [bits]	PSNR [dB]	Bits [bits]
2	38.5436	58,208	38.5436	58,208
3	37.2056	21,016	36.9255	17,968
4	37.7067	42,920	37.7067	42,920
0	38.4674	57,848	38.4674	57,848
3	37.0546	33,032	36.6693	28,552
6	37.4065	76,376	37.4065	76,376

Table 2 shows the coding results of B_{ANC} frames of “Ballroom” sequences with different QPs and reference frames. In this test, only the first frame of each sequence is coded. The first frame of view-number 3 sequence is coded as B_{ANC} frame. In the upper half of Table 2, B_{ANC} frame refers to the first frames of view-number 2 and 4 sequences as shown in Fig. 4(a). In the lower half of Table 2, B_{ANC} frame refers to the first frames of view-number 0 and 6 sequences as shown in Fig. 4(b). The QP for reference frames is 34. Because of the arrangement of cameras, the correlation between view-number 2 and 4 sequences is higher than that of view-number 0 and 6 sequences. When QP for B_{ANC} frame is fixed, the results in the upper-half table show higher coding efficiency than the lower-half table. Also, when QP for B_{ANC} frame is changed from 35 to 36, the results in the upper-half table shows less PSNR decrease than the lower-half table.



(a) High Correlation between Reference Frames



(b) Low Correlation between Reference Frames

Figure 4. Different Reference Frames for B_{ANC} Frame

Therefore, we can see that a B_{ANC} frame is coded more efficiently and is less sensitive to QP change, as the correlation between the reference frames is higher. In contrast, when the correlation between the reference frames is low, the B_{ANC} frame has to be coded with more importance and assigned smaller QP. The proposed algorithm utilizes this correlation between reference frames to estimate the importance of B_{ANC} frames. We determine QP for B_{ANC} frame based on its importance.

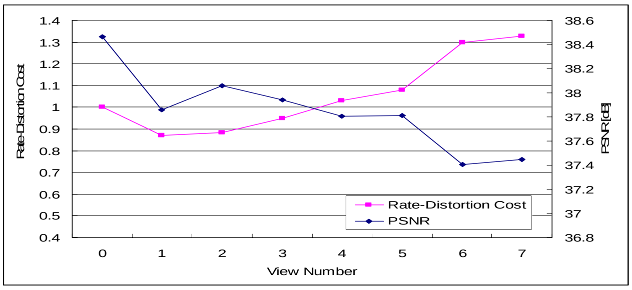


Figure 5. Coding Results of “Exit” Sequences

Fig. 5 presents the coding results of the first anchor frames of “Exit” sequences with basis QP = 31. The anchor frame of view-number 0 sequence is coded as I frame. Other frames are coded as P frames referring to the I frame. The rate-distortion cost of each frame is calculated by

$$RDCost = D + \lambda R \tag{1}$$

and every value is divided by the rate-distortion cost of the I frame for normalization. In Eq. (1), D is distortion caused by encoding each frame and it is calculated by

$$D = \frac{\sum_{x \in X} \sum_{y \in Y} \{org(x, y) - rec(x, y)\}^2}{width \times height \times 1.5} \tag{2}$$

where $org(x,y)$ and $rec(x,y)$ stand for pixel values of the original frame and the reconstructed frame, respectively. R

represents the number of bits per pixel after encoding. It is obtained by

$$R = \frac{total\ bits}{width \times height \times 1.5} \quad (3)$$

λ is the Lagrangian multiplier for rate-distortion cost, which is calculated by

$$\lambda = 0.85 \times 2^{\frac{\min\{52, QP\}}{3} - 4} \quad (4)$$

As the distance between a frame and its reference frame increases, the correlation between them decreases. Therefore, as shown in Fig. 5, the PSNR values decrease and the rate-distortion costs increase. Based on this result, the proposed algorithm estimates the correlation between frames using the rate-distortion costs. QP for B_{ANC} frames is obtained using the rate-distortion cost by

$$QP_{B_{ANC}} = \max\{0, \min\{51, QP_{REF} + dQP\}\} \quad (5)$$

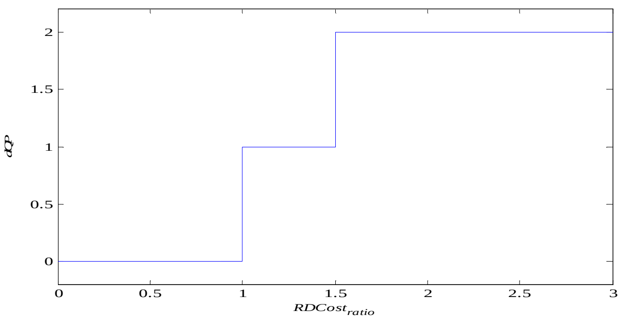
where $QP_{B_{ANC}}$ is the QP for B_{ANC} frames and it is bounded to lie between 0 and 51. QP_{REF} is the basis QP to be used for I frames and P frames. dQP is delta QP value for B_{ANC} frames. dQP is obtained by

$$dQP = \mathrm{ceil}\left(\max\left\{\sqrt{\alpha(RDCost_{ratio} - \beta)}, 0\right\}\right) \quad (6)$$

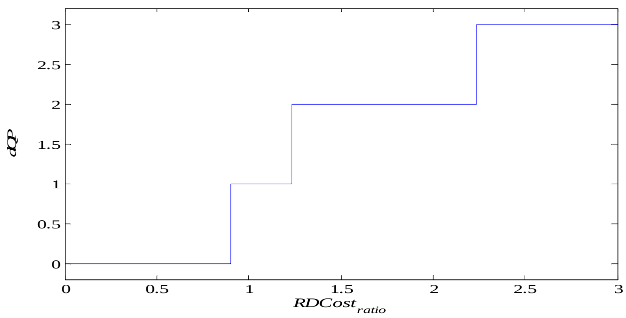
where ceil is the ceiling operation and $RDCost_{ratio}$ is the ratio of the rate-distortion costs which is calculated by

$$RDCost_{ratio} = RDCost_{I} / RDCost_{P} \quad (7)$$

where $RDCost_P$ is the rate-distortion cost of the second P frame which is referred by the B_{ANC} frame. $RDCost_I$ is the rate-distortion cost of the I frame which is at the same time instance as the B_{ANC} frame. When a B_{ANC} frame refers to an I frame and a P frame, α is 2 and β is 1 and when a B_{ANC} frame refers to two P frames, α is 3 and β is 0.9. Fig. 6 shows the variations of dQP.



(a) $\alpha = 2$, $\beta = 1$



(b) $\alpha = 3$, $\beta = 0.9$

Figure 6. Variations of dQP

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed algorithm, many test sequences are used. Characteristics of test sequences are listed in Table 3. Basis QP settings for CE are used for experiments.

Table 3. Characteristics of Test Sequences

Test sequence	Property	Camera Arrangement
Ballroom, Exit	640x480, 25fps	8 cameras with 20cm spacing, 1D/parallel
Uli	1024x768, 25fps	8 cameras with 20cm spacing, 1D/parallel convergent
Race1	640x480, 30fps	8 cameras with 20cm spacing, 1D/parallel
Rena	640x480, 30fps	100 cameras with 5cm spacing, 1D/parallel
Breakdancers	1024x768, 15fps	8 cameras with 20cm spacing, 1D/arc

Table 4 shows the experimental results for “Ballroom” sequences with basis QP = 34, 31, and 29. The proposed algorithm achieved 0.09 dB average quality improvement compared to JSVM. Fig. 7 shows the rate and distortion curve for “Ballroom” sequences. The rate and distortion curve of the proposed algorithm is located upper than that of JSVM at every bitrate.

Table 4. Experimental Results for “Ballroom”

QP	Proposed		JSVM	
	Bitrate [kbps]	PSNR [dB]	Bitrate [kbps]	PSNR [dB]
34	272.58	33.24	269.83	33.17
31	397.37	34.87	392.61	34.78
29	511.65	35.88	503.34	35.77

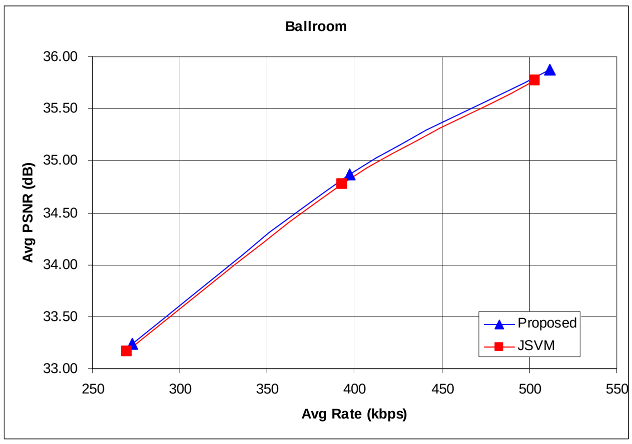


Figure 7. Rate and Distortion Curve for “Ballroom”

Table 5 shows the experimental results for “Exit” sequences with basis QP = 31, 29, and 26. The proposed algorithm achieved 0.13 dB average quality improvement compared to JSVM. Fig. 8 shows the rate and distortion curve for “Exit” sequences. The rate and distortion curve of the proposed algorithm is located upper than that of JSVM.

Table 5. Experimental Results for “Exit”

	Proposed		JSVM	
QP	Bitrate [kbps]	PSNR [dB]	Bitrate [kbps]	PSNR [dB]
31	198.32	36.98	194.48	36.85
29	252.54	37.73	246.63	37.60
26	389.27	38.73	378.74	38.60

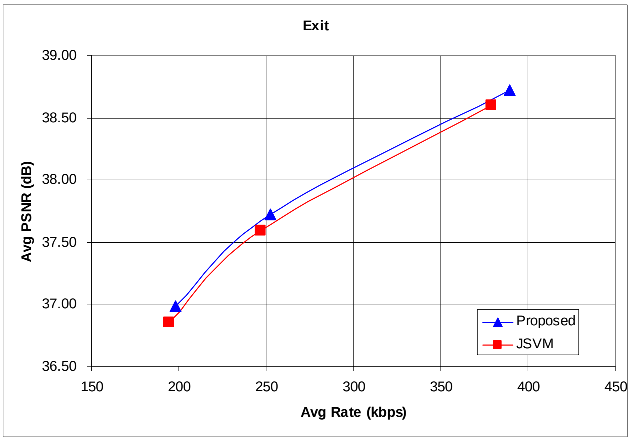


Figure 8. Rate and Distortion Curve for “Exit”

Table 6 shows the experimental results for “Breakdancers” sequences with basis QP = 31, 26, and 22. The proposed algorithm achieved 0.15 dB average quality improvement compared to JSVM. Fig. 9 shows the rate and distortion curve for “Breakdancers” sequences. The rate and distortion curve of the proposed algorithm is located upper than that of JSVM.

Table 6. Experimental Results for “Breakdancers”

	Proposed		JSVM	
QP	Bitrate [kbps]	PSNR [dB]	Bitrate [kbps]	PSNR [dB]
31	245.56	37.59	246.45	37.46
26	476.78	39.07	489.10	38.92
22	977.48	40.02	1040.54	39.86

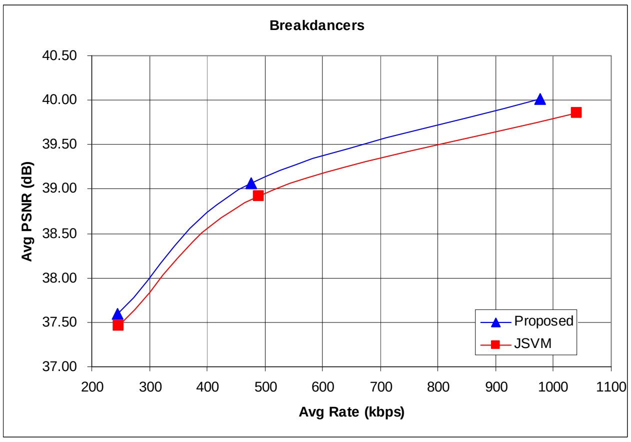


Figure 9. Rate and Distortion Curve for “Breakdancers”

Table 7 shows the experimental results for “Uli” sequences with basis QP = 36, 30, and 28. The proposed algorithm achieved 0.12 dB average quality improvement compared to JSVM. Fig. 10 shows the rate and distortion curve for “Uli” sequences. The rate and distortion curve of the proposed algorithm is located upper than that of JSVM.

Table 7. Experimental Results for “Uli”

	Proposed		JSVM	
QP	Bitrate [kbps]	PSNR [dB]	Bitrate [kbps]	PSNR [dB]
36	778.17	32.29	765.12	32.15
30	1595.03	35.57	1570.64	35.46
28	2012.97	36.57	1988.52	36.47

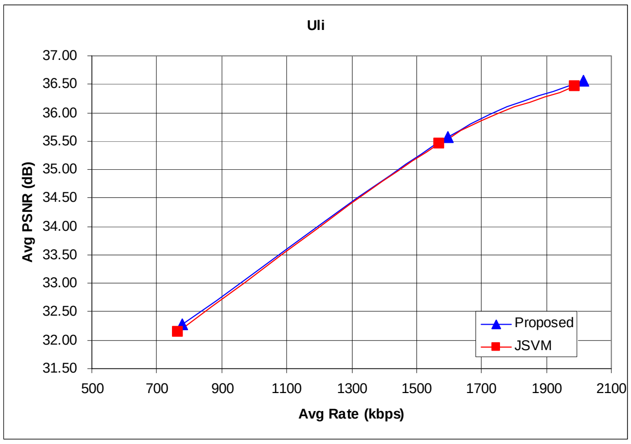


Figure 10. Rate and Distortion Curve for “Uli”

Table 8. Experimental Results for “Rena”

	Proposed		JSVM	
QP	Bitrate [kbps]	PSNR [dB]	Bitrate [kbps]	PSNR [dB]
33	133.28	38.03	131.36	37.96
28	257.67	41.11	254.66	41.03
23	529.82	44.01	518.92	43.90

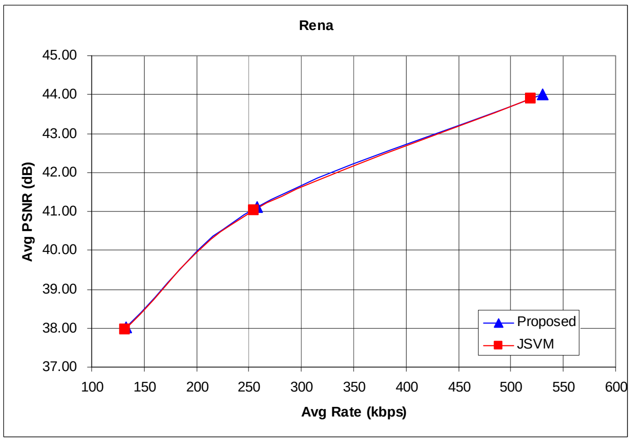


Figure 11. Rate and Distortion Curve for “Rena”

Table 8 shows the experimental results for “Rena” sequences with basis QP = 33, 28, and 23. The proposed algorithm achieved 0.09 dB average quality improvement compared to JSVM. Fig. 11 shows the rate and distortion curve for "Rena" sequences. The rate and distortion curve of the proposed algorithm is located upper than that of JSVM at every bitrate.

Table 9 shows the experimental results for "Race1" sequences with basis QP = 28, 26, and 24. The proposed algorithm achieved 0.16 dB average quality improvement compared to JSVM. "Race1" sequences show the best results amid the test sequences. Fig. 12 shows the rate and distortion curve for "Race1" sequences. The rate and distortion curve of the proposed algorithm is located upper than that of JSVM.

Table 9. Experimental Results for "Race1"

QP	Proposed		JSVM	
	Bitrate [kbps]	PSNR [dB]	Bitrate [kbps]	PSNR [dB]
28	412.57	37.61	405.12	37.46
26	548.07	38.62	536.05	38.45
24	740.26	39.65	724.22	39.50

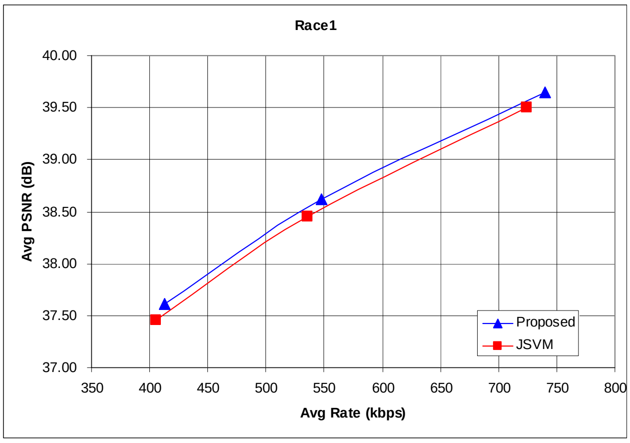


Figure 12. Rate and Distortion Curve for "Race1"

Table 10. Variances of PSNR Values of All Sequences

Test sequence	QP	Proposed	JSVM
Ballroom	34	0.1606	0.2186
	31	0.1622	0.2634
	29	0.1515	0.2843
Exit	31	0.1136	0.1828
	29	0.0942	0.1590
	26	0.0766	0.1312
Breakdancers	31	0.0843	0.2056
	26	0.0487	0.1429
	22	0.0348	0.0946
Uli	36	1.2339	1.7147
	30	0.9824	1.2977
	28	0.8900	1.1422
Rena	33	0.2701	0.3500
	28	0.2514	0.3500
	23	0.2051	0.3200
Race1	28	0.5100	0.5523
	26	0.4955	0.5311
	24	0.4612	0.4930

In Table 10, the variances of the PSNR values of all sequences are listed. By the proposed algorithm, the sequences whose anchor frames are coded as B_{ANC} frames are coded with higher quality than the results by JSVM. Consequently, the proposed algorithm reduced the variances of PSNR values about 33% in average.

5. CONCLUSION

The proposed algorithm adaptively selects QP for B_{ANC} frames by considering the characteristics of the hierarchical B picture structure and the correlation between reference frames. B_{ANC} frames are coded using two reference frames. The proposed algorithm estimates the correlation of reference frames according to the rate-distortion costs of them. QP for B_{ANC} frames is then calculated using this correlation. Thus, the proposed algorithm improves the quality of B_{ANC} frames and coding efficiency of the corresponding sequences. Consequently, compared to the results of JSVM, the proposed algorithm showed 0.07~0.16dB of average PSNR improvement. Also, the variances of the PSNR values of all sequences have decreased about 33% in average.

ACKNOWLEDGEMENTS

This work was supported in part by the Information Technology Research Center (ITRC) through the Realistic Broadcasting Research Center (RBRC) at Gwangju Institute of Science and Technology (GIST), and in part by the Ministry of Education (MOE) through the Brain Korea 21 (BK21) project.

REFERENCES

[1] Survey of Algorithm used for Multi-view Video Coding (MVC), ISO/IEC JTC1/SC29/WG11 N6909, 2005.
[2] Call for Proposals on Multi-view Video Coding, ISO/IEC JTC1/SC29/WG11 N7327, 2005.
[3] Description of Core Experiments in MVC, ISO/IEC JTC1/SC29/WG11 W8019, 2006.
[4] Requirements on Multi-view Video Coding, ISO/IEC JTC1/SC29/WG11 N7539, 2005.
[5] Multiview Coding using AVC, ISO/IEC JTC1/SC29/WG11 M12945, 2006.