Bitrate Efficient 3DoF+ 360 Video View Synthesis for Immersive VR Video Streaming

JongBeom Jeong
Gachon University
(Dept. of Computer Engineering)
Seongnam, South Korea
uof4949@gc.gachon.ac.kr

Dongmin Jang
Gachon University
(Dept. of Computer Engineering)
Seongnam, South Korea
dogzz9445@gc.gachon.ac.kr

Jang-Woo Son
Gachon University
(Dept. of Computer Engineering)
Seongnam, South Korea
sjw6757@gc.gachon.ac.kr

Eun-Seok Ryu
Gachon University
(Dept. of Computer Engineering)
Seongnam, South Korea
esryu@ gachon.ac.kr

***Abstract*— Recently, as the demand for Virtual Reality(VR) is increasing, it is not hard to experience immersive contents with VR. However, a tremendous amount of calculation and bandwidth are required when processing 360 videos. Moreover, to enjoy stereoscopic 360 contents, additional information such as the depth of the video must be provided. Therefore, in this paper, it proposes the efficient method of streaming high quality 360 videos. To reduce the bandwidth when streaming and synthesizing the 3DoF+ 360 video which supports limited movements of the user, the proper down-sampling ratio and quantization parameter are offered with the graph between bitrate and Peak Signal-to-Noise Ratio(PSNR). To encode and decode the 360 video, High-Efficiency Video Coding(HEVC) is used. And view synthesizer produces the video of intermediate view to provide the user with an immersive experience.**

Keywords—VR, 3DoF+, HEVC, view synthesis, VSRS, RVS, FTV, WS-PSNR, Multi-view video coding

I. INTRODUCTION

As the VR market is getting bigger, efficient immersive VR technology is needed. To play high-quality VR video through Head-Mounted Display(HMD), the resolution of the video must be 4K, at least. In this case, the amount of data to be processed from HMD increases rapidly. That's why the Moving Picture Experts Group(MPEG) proposed the technology which processes viewport the user views named Motion-Constrained Tile Set(MCTS)[1] in 2016, and one paper contains MCTS implementation for VR streaming was submitted [2]. Also, to provide the user with high quality 360 videos, region-wise packing[3] was proposed. It encodes Region of Interest(ROI) with high quality, and the other is encoded with low quality.

To support the immersive media, MPEG-I group, established by MPEG, divided the standardization associated with VR into three phases, 3 Degree of Freedom (3DoF), 3DoF+, and 6DoF[4]. In 3DoF+ and 6DoF, multi-view 360 videos are needed, and it consists of texture and depth images to support 3D video[5]. Since those provide 360 videos in response to user's movement, it is inevitable to synthesize the immediate views using existing views. View Synthesis Reference Software(VSRS) for 360 videos [6], Reference View Synthesizer(RVS)[7] and WS-PSNR for 360 video quality evaluation[8] were proposed to MPEG to create virtual views and evaluate them.

When transmitting 3DoF+ or 6DoF 360 videos, it requires lots of bandwidth since those need both high-resolution texture and depth. To overcome this problem, down-sampling or region-wise packing could be applied. In this paper, it proposes the appropriate down-sampling ratio and quantization parameter of 3DoF+ texture and depth in view synthesis. It introduces a pilot test with Super Multiview Video(SMV)[9]. Finally, it provides the graph of bitrate and PSNR obtained by 3DoF+ test sequences using 360lib with HEVC.

II. RELATED WORK

A. Standards of 360 video in MPEG

During the 116th meeting of MPEG, MPEG-I group was established due to the support of immersive media. They started to standardize the format of immersive, omnidirectional video in 2017[10]. Fig.1 shows the standardization roadmap of MPEG. They divided the standardization into 3 phases[11][12]. Phase 1a aims to provide 360 video and contents including stitching, projection, and video encoding.

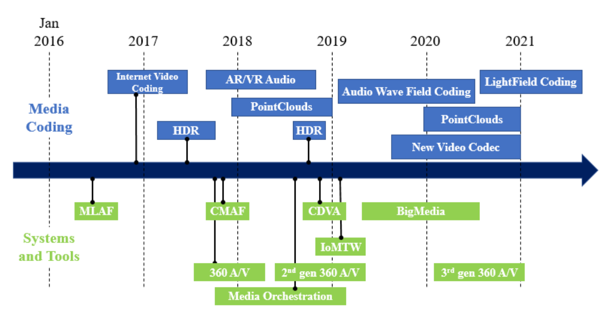


Fig 1. MPEG standardization roadmap

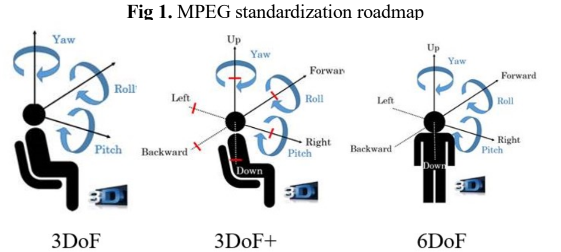


Fig 2. Viewing angle and the degree of freedom of 3DoF, 3DoF+, and 6DoF

Fig.2 shows the 3DoF, 3DoF+, and 6DoF viewing angle and degree of freedom. If a person watches the stereoscopic video, the movement of the user is defined into 3 directions, which mean yaw, pitch, and roll. However, in 3DoF video, it can't represent things behind the objects, which means the limited experience of VR.

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2018-0-00765,Development of Compression and Transmission Technologies for Ultra High Quality Immersive Videos Supporting 6DoF)

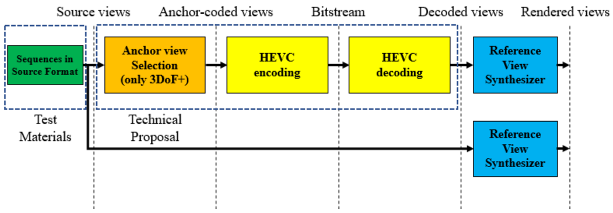


Fig 3. Anchors generation structure of 3DoF+

To overcome the limitation of 3DoF, the concept of 3DoF+, part of phase 1b in MPEG-I, was proposed. 3DoF+ provides limited movements of yaw, pitch, roll, described in Fig.2. So it has increased the degree of freedom in 3DoF, relatively.

In 3DoF+, VR device has to offer the video of view which the user watches. If the view user wants to see is not included in the original video, 3DoF+ system synthesizes the view that doesn't exist before. So, Reference Intermediate View Synthesizer(RIVS)[13] is needed. Also, to synthesize virtual views, additional depth information, contains distances between camera and objects, must be supplied. Since it requires lots of data to be sent, optimization for data transmits and compression ought to be proposed. For this reason, enhanced technologies of communication such as 5G mobile technology[14] and mobile data offloading[15] are announced recently. View synthesis assumes video transmission from the server to client, so the video needs to be compressed as we can see at Fig.3. The anchor view is used in view synthesis, and it has to be encoded and decoded.

Subsequently, phase 2 of MPEG-I deals with 6DoF, which means 3DoF+ with translational movements to X, Y, and Z axes. It supports user's movements including walking, as described in Fig.2.

B. Multi-view video Coding

Multi-view video enables the user to have a three-dimensional and immersive experience. It provides diverse views gained from one scene simultaneously. Especially, three-dimensional multi-view video includes both texture and depth information. It enables users to have multiple views which they want to watch. MPEG defined the three-dimensional video system[16] which is a part of Free-viewpoint TV(FTV), and it contains multi-view video acquisition, encoding, transmission, decoding, and display.. To process the multi-view video efficiently, multi-view video coding[17][18] is required.

Multi-view videos have common features since they contain the same scene at the same time. The difference between each view is that they have an indigenous point of view. What it implies is a multi-view video of one viewpoint can be made by referencing another view.

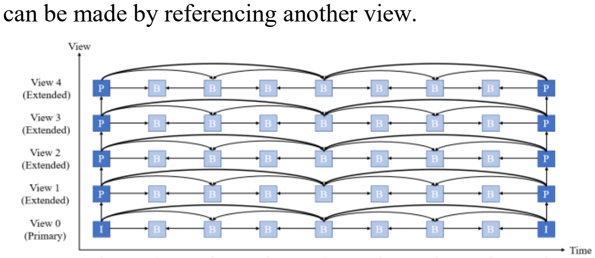


Fig 4. Multi-view video encoding view reference structure

Fig.4 shows the hierarchial B frame multi-view video encoding structure between primary view and extended views. The blue box means a key frame which can be referenced by B frame. The I frame is able to be reconstructed while the P frame references one frame. The B frame references two frames when predicting. Joint Multiview Video Model (JMVM)[19] for reference software model of multi-view video coding was proposed to compress multi-view video while containing compatibility with H.264.

C. View Synthesis

Even though multi-view video provides some views, it cannot offer a view which is out of source views. Since multi-view video coding requires lots of computing power to process and data, the number of views multi-view video can support is limited. Accordingly, view synthesis for multi-view video[20][21] was developed to overcome the limitation of multi-view video coding. When we use view synthesis, we don't have to send all of the source views because it synthesizes dropped views which we didn't send. Also, if the video provider didn't acquire many source views due to the limitation of a resource such as a camera and the amount of data, still we can synthesize the other views the provider didn't offer.

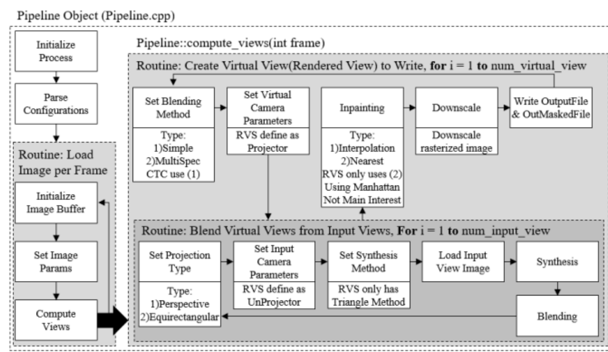


Fig 5. View Synthesis Pipeline of RVS

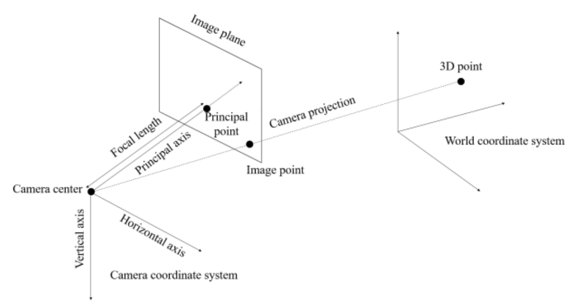


Fig 6. Image projection in a pinhole camera

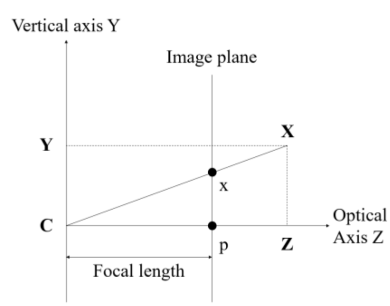


Fig 7. Image plane in camera coordinate system

Fig.5 describes how to synthesize the intermediate views with Reference View Synthesizer (RVS) 1.0.2[22]. It requires texture video, depth map, and a camera parameter. Depth map[23][24] represents the distance between the camera and the object shown in texture video. If the depth map format is 8bit, the range of the depth value is between 0 and 255. The depth map can be gained by a depth camera which uses a depth sensor. Otherwise, it is able to be generated by depth estimation software. MPEG-4 group proposed Depth Estimation Reference Software (DERS) [25][26] to make the depth map from the texture video efficiently.

Generally, the multi-view video is obtained from a pinhole camera. It projects the actual object into a 2D plane image, as shown in Fig.6. To implement the projection, world coordinate system and camera coordinate system is needed. World coordinate system presents three-dimensional space. The camera is located in the world coordinate system, and it also has a three-dimensional coordinate system. The camera center point means the location of the camera in the world coordinate system. Camera coordinate system has X, Y, Z axis. X-axis means horizontal axis, Y axis means the vertical axis, and Z axis means the optical axis, in other words, principal axis. The optical axis is the direction of the camera ray. The principal point is an intersection point between the principal axis and the image plane. The distance from the camera center to the principal is called focal length, as shown in Fig.7. Each point of the object in three-dimensional space is projected onto a two-dimensional image plane by the camera. At that time, explained parameters are used.

To make the intermediate view, the conversion of point coordinates from reference views into synthesized view must be done. Each reference views which is used to synthesize the intermediate view has its own camera coordinate system. If we know the camera parameter of reference views and intermediate view, it is possible to generate camera coordinate system of intermediate view using world coordinate system. Once the conversion is complete, the texture mapping from the reference views to intermediate view can be done.

III. Pilot Test

A. Pilot test with FTV multi-view sequences

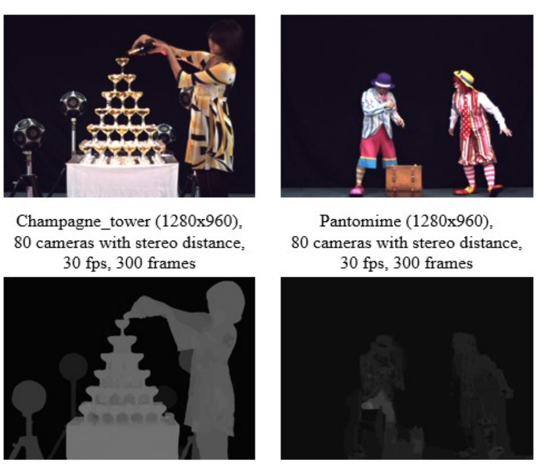


Fig 8. FTV test sequences from Nagoya University

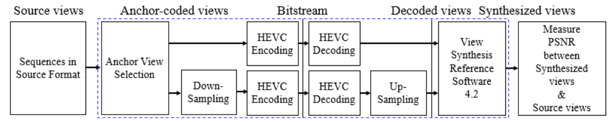


Fig 9. Proposed system architecture with FTV test sequences

TABLE I. COMBINATION OF VIEW SYNTHESIS

Left View	Synthesized View	Right View
37	38	39
37	39	41
39	40	41

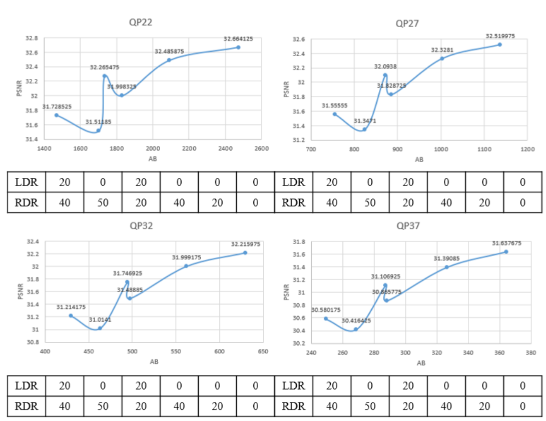
TABLE II. COMBINATION, RESOLUTION OF DOWNSAMPLING RATIO

Down-sampling Ratio	0%	20%	40%	50%	75%
Champagne_tower	1280x960	1024x768	768x576	640x480	320x240
Pantomime	1280x960	1024x768	768x576	640x480	320x240

To reduce the bitrate when transmitting multi-view video, this paper proposes low complexity multi-view video transmit system including down-sampling and up-sampling. The feasibility of this method was proved by a pilot test with FTV multi-view sequences[27]. Champagne_tower and pantomime sequences, as shown in Fig.8, were used. Their resolution is 1280x960, acquired from 80 cameras, and the number of frames is 300.

Fig.9 describes proposed system architecture with FTV multi-view test sequences. Firstly, it selects the anchor view, which means the source view used to synthesize the intermediate view. Test sequences provide the depth map of 37, 39, 41 views, mean anchor view which requires both texture and depth. The combination of view synthesis is represented in Table I. Second, it down-samples the selected anchor views. The down-sampling ratio is 0, 20, 40, 50, and 75(%), as shown in Table II. For down-sampling and up-sampling, Joint Scalable Video Model (JSVM)[28] was used.

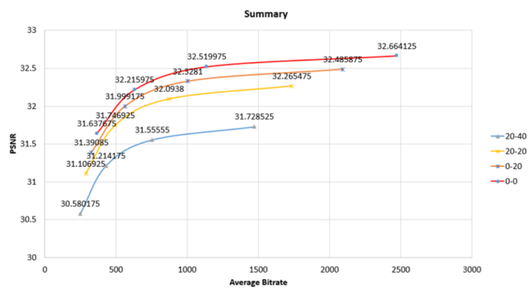
Third, it encodes and decodes the down-sampled views. For encoding and decoding, HEVC reference software (HM) version 16.16 [29] was used. VSRS 4.2 [30] was used to synthesize the intermediate view. Fourth, it up-samples the decoded views. Fifth, it synthesizes the intermediate view by referencing up-sampled anchor views. Finally, it measures Peak Signal-to-Noise Ratio (PSNR) between original intermediate views and synthesized views for objective quality evaluation. For PSNR measurement, JSVM was used.



LDR	20	0	20	0	0	0	LDR	20	0	20	0	0	0
RDR	40	50	20	40	20	0	RDR	40	50	20	40	20	0

LDR	20	0	20	0	0	0	LDR	20	0	20	0	0	0
RDR	40	50	20	40	20	0	RDR	40	50	20	40	20	0

Fig 10. RD-curve between PSNR and average bitrate with different QPs



QP	37	32	27	22

Fig 11. RD-curve between PSNR and average bitrate with different down-sampling ratio combinations

For encoding, the Quantization Parameter (QP) values are 22, 27, 32, and 37. In a pilot test with FTV multi-view sequences, experiment for every combination of down-sampling ratio, QP, and view synthesis was executed. The pilot test result is shown in Fig.10. It describes RD-curve between PSNR and average bitrate with different QPs. The reason why the graph shows the combination 0-0 to 20-40 is it only includes the combinations whose difference values with original view combination(0-0) are under 1.

Even though the average down-sampling ratio of the combination 0-40 (left view-right view) is equal to 20-20, the PSNR value of 20-20 was higher than 0-20. What is more, the average bitrate of 20-20 was smaller than 0-40. Fig.10 implies that uniform down-sampling ratio of left and right view is better. Also, the performance of 20-40 was better than 0-50 because of the equality of down-sampling ratio.

Fig.11 describes the RD-curve between PSNR and average bitrate with different down-sampling ratio combinations. In the case of 20-20, the difference value between QP=27 and QP=22 is 0.171615, which is very low while the difference value of bitrate is 862.6038, which is very high.

B. Pilot test with 3DoF+ test sequences

For the 3DoF+ experiment, MPEG provides Classroom-Video[31], TechnicolorMuseum, and TechnicolorHijack as test sequences described in Fig.12. The pilot test was conducted to ClassroomVideo. View v0 was defined as a synthesized view, v1, v2, v3, v4, v5, v6 were named near views, and v9, v10, v11, v12, v13, v14 were called far views, as shown in Fig.13.

Fig 12. 3DoF+ test sequences

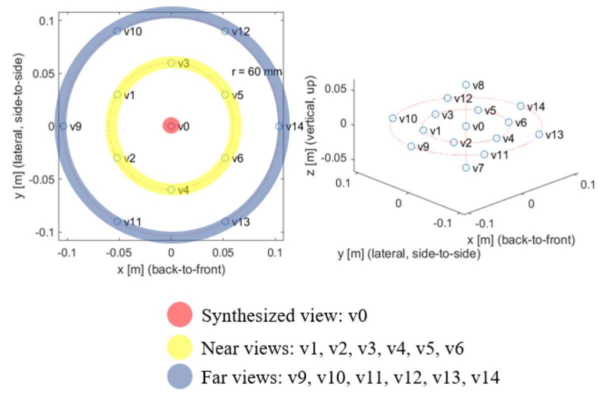


Fig 13. ClassroomVideo viewpoint definition for objective quality evaluation

TABLE III. PSNR FOR SYNTHESIZED VIEWS OF CLASSROOMVIDEO

Input Views (Down-sampling Ratio: 50%)	WS-PSNR_Y	WS-PSNR_U	WS-PSNR_V
(1) nearOrg+farDown	31.8250	48.9032	51.4948
(2)nearOrg+farTextureDown	31.4879	47.8435	50.6719
(3) nearDown+farOrg	31.4080	48.5565	51.1620
(4)nearTextureDown+farOrg	31.4425	43.7384	50.5797
(5) nearOrg+farOrg	32.7309	49.9141	52.4917
(6) nearOrg	31.8293	48.9084	51.5008
(7) farOrg	31.4247	48.5634	51.1593

The distances between the synthesized view and the near views are same, the far views as well. For objective quality evaluation, WS-PSNR tool [32] was used.

Table III shows the PSNR for synthesized views of ClassroomVideo. PSNR value of (6) was higher than (1) even though (6) has fewer views. Adding more views which are down-sampled isn't good for the quality of the synthesized view. If the input views are closer to the synthesized view, its PSNR value was higher as we can see by comparing (1) and (3). Interestingly, the PSNR value of (1) was higher than (2) despite the depth maps of (2) weren't down-sampled while (1) didn't.

IV. Experiment

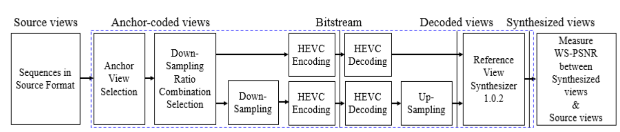


Fig 14. Proposed system architecture

Fig.14 describes the proposed system architecture for 3DoF+ multi-view video transmission including anchor view selection, down-sampling ratio combination selection, down-sampling, encoding, decoding, up-sampling, view synthesis, and measuring WS-PSNR. In 3DoF+ Common Test Condition (CTC), the QPs used for texture and depth is shown in Table IV. The difference value between texture and depth QP is 5, which was decided by an experiment[33]. Table V shows the resolution of down-sampling ratio for ClassroomVideo. The down-sampling is applied to both texture and depth. 360ConvertStatic of 360lib 5.1 was used in down-sampling. Table VI shows the anchor-coded views per class or ClassroomVideo. Class A1 uses all views for view synthesis, while class A2 and A3 uses the subset of views.

TABLE IV. QPs USED FOR TEXTURE AND DEPTH

	R1	R2	40%	50%
Texture QP	22	27	32	37
Depth QP	17	22	27	32

TABLE V. RESOLUTION OF DOWNSAMPLING RATIO

Down-sampling Ratio	0%	12.5%	25%	37.5%	50%
ClassroomVideo	4096x2048	3584x1792	3072x1536	2560x1280	2048x1024

TABLE VI. ANCHOR-CODED VIEWS PER CLASS

Test class	Sequence Name	# of source views	# of anchor-coded views	Anchor-coded views
A1	ClassroomVideo	15	15	All
A2	ClassroomVideo	15	9	v0,v7…v14
A3	ClassroomVideo	15	1	v0

TABLE VII. 3DoF+ VIEW SYNTHESIS FRAME RANGE

Test class	Sequence Name	Frames
A1	ClassroomVideo	89-120
A2	ClassroomVideo	89-120
A3	ClassroomVideo	89-120

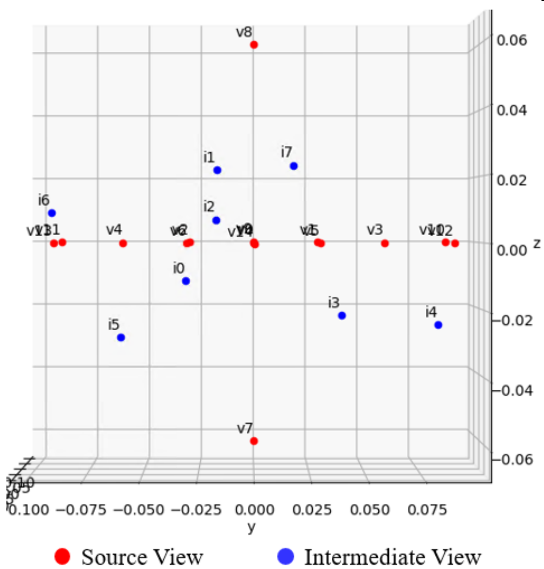


Fig 15. Source and intermediate view position visualization from X axis

To reduce the RVS runtime, frame ranges for view synthesis was set in CTC as shown in Table VII. Since proposals for 3DoF+ are required to generate ERP video for all intermediate view positions, the experiment was designed to synthesize the intermediate views using A1, A2, A3 class views. Fig.15 shows the positions of source and intermediate views, which weren't exist before.

The used version of RVS is 1.0.2 with opencv 3.4.1, and the server used for experiment has 2 Intel Xeon E5-2687w v4 CPU and 128GB. Table VIII shows the summary of WS-PSNR_Y with different down-sampling ratios for regular outputs and masked outputs in synthesizing the intermediate views.

TABLE VIII. WS-PSNR_Y OF SYNTHESIZED VIEWS

WS-PSNR_Y	Regular output			Masked output		
Down-Sampling Ratio	ClassroomVideo					
	A1	A2	A3	A1	A2	A3
0	39.46dB	38.32dB	29.16dB	39.34dB	37.35dB	26.70dB
12.5	37.92dB	37.25dB	29.10dB	37.78dB	36.54dB	26.86dB
25	37.33dB	36.71dB	29.03dB	37.21dB	36.11dB	26.84dB
37.5	36.55dB	36.00dB	28.90dB	36.45dB	35.51dB	26.83dB
50	35.30dB	34.85dB	28.70dB	35.22dB	34.49dB	26.86dB

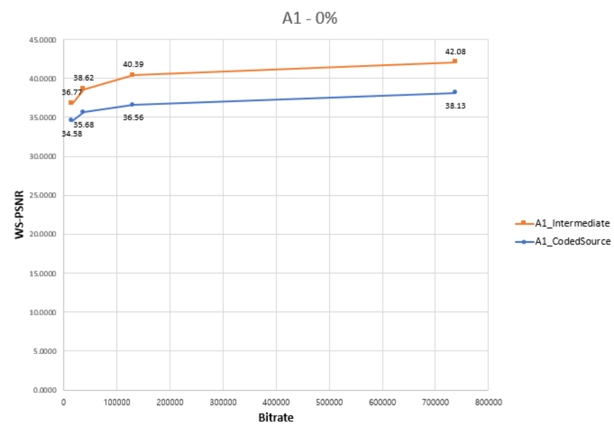


Fig 16. RD-curve between WS-PSNR_Y and Bitrate of A1 with 0% down-sampling

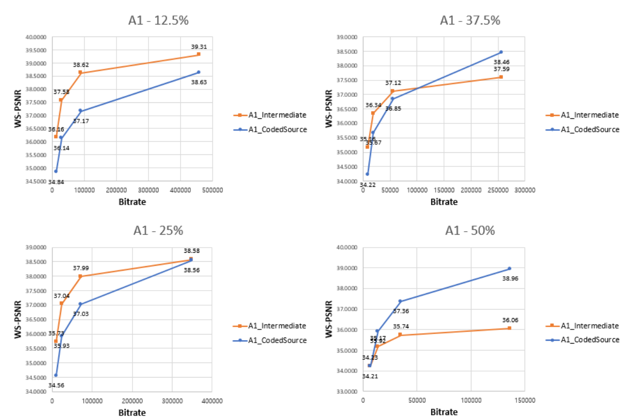


Fig 17. RD-curve between WS-PSNR_Y and Bitrate of A1 with 12.5%, 25%, 37.5%, and 50% down-sampling

Table VIII shows the WS-PSNR_Y values of synthesized intermediate views. Result of the regular output was better than the masked output's. Also, class A2and A3, which discarded some source views showed low WS-PSNR. If the anchor views must be down-sampled, the ratio 12.5% is reasonable.

Fig.16 shows the RD-curve between WS-PSNR_Y and bitrate of A1 with 0% down-sampling. The values of the X axis used 37, 32, 27, 22 for texture QP, and 32, 27, 22, 17 for depth QP. QP 27 can be used instead of 22 for texture since the gap between QP=27 and QP=22 wasn't high.

With QP=27 and 12.5% of down-sampling ratio, it saved about 87.81% bitrate while losing only 8% WS-PSNR.

V. CONCLUSION

This study proposes a bitrate-reducing method for 3DoF+ video synthesis and transmission. Especially, by down-sampling and up-sampling the texture and depth, the proposed method saved lots of bitrate while losing only a few WS-PSNR value. However, since the condition of the experiment wasn't diverse enough to deduct the optimal parameter for view synthesis, the experiment using video compression methods such as region-wise packing[34] have to be progressed.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No.2018-0-00765,Development of Compression and Transmission Technologies for Ultra High-Quality Immersive Videos Supporting 6DoF).

REFERENCES

[1] Ye-Kui Wang, Hendry, Marta Karczewicz, "Viewport dependent processing in VR: partial video decoding", 116th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG 116/ m38559, 2016.

[2] Jang-Woo Son, Dongmin Jang, Eun-Seok Ryu, "Implementing Motion-Constrained Tile and Viewport Extraction for VR Streaming", ACM Network and Operating System Support for Digital Audio and Video 2018 (NOSSDAV2018), 2018.

[3] Sejin Oh, Soojin Hwang, "OMAF: Generalized signaling of region-wise packing for omnidirectional video", 118th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2017/ m40423, 2017.

[4] Joel Jung, Bart Kroon, Renaud Doré, Gauthier Lafruit, Jill Boyce, "Update on N17618 v2 CTC on 3DoF+ and Windowed 6DoF", 123rd MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG123/m43571.

[5] Masayuki Tanimoto, Toshiaki Fujii, "FTV – Free Viewpoint Television", 61th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2002/m8595, 2002.

[6] Takanori Senoh, Nobuji Tetsutani, Hiroshi Yasuda, "MPEG-I Visual: View Synthesis Reference Software (VSRSx)", 123rd MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2018/m42911, 2018.

[7] Bart Kroon, Gauthier Lafruit, "Reference View Synthesizer (RVS) 2.0 manual", 123rd MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2018/n17759, 2018.

[8] Yule Sun, Ang Lu, Lu Yu, "WS-PSNR for 360 video quality evaluation", 115th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2016/m38551, 2016.

[9] Takanori Senoh, Krzysztof Wegner, Olgierd Stankiewicz, Gauthier Lafruit, Masayuki Tanimoto, "FTV test material summary", 116th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2016/n16521, 2016.

[10] WG11(MPEG), "MPEG Strategic Standardisation Roadmap", 115th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2016/n16316, 2016.

[11] Mary-Luc Champel, Rob Koenen, Gauthier Lafruit, Madhukar Budagavi, "Working Draft 0.4 of TR: Technical Report on Architectures for Immersive Media", 120th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2017/n17264, 2017.

[12] Yo-Sung Ho, "MPEG-I 표준과 360 도 비디오 콘텐츠 생성", The Magazine of the IEEE 44(8), 52-57, Aug 2017.

[13] Renaud Doré, Julien Fleureau, Bertrand Chupeau, Gerard Briand, "3DoFPlus Intermediate View Synthesizer proposal", 122nd MPEG meeting of ISO/IEC JTC1/WG11, MPEG2018/m42486, 2018.

[14] Rupendra Nath Mitra, Dharma P. Agrawal, "5G mobile technology: A survey", ICT Express 1 (2015) 132-137, 2016.

[15] Yeongjin Kima, Joohyun Leeb, Jaeseong Jeong, Song Chong, "Multi-flow management for mobile data offloading", ICT Express 3 (2017) 33-37, 2016.

[16] ISO/IEC JTC1/SC29/WG11, "Introduction to 3D Video", MPEG2008/n9784, 2008.

[17] Philipp Merkle, Aljoscha Smolic, Karsten Müller, and Thomas Wiegand, "Multi-view video plus depth representation and coding", 2007 IEEE International Conference on Image Processing, pp. 201-204, 2007.

[18] Karsten Müller, Heiko Schwarz, Detlev Marpe, Christian Bartnik, Sebastian Bosse, Heribert Brust, Tobias Hinz, Haricharan Lakshman, Philipp Merkle, Franz Hunn Rhee, Gerhard Tech, Martin Winken, and Thomas Wiegand, "3D High-Efficiency Video Coding for Multi-View Video and Depth Data", IEEE Transactions on Image Processing, VOL. 22, No.9, 2013.

[19] Anthony Vetro, Purvin Pandit, Hideaki Kimata, Aljoscha Smolic, "Joint Multiview Video Model (JMVM) 7.0", 83rd MPEG meeting of ISO/IEC JTC1/WG11, n9578, 2008.

[20] Emin Martinian, Alexander Behrens, Jun Xin, Anthony Vetro, "View Synthesis for Multiview Video Compression", Picture Coding Symposium, 2006.

[21] Sehoon Yea, Anthony Vetro, "View synthesis prediction for multiview video coding", Signal Processing: Image Communication, Volume 24, Issues 1-2, pp. 89-100, 2008.

[22] Sarah Fachada, Bart Kroon, Daniele Bonatto, Bart Sonneveldt, Gauthier Lafruit, "Reference View Synthesizer (RVS) 1.0.2 manual", 123rd MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG123/m42945.

[23] Jun Zhang, Miska M. Hannuksela, Houqiang Li, "Joint Multiview Video Plus Depth Coding", 2010 IEEE International Conference on Image Processing, Hong Kong, pp. 2865-2868, 2010.

[24] Yo-Sung Ho, Kwan-Jung Oh, Cheon Lee, SangBeom Lee, San-Tae Na, "Depth Map Generation and Depth Map Coding for MVC", 81th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2007/m14638, 2007.

[25] Masayuki Tanimoto, Toshiaki Fujii, Kazuyoshi Suzuki, "Depth Estimation Reference Software (DERS) with Image Segmentation and Block Matching", 87th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2009/m16092, 2009.

[26] Masayuki Tanimoto, Toshiaki Fujii, Mehrdad Panahpour Tehrani, Kazuyoshi Suzuki, Menno Wildeboer, "Depth Estimation Reference Software (DERS) 3.0", 88th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2009/m16390, 2009.

[27] ISO/IEC JTC1/SC29/WG11, "FTV test material summary", MPEG2016/n16521, 2016.

[28] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Joint Scalable Video Model, Doc. JVT-X202, 2007.

[29] Joint Collaborative Team on Video Coding (JCT-VC), "HEVC reference software version HM16.16," downloadable at https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.16.

[30] Takanori Senoh, Kenji Yamamoto, Nobuji Tetsutani, Hiroshi Yasuda, Krzysztof Wegner, "View Synthesis Reference Software (VSRS) 4.2 with improved inpainting and hole filling", 118th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2017/m40657, 2017.

[31] Bart Kroon, "3DoF+ test sequence ClassroomVideo", 122nd MPEG meeting of ISO/IEC JTC1/WG11, MPEG2018/m42415, 2018.

[32] ISO/IEC JTC1/SC29/WG11, "ERP WS-PSNR Software Manual", w17760, 2018.

[33] Bin Wang, Yule Sun, Lu Yu, "On Depth Delta QP in Common Test Condition of 3DoF+ Video", 123rd MPEG meeting of ISO/IEC JTC1/WG11, MPEG2018/m43801, 2018.

[34] Sejin Oh, Jangwon Lee, "OMAF: Signaling of Projection/Region-wise Packing Information of Omnidirectional Video in ISOBMFF", 117th MPEG meeting of ISO/IEC JTC1/WG11, MPEG2018/m39865, 2017.