

Special Paper

방송공학회논문지 제24권 제7호, 2019년 12월 (JBE Vol. 24, No. 7, December 2019)

<https://doi.org/10.5909/JBE.2019.24.7.1209>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

# A Study on Projection Conversion for Efficient 3DoF+ 360-Degree Video Streaming

Jong-Beom Jeong<sup>a)</sup>, Soonbin Lee<sup>b)</sup>, Dongmin Jang<sup>a)</sup>, Sungbin Kim<sup>b)</sup>, Sangsoon Lee<sup>b)</sup>, and Eun-Seok Ryu<sup>a)‡</sup>

## Abstract

The demand for virtual reality (VR) is rapidly increasing. Providing the immersive experience requires much operation and many data to transmit. For example, a 360-degree video (360 video) with at least 4K resolution is needed to offer an immersive experience to users. Moreover, the MPEG-I group defined three degrees of freedom plus (3DoF+), and it requires the transmission of multiview 360 videos simultaneously. This could be a burden for the VR streaming system. Accordingly, in this work, a bitrate-saving method using projection conversion is introduced, along with experimental results for streaming 3DoF+ 360 video. The results show that projection conversion of 360 video with *360lib* shows a Bjontegaard delta bitrate gain of as much as 11.4%.

Keyword : 360-Degree Video, 3DoF+, HEVC, Virtual Reality, 360lib, Projection

---

a) Department of Computer Education, Sungkyunkwan University

b) Department of Computer Engineering, Gachon University

‡ Corresponding Author : Eun-Seok Ryu

E-mail: [esryu@skku.edu](mailto:esryu@skku.edu)

Tel: +82-2-760-0677

ORCID: <https://orcid.org/0000-0003-4894-6105>

※ This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2017-0-00307, Development of Tiled Streaming Technology for High Quality VR Contents RealTime Service). This work was also supported by Institute of Information & communications Technology Planning & Evaluation (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).

· Manuscript received October 1, 2019; Revised October 24, 2019; Accepted October 24, 2019.

## 1. Introduction

Recently, the demand for efficient virtual reality (VR) technology has been increasing, because the scale of the VR market has expanded. For an immersive experience for users with head-mounted display (HMD) in VR, at least a video resolution of 4K is required. As a result, the number of transmission data and amount of calculation become a huge burden for the system. Especially on mobile platforms, because the amount of computing power is limited, asymmetric multicore processing[9] and 360-degree video (360 video) streaming over millimeter-wave communica-

tion[5] have been proposed. Similarly, tile-based 360 video streaming[11][13] has been proposed to process tiles that are needed by users. To overcome the difficulties, the moving picture experts group (MPEG) established MPEG-I to support standardization for immersive media until 2021. For example, bitrate-efficient technologies such as a motion-constrained tile set (MCTS)[10], were proposed for MPEG-I and implementation was reported[12].

In MPEG-I, the degrees of freedom (DoF) for immersive media is divided into 3DoF, 3DoF+, and 6DoF. Among them, 3DoF+ enables a user sitting in a chair to watch a video with limited movement. For this, multiple 360 video transmission is required, which means that the individual quality of video is declines, because there are a number of videos to be sent, and the bandwidth is limited. Accordingly, a method of saving the bitrate by down-sampling the 3DoF+ 360 video was proposed[2]. However, if the 360 video is down-sampled, loss of the information of the original video must occurs. To minimize this problem, a bitrate-saving method is proposed for 3DoF+ 360 video transmission with appropriate projection conversion in 360 omnidirectional video. Because the 360 video can be represented by various projections, it is possible to reduce the size of 360 video by projection conversion. In this study, 360 library was used to convert the projection of 360 video, and a high-efficiency video coding (HEVC) test model

(HM) was adopted to encode and decode.

As shown in Fig. 1, multiview 360 video compression with projection conversion is proposed. first, multiple 360 videos are acquired from 360 cameras. In pre-processing, the projection type of the videos is converted to reduce the size of the videos. Subsequently, the videos are compressed and transmitted through a network. In post-processing, the compressed videos are decompressed, and converted to the original projection.

This work consists of five sections: section II introduces the related work. For instance, 360 video standardization from MPEG, 360 video projection, and view synthesis. Section III shows the overall experiment, including the bitrate reducing method. Section IV summarizes the experimental results for the proposed method. Finally, in Section 5, the conclusions are presented, and future work is discussed.

## II. Related Work

### 1. 360 Video Standardization from MPEG

The MPEG-I ad-hoc group was established to take charge of 360 video standardization. As mentioned in Section I, in MPEG-I, DoF for immersive media is divided into 3DoF+, 3DoF+, and 6DoF[16], as shown in Fig. 2.

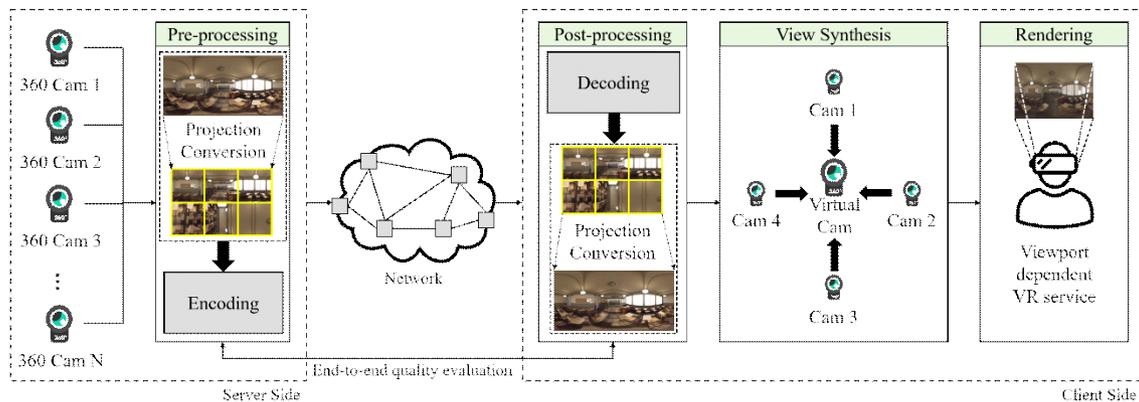


Fig. 1. Multiview 360 video transmission with projection conversion conceptual diagram

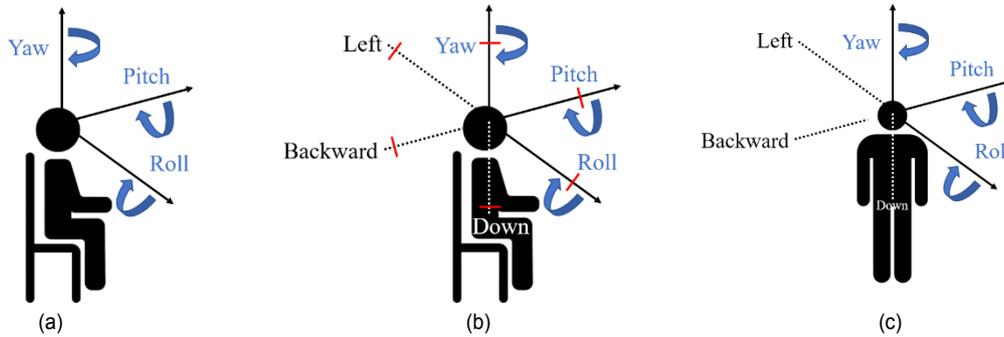


Fig. 2. Viewing angle and degree of freedom : (a) 3DoF, (b) 3DoF+, and (c) 6DoF

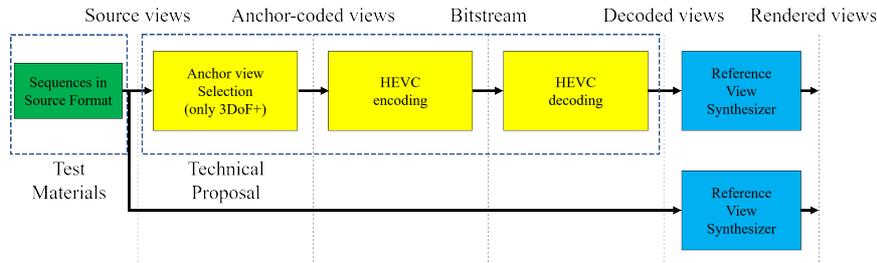


Fig. 3. Definition of the anchor in 3DoF+

In 3DoF, an user sitting in a chair can watch the 360 video acquired from one viewpoint, which restricts the user's immersive experience. In 3DoF+, an user is also supposed to sit in a chair. However, it provides a viewport in accordance with the user's head movement. Finally, in 6DoF, it makes it possible for the user to walk around while watching 360 video, with support of the user's movement.

In 3DoF+ and 6DoF, it is essential to synthesize the intermediate view which did not exist before with source views. Source views mean the videos acquired from cameras. Because the bandwidth to transmit 360 videos is limited, it is inevitable that the quality of each 360 video declines. Otherwise, sending the subset of source views is considerable in 3DoF+ and 6DoF. That is why MPEG-I defined the anchor view, which means the source view to transmit, as shown in Fig. 3.

## 2. 360 Video Projection

360 video is mapped into a sphere. However, when

transmitting the 360 video, it is necessary to map it to the 2D plane, and this is called "projection". Omnidirectional media format (OMAF)[8], defines the standards for omnidirectional 360 video. For instance, equirectangular projection (ERP), cubemap projection (CMP), adjusted cubemap projection (ACP), octahedron projection (OHP), segmented sphere projection (SSP), and rotated sphere projection (RSP) are the projection types for 360 video defined in OMAF.

Among them, ERP is the general and most widely used format for representing 360 video. It projects the video on the sphere into the rectangular 2D plane, as shown in Fig. 4. However, the top and bottom areas have distortion because pixels of those areas mapped in the sphere overlap when they are projected onto the rectangular 2D plane.

In CMP, which is widely used, there are six square faces, as shown in Fig. 4. By inscribing the sphere with video in a regular hexahedron, each face is projected into the squares. In some cases, it can reduce the size of the video compared with the ERP. However, there are distortions on each edge of the squares. Video with CMP inefficiently

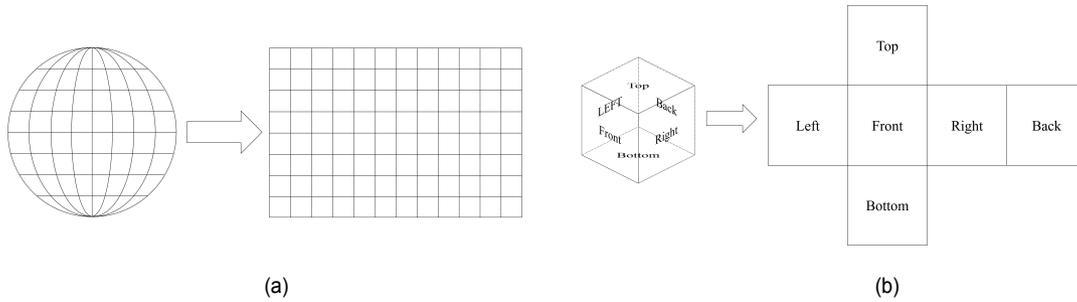


Fig. 4. Examples of 360 video projection format: (a) Equirectangular projection and (b) Cubemap projection

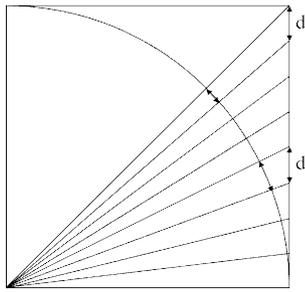


Fig. 5. Cubemap sampling

represents the sphere because it samples the pixels nonuniformly. In Fig. 5, the areas closer to the cube side edges are more densely sampled. Consequently, ACP was developed to overcome these problems. It processes approximate uniform sphere sampling while preserving the packing scheme. As a result, ACP enlarges the center of

the cubemap in CMP, as shown in Fig. 6.

### 3. View Synthesis

3DoF+, as described in Section II, Subsection 1, supports an user's head movements while the user is sitting in a chair. Therefore, in 3DoF+, it is necessary to synthesize the video of the user's viewpoint, which is called "view synthesis". Previously, free viewpoint television (FTV)[7] from MPEG, adopted view synthesis reference software (VSRS)[6] as a tool for view synthesis in multi-view video system. In 2018, VSRSx[15], which supports 360 video view synthesis, was proposed, and it supports two or four input views to synthesize a virtual view. However, there are more than four videos in 3DoF+ test sequences which are introduced in common

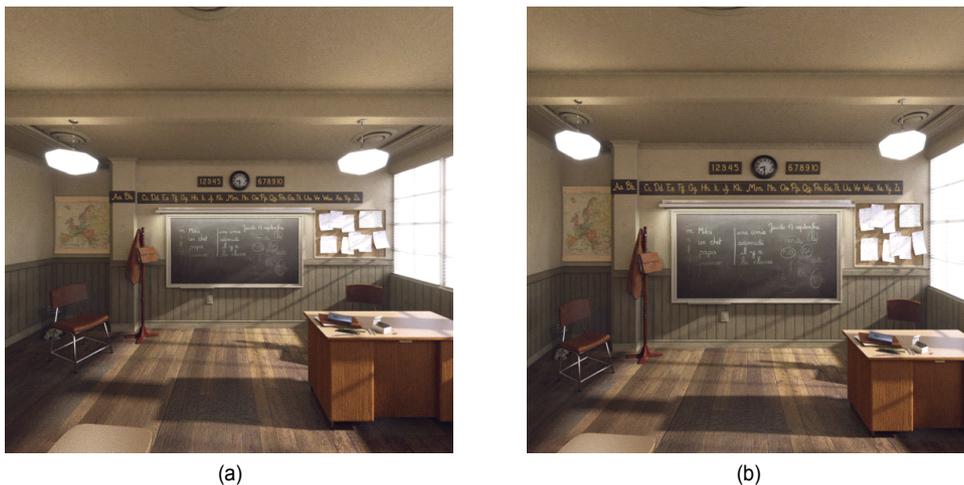


Fig. 6. Front face representations of CMP and ACP: (a) 3x2 cubemap projection (CMP) and (b) 3x2 adjusted cubemap projection (ACP)

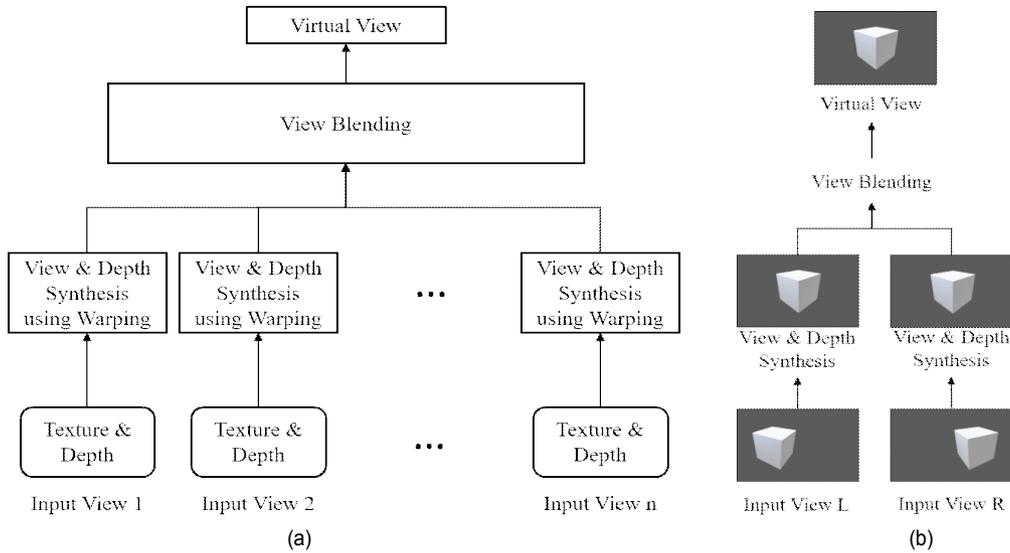


Fig. 7. (a) Conceptual diagram of RVS and (b) simple example of view synthesis in RVS

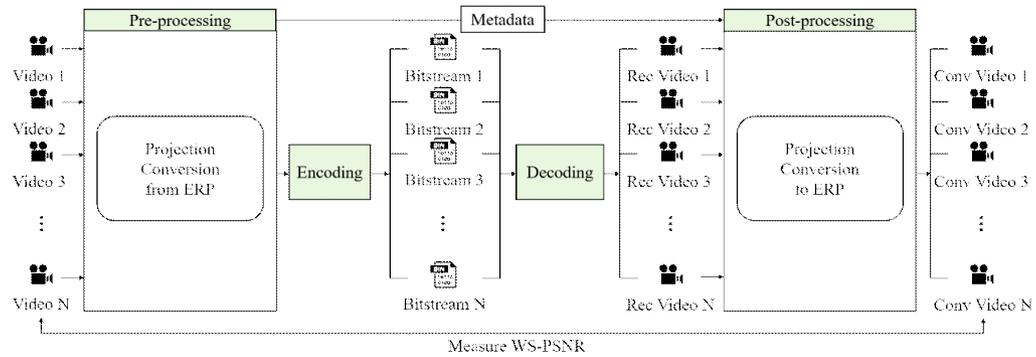


Fig. 8. 3DoF+ 360 video transmission bitrate saving with projection conversion

test conditions (CTC) for 3DoF+[3].

Consequently, a reference view synthesizer (RVS)[1] was adopted as a tool for view synthesis in 3DoF+, because it supports more input views than VSRS. Fig. 7. shows the conceptual diagram and example of view synthesis in RVS. It obtains texture and depth as input views. “Texture” means the video that represents the color of the objects, and “depth” means the distance between the camera and the objects. If a number of input views and the location of the target view are provided to the RVS, it generates target virtual view texture and depth for each input view using warping. Then, it blends the synthesized virtual views.

### III. 3DoF+ 360 Video Transmission Bitrate Saving with Projection Conversion

In this section, a 3DoF+ 360 video bitrate saving method with projection conversion when transmitting, as shown in Fig. 8, is explained. In pre-processing,  $n$  360 ERP videos are converted with different projection techniques such as CMP. The purpose of this phase is to reduce the size of the 360 videos with minimum loss of information. Then, the converted videos are encoded by an HEVC encoder for transmission. Next, the encoded bitstreams are decoded and delivered to the next phase. In post-processing, the projection of the reconstructed videos is converted to ERP, which is the

original projection format of the source videos. Finally, the weighted-to-spherically-uniform peak signal-to-noise ratio (WS-PSNR)[14][17] between the source views and the converted views is measured. Because the source view projection is ERP, WS-PSNR, a weighted metric to measure the quality of reconstructed video, is selected as a tool for objective quality evaluation.

### 1. Projection Conversion with 3DoF+ Test Sequences

Here, projection conversion with a 3DoF+ test sequence is introduced. MPEG has approved ClassroomVideo[4], TechnicolorMuseum, and TechnicolorHijack for 3DoF+ test materials, as shown in Fig. 9. Table 1 contains the general characteristics of the test sequences. ClassroomVideo was

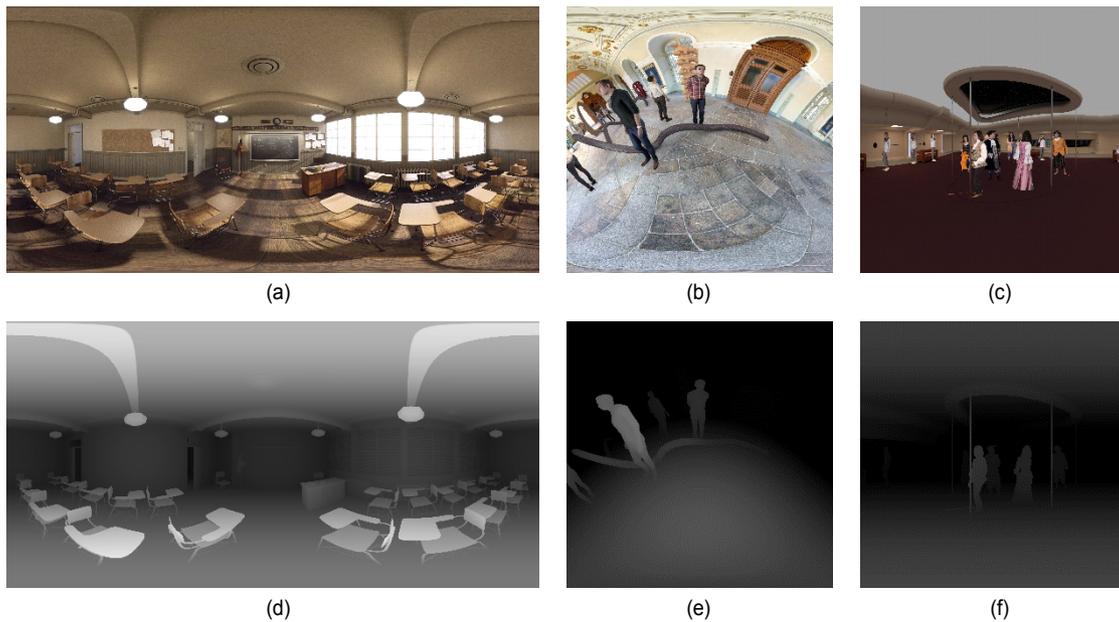


Fig. 9. (a) ClassroomVideo texture, (b) TechnicolorMuseum texture, (c) TechnicolorHijack texture, (d) ClassroomVideo depth, (e) TechnicolorMuseum depth, and (f) TechnicolorHijack depth

Table 1. General characteristics of 3DoF+ test sequences

Name	Resolution	No. of source views	Frame length	View FoV	Global FoV	Frame rate	Bit depth
<i>ClassroomVideo</i>	4096×2048	15	120	360°×180°	360°×180°	30 fps	10-bit
<i>TechnicolorMuseum</i>	2048×2048	24	300	180°×180°	360°×180°	30 fps	10-bit
<i>TechnicolorHijack</i>	4096×4096	10	300	180°×180°	180°×180°	30 fps	10-bit

Table 2. Projection conversion results for texture

View number	Projection type	Original resolution	Converted view resolution	Processing time (s)
v0	CMP3×2	4096×2048	3456×2304	118
	ACP	4096×2048	3456×2304	118
	COHP1	4096×2048	2688×3136	157
	COHP2	4096×2048	6208×1344	155

selected as the test material in this study, because it is the omnidirectional 360 video among the test sequences, which is appropriate for the projection conversion. An experiment for projection conversion was conducted to find an appropriate projection format that shows a few difference values of WS-PSNR in comparison with ERP, the original projection format. Another purpose of this experiment was to find the projection format that is suitable for real-time processing in 3DoF+, because it requires the ability to process a number of 360 videos simultaneously.

For appropriate projection format selection, CMP3×2, ACP, compact OHP option 1 (COHP1), and compact OHP option 2 (COHP2) were selected for the experiment. Fig.

10 shows the converted pictures of position v0 with CMP3×2, ACP, COHP1, and COHP2, and Table. 2 shows the projection conversion results including converted view resolution and processing time.

#### IV. Experimental Results

In Section III, projection conversion with CMP3×2, ACP, COHP1, and COHP2 was mentioned. The *360lib* 5.0, WS-PSNR, and HM 16.16 are used as, discussed in Section III. In this experiment, a Linux server with 2 Intel Xeon E5-2687w v4 CPU and 128GB memory was used. Table 3

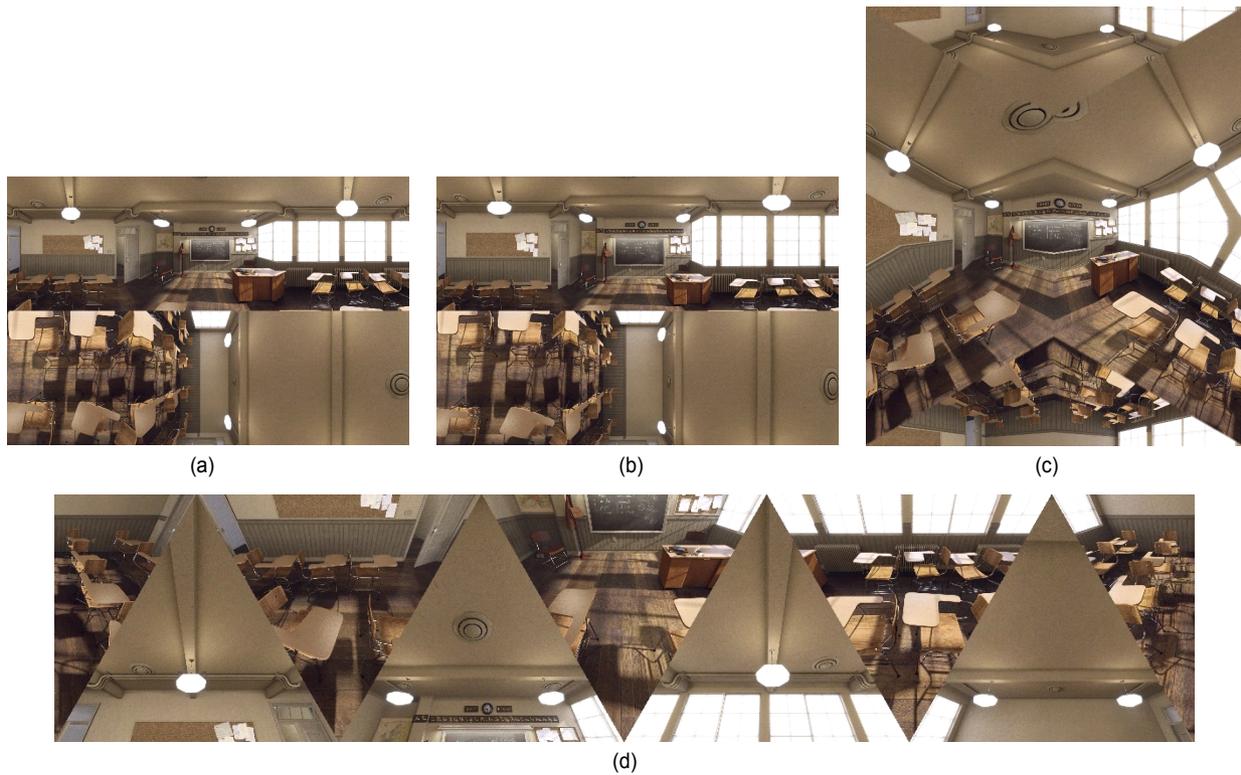


Fig. 10. Converted projection. (a) CMP3×2, (b) ACP, (c) COHP1, (d) COHP2

Table 3. Anchor views per class

Sequence name	Test class	No. of source views	No. of anchor-coded views	Anchor-coded views
ClassroomVideo	A1	15	15	All
	A2	15	9	v0, v7, ..., v14

shows the anchor views for each class of Classroom Video. Class A1 contains all the views, while A2 contains a subset of source views. In this study, an experiment for A1 was introduced to check the result for all the source views.

Tables 4 and 5 summarize the experimental results. Table 4 shows the Bjontegaard delta bitrate (BD-rate), encoding time saving, and pixel rate saving for each projection compared with the anchor, which is ERP. ACP showed the best results among the introduced projection types. For the Y value, it showed an 11.4% BD-rate saving compared with the anchor. In terms of the complexity, encoding time and pixel rate were measured. For example, CMP3×2 showed a 0.03% encoding time increase, which

is still considerable as long as the BD-rate is better than the anchor. The pixel rate saving is the reduced number of pixels. CMP3×2 and ACP showed a 0.05% pixel saving, which decreased the burden for the decoders at the client side. In Table 5, the details of experimental results for both texture and depth projection conversion results are shown. Fig. 11 shows the rate distortion (RD)-curve between the bitrate and WS-PSNR\_Y, which is the luma value of the WS-PSNR, to visualize the results individually, and Fig. 11(e) integrates the results shown in Fig. 11. Among the projection types, ACP showed the best performance considering the bitrate and WS-PSNR\_Y. Because ACP considers the pixel distortion of CMP3×2, it showed a better

Table 4. Results for texture between ERP and the proposed projection types

Projection type	BD-rate			Encoding time saving	Pixel rate saving
	Y	U	V		
CMP3×2	3.0%	-22.0%	-27.3%	0.03%	-0.05%
ACP	-11.4%	-30.4%	-35.5%	0.04%	-0.05%
COHP1	10.9%	-2.0%	-10.2%	0.08%	0.04%
COHP2	9.5%	-8.6%	-17.0%	0.07%	-0.00%

Table 5. Bitrate and WS-PSNR\_Y of A1 texture and depth

A1 Projection type	Texture			Depth		
	QP	Bitrate (Mbps)	WS-PSNR_Y (dB)	QP	Bitrate (Mbps)	WS-PSNR_Y (dB)
ERP (Anchor)	22	724.41	38.13	12	25.34	62.22
	27	120.85	36.56	17	17.14	59.42
	32	30.87	35.67	22	11.45	56.19
	37	11.37	34.57	27	7.14	52.68
CMP3×2	22	515.96	37.42	12	40.52	51.58
	27	94.08	36.37	17	24.71	51.20
	32	26.24	35.60	22	14.64	50.49
	37	10.19	34.62	27	7.95	49.16
ACP	22	500.66	37.58	12	39.99	52.70
	27	90.73	36.47	17	24.39	52.20
	32	25.88	35.69	22	14.18	51.27
	37	10.41	34.72	27	7.65	49.67
COHP1	22	542.79	37.53	12	43.11	51.19
	27	102.64	36.42	17	26.30	50.81
	32	30.87	35.61	22	15.32	50.05
	37	12.57	34.57	27	8.23	48.70
COHP2	22	535.87	37.49	12	45.59	51.19
	27	100.79	36.40	17	27.94	50.80
	32	30.33	35.63	22	16.82	50.05
	37	12.66	34.65	27	9.78	48.73

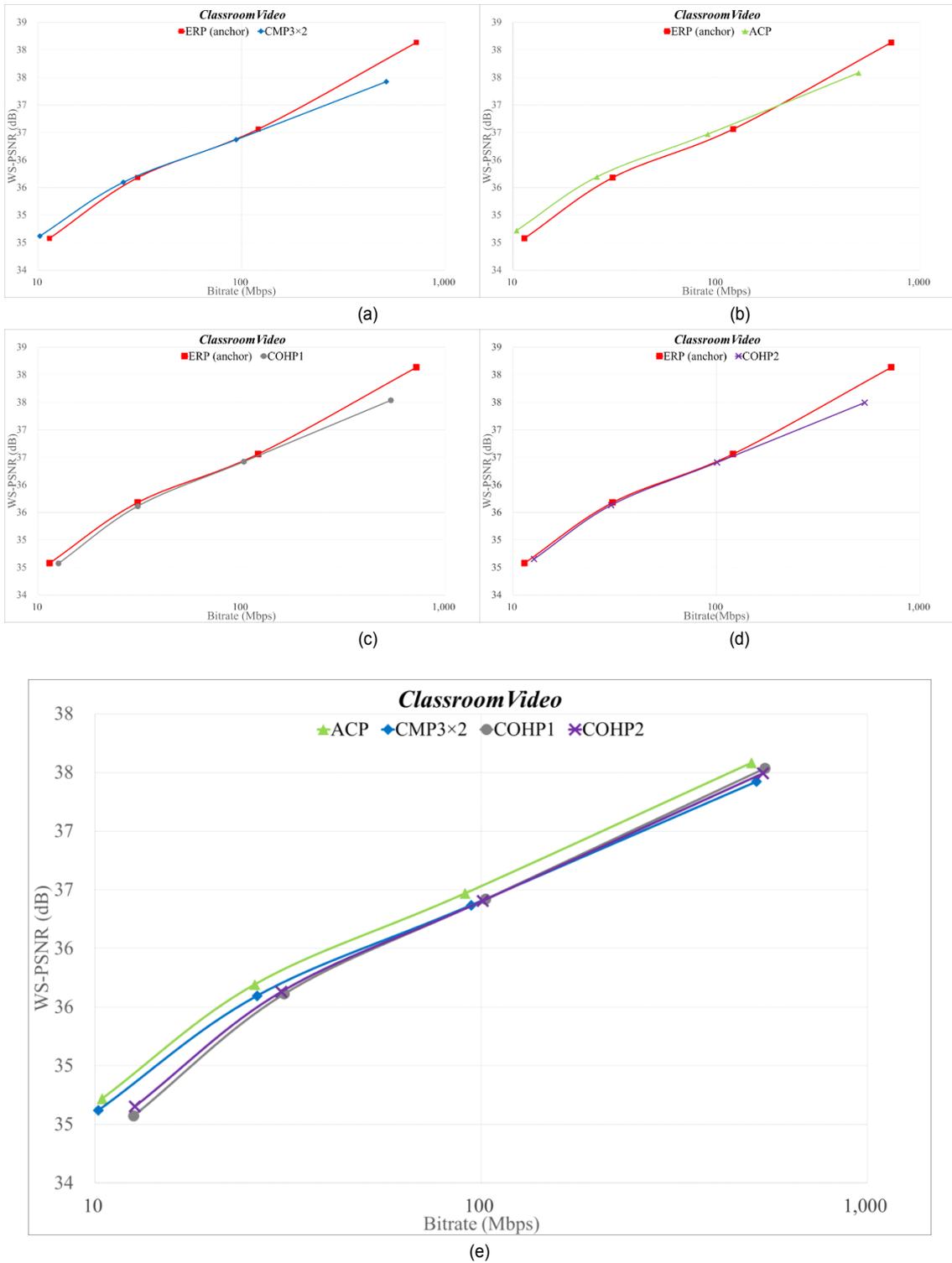


Fig. 11. RD-curve for texture between bitrate and WS-PSNR<sub>Y</sub> for each projection type. (a) CMP3x2, (b) ACP, (c) COHP1, (d) COHP2, and (e) summary

result than the traditional cubemap. The result of CMP3×2 was better than that of the anchor, which was encoded with the ERP format, when the QP value was high. However, COHP1 and COHP2 showed lower WS-PSNR\_Y values for all QPs compared with the anchor. Each side of the COHP1 and COHP2 is not represented as vertical or horizontal lines, which lowers the encoding efficiency. Therefore, the converted video for the aforementioned projections showed less promising results compared with the anchor.

## V. Conclusion

Projection conversion based a bitrate-saving method was proposed. In detail, the original ERP video is converted by the *360lib* converter to reduce the size of resolution, and an HEVC encoder performs encoding for transmission. From the client side, the bitstreams are decoded using an HEVC decoder, and decoded videos are reconstructed by the *360lib* converter. Among the introduced projections, ACP showed an 11.4% BD-rate gain for luma value and a 0.05% pixel rate saving. However, the introduced method causes pixel loss, which leads to the loss of WS-PSNR. Consequently, research about an efficient projection type that represents the original video with less pixel loss is needed, and intensive experiments must be performed to deduce an equation that defines the relation between the original and converted videos.

## References

- [1] Bart Kroon, Gauthier Lafruit. 2018. Reference view synthesizer (rvs) 2.0 manual. 123th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG123/n17759.
- [2] Jong-Beom Jeong, Dongmin Jang, Jangwoo Son, and Eun-Seok Ryu. 2018. 3DoF+ 360 Video Location-Based Asymmetric Down-Sampling for View Synthesis to Immersive VR Video Streaming. *Sensors* 18, 9(2018), 3148.
- [3] Joel Jung, Bart Kroon, Renaud Doré, Gauthier Lafruit, Jill Boyce. 2018. CTC on 3DoF+ and Windowed 6DoF (v2). 123th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG123/n17726.
- [4] Bart Kroon. 2018. 3DoF+ test sequence ClassroomVideo. 122th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2018/m42415.
- [5] Tuan Thanh Le, Dien Ngyuen Van, and Eun-Seok Ryu. 2018. Real-time 360-degree video streaming over millimeter wave communication. 2018 International Conference on Information Networking (ICOIN), IEEE. pp. 857-862.
- [6] Masayuki Tanimoto, Toshiaki Fujii, Kazuyoshi Suzuki. 2009. View Synthesis Algorithm in View Synthesis Reference Software 2.0 (VRS2.0). 87th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2009/m16090.
- [7] Masayuki Tanimoto, Toshiaki Fujii. 2010. FTV - Free Viewpoint Television. 61th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2010/m8595.
- [8] Sejin Oh. 2017. Projections under Considerations for ISO/IEC 23090-2 Omnidirectional Media Format. 118th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG2017/n16828.
- [9] Hyun-Joon Roh, SungWon Han, and Eun-Seok Ryu. 2017. Prediction complexitybased HEVC parallel processing for asymmetric multicores. *Multimedia Tools and Applications* 76, 23 (2017), 25271 - 25284.
- [10] Robert Skupin, Yago Sanchez, Karsten Schäjhing, Thomas Schierl, Eun-Seok Ryu, and Jangwoo Son. 2018. Temporal MCTS Coding Constraints Implementation. 122th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 122/m42423.
- [11] Jang-Woo Son, Dongmin Jang, Eun-Seok Ryu. 2018. Implementing 360 video tiled streaming system. *Proceedings of the 9th ACM Multimedia Systems Conference*. ACM. pp. 521-524.
- [12] Jang-Woo Son, Dongmin Jang, Eun-Seok Ryu. 2018. Implementing Motion-Constrained Tile and Viewport Extraction for VR Streaming. *ACM Network and Operating System Support for Digital Audio and Video 2018 (NOSSDAV2018)*. pp. 61-66.
- [13] Jang-Woo Son, Eun-Seok Ryu. 2018. Tile-Based 360-Degree Video Streaming for Mobile Virtual Reality in Cyber Physical System. Elsevier, *Computers and Electrical Engineering*.
- [14] Yule Sun, Lu Ang, Lu Yu. 2017. Weighted-to-spherically-uniform quality evaluation for omnidirectional video. *IEEE signal processing letters* 24.9. pp. 1408-1412.
- [15] Takanori Senoh, Nobuji Tetsutani, Hiroshi Yasuda. 2018. MPEG-I-Visual: View Synthesis Reference Software (VRS2x). 123th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 123/m42911.
- [16] Xin Wang, Lulin Chen, Shuai Zhao and Shawmin Lei. 2017. From OMAF for 3DoF VR to MPEG-I Media Format for 3DoF+, Windowed 6DoF and 6DoF VR. 119th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 119/m41197.
- [17] Yule Sun, Bin Wang, Lu Yu. 2018. ERP WS-PSNR Software Manual. 123th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 123/n17760.

---

저 자 소 개

---



Jong-Beom Jeong

- 2018. 8. : Received B.S. degree in Department of Computer Engineering from Gachon University
- 2018. 9. ~ 2019. 8. : Pursued M.S. degree in Department of Computer Engineering from Gachon University
- 2019. 9. ~ Current : Pursuing M.S. degree in Department of Computer Education from Sungkyunkwan University (SKKU)
- ORCID : <https://orcid.org/0000-0002-7356-5753>
- Research of Interest : Multimedia communication and system, video compression standard



Soonbin Lee

- 2014. 3. ~ Current : Pursuing B.S. degree in Department of Computer Engineering from Gachon University
- ORCID : <https://orcid.org/0000-0002-8951-0335>
- Research of Interest : Multimedia communication and system, video compression standard



Dongmin Jang

- 2019. 2. : Received B.S. degree in Department of Computer Engineering from Gachon University
- 2019. 3. ~ 2019. 8. : Pursued M.S. degree in Department of Computer Engineering from Gachon University
- 2019. 9. ~ Current : Pursuing M.S. degree in Department of Computer Education from Sungkyunkwan University (SKKU)
- ORCID: <https://orcid.org/0000-0001-7920-8637>
- Research of Interest : Multimedia communication and system, video compression standard



Sungbin Kim

- 2013. 3. ~ Current : Pursuing B.S. degree in Department of Computer Engineering from Gachon University
- ORCID : <https://orcid.org/0000-0003-0956-7583>
- Research of Interest : Multimedia communication and system, video compression standard



Sangsoon Lee

- 1982. 2. : Received B.S. degree in Department of Electronic Engineering from Inha University
- 1986. 2. : Received M.S. degree in Department of Computer Engineering from Inha University
- 2005. 2. : Received Ph.D. degree in Department of Computer Engineering from Incheon University
- 1994. 2. ~ Current : Associate professor in Department of Computer Engineering from Gachon University
- ORCID : <https://orcid.org/0000-0001-6680-2637>
- Research of Interest : Computer network, system software, IoT

---

저 자 소 개



Eun-Seok Ryu

- 1999. 8. : Received B.S. degree in Department of Computer Science from Korea University
- 2001. 8. : Received M.S. degree in Department of Computer Science from Korea University
- 2008. 2. : Received Ph.D. degree in Department of Computer Science from Korea University
- 2008. 3. ~ 2008. 8. : Research professor from Korea University
- 2008. 9. ~ 2010. 12. : Postdoctoral Research Fellow in the School of Electrical and Computer Engineering from Georgia Centers for Advanced Telecommunications Technology (GCATT)
- 2011. 1. ~ 2014. 2. : Staff engineer from InterDigital Labs
- 2014. 3. ~ 2015. 2. : Principal Engineer from Samsung Electronics
- 2015. 3. ~ 2019. 8. : Associate professor in Department of Computer Engineering from Gachon University
- 2019. 9. ~ Current : Associate professor in Department of Computer Education from Sungkyunkwan University (SKKU)
- ORCID: <https://orcid.org/0000-0003-4894-6105>
- Research of Interest : Multimedia communication and system, video compression and international standard, application field of HMD/VR