Special Paper
방송공학회논문지 제25권 제7호, 2020년 12월 (JBE Vol. 25, No. 7, December 2020)
https://doi.org/10.5909/JBE.2020.25.7.1073
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

Implementing VVC Tile Extractor for 360-degree Video Streaming Using Motion-Constrained Tile Set

Jong-Beom Jeong[a)], Soonbin Lee[a)], Inae Kim[a)], Sangsoon Lee[b)], and Eun-Seok Ryu[a)‡]

Abstract

360-degree video streaming technologies have been widely developed to provide immersive virtual reality (VR) experiences. However, high computational power and bandwidth are required to transmit and render high-quality 360-degree video through a head-mounted display (HMD). One way to overcome this problem is by transmitting high-quality viewport areas. This paper therefore proposes a motion-constrained tile set (MCTS)-based tile extractor for versatile video coding (VVC). The proposed extractor extracts high-quality viewport tiles, which are simulcasted with low-quality whole video to respond to unexpected movements by the user. The experimental results demonstrate a savings of 24.81% in the bjøntegaard delta rate (BD-rate) saving for the luma peak signal-to-noise ratio (PSNR) compared to the rate obtained using a VVC anchor without tiled streaming.

Keyword : VVC, Tiled Streaming, 360 video, MCTS, Extractor

Ⅰ. Introduction

Virtual reality (VR) technology has become widely popular, and many head-mounted displays (HMDs) and VR-ready mobile devices are now available on the market. To provide an immersive experience via a HMD, high-quality video is required with resolution approaching 12K and 90 fps[1], which is challenging to stream. The moving picture experts group (MPEG) and the ITU-T video coding experts group (VCEG) thus established a joint video experts team (JVET) to develop a next-generation video codec. JVET subsequently defined a common test condition (CTC) and evaluation method for 360-degree video streaming[2]. Furthermore, several 360-degree video streaming technologies have been proposed. For example, user location-based adaptive down-sampling[3] and redundancy removal[4] for three degrees of freedom plus (3DoF+) videos were proposed. Further, because only a portion of the 360-degree video is displayed to the HMD, viewport-dependent tile-based streaming[5-8] was proposed to save bandwidth.

To implement tiled streaming, a motion-constrained tile set (MCTS) and bitstream-level tile extractor with high-efficiency video coding (HEVC) have been proposed to guar-

a) Department of Computer Education, Sungkyunkwan University
b) Department of Computer Engineering, Gachon University
‡ Corresponding Author : Eun-Seok Ryu
E-mail: esryu@skku.edu
Tel: +82-2-760-0677
ORCID: https://orcid.org/0000-0003-4894-6105

※ This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2019 R1A2C1010476).
· Manuscript received October 29, 2020; Revised December 2, 2020; Accepted December 2, 2020.

antee independent extraction and decoding of each tile in a picture. This paper proposes a versatile video coding (VVC)-compliant tile extractor-based viewport-dependent streaming system. Figure 1 provides an overview of the proposed tiled streaming system. The server encodes the input video with two layers: the tile layer contains a high-quality MCTS bitstream, and the base layer covers the entire video and contains a low-quality non-MCTS bitstream. At the client side, the viewport tile selector chooses the viewport tiles based on the viewport orientation and then transmits the tile indices to the server. At the server side, the proposed tile extractor extracts the viewport tiles from the tile layer and transmits them with the base layer to the client. Finally, the client decodes the two layers and generates the viewport video.

The remainder of this paper is divided into the following sections. Section 2 introduces the background. Section 3 describes the proposed tile extractor scheme. Section 4 introduces the experimental settings and the analyses of the results. Finally, Section 5 presents the conclusion.

II. Background

This section introduces the background of the proposed method. The HMD renders only a portion of the entire 360-degree video, and the area that is rendered changes depending on the user's movements. Once the area to be rendered is determined, transmitting only that area can save bandwidth, which is not a new concept as a streaming strategy. However, encoding a subset of the 360-degree video at the raw video level places a burden on the server. To extract a rectangular area from the bitstream, tile can be used in HEVC[9]. In HEVC, a picture is composed of one or more slices, each of which may contain one or more coding tree units (CTUs). Figure 2(a) shows a picture composed of four slices, where slice number 0 is represented as slice #0. Even though slice #0 is damaged in this figure, the other slices can still be decoded because there is no correlation between the slices. Meanwhile, as shown in Figure 2(b), a slice may contain a tile that forms a rectangular area of the video. Because the tiles can be used both to implement parallel decoding and for individual extraction and streaming, the correlation between them should be removed. MCTS limits the temporal prediction range within the edge of each tile, enabling the extraction and decoding of the tiles. [9] showed that viewport-dependent tiled streaming can save 35.42% of the bjøntegaard delta rate (BD-rate) as compared to non-tiled streaming.

In VVC, slice and tile are also available, and rectangular

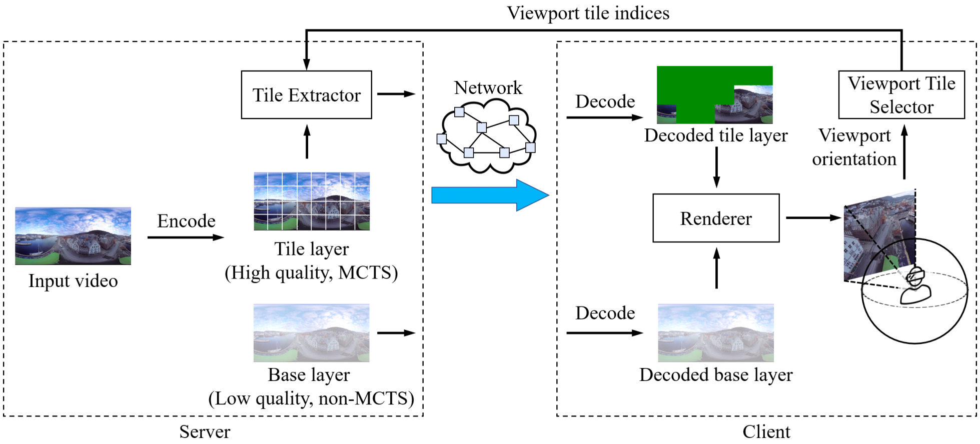


Fig. 1. Overview of viewport-dependent VVC-compliant 360-degree video tiled streaming

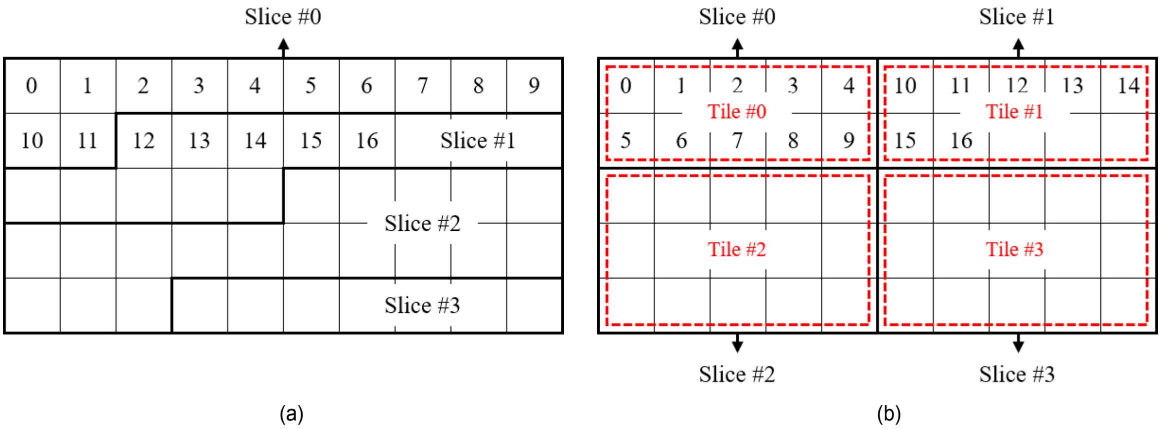


Fig. 2. Illustration of slices and tiles: (a) four slices in raster-scan order, (b) four slices contained within each tile in tile-scan order

slice has been proposed[10]. That is, a picture may contain one or more rectangular slices representing rectangular areas. Tile is included in VVC, and it can be used with MCTS. Further, to provide a more versatile partitioning structure of a picture, the use of sub-picture has been proposed[11]. Each sub-picture in a picture can be decoded independently, and thus flexible picture partitioning is available in VVC. However, sub-picture or tile extraction software is not included in VVC test model (VTM) 7.3, and therefore, an extraction software that uses the high-level syntax is needed.

Ⅲ. VVC Tile Extractor using Motion-Constrained Tile Set

This section describes the architecture of the proposed tile extractor for VVC. Figure 3 shows a functional flow chart of the tile extractor, which is based on VTM 7.3. Before tile extraction, the input bitstream should be encoded while the EnablePicPartitioning and MCTSEnc Constraint options are set to 1. Further, the size of each tile should be declared, and the in-loop filter across slices should be deactivated. Each slice contains one tile, and the proposed extractor obtains the target tile index. Because the VVC bitstream is composed of several network abstraction layer (NAL) units, the proposed extractor parses each NAL unit and stores the information that is required when extracting a tile. In HEVC, the MCTS bitstream includes an extraction information sets (EIS) supplemental enhancement information (SEI) message that contains the output parameter sets for each tile. The HEVC tile extractor parses the output video parameter set (VPS), sequence parameter set (SPS), and picture parameter set (PPS) from the EIS SEI message, and then stores them as NAL units. Meanwhile, when using the SEI message in VTM 7.3, a flag HEVC_SEI should be activated, which is disabled by default. Therefore, the proposed tile extractor parses the VPS, SPS, PPS, adaptation parameter set (APS), and picture header (PH) from the input bitstream and modifies them using the CTU address of the target tile. The APS and PH are newly added in VVC. The proposed method can reduce the bitrate because it does not require the SEI message for the output parameter sets.

After obtaining the parameter sets, the proposed extractor receives their information. For example, the PPS contains the number of tiles, the size of the picture and CTU, and arrays of the tile sizes. After parsing the original PPS, the proposed extractor finds a slice that contains the target tile using the CTU address. Some parameters in the parsed original parameter sets then need to be modified. For example, the proposed extractor modifies the picture

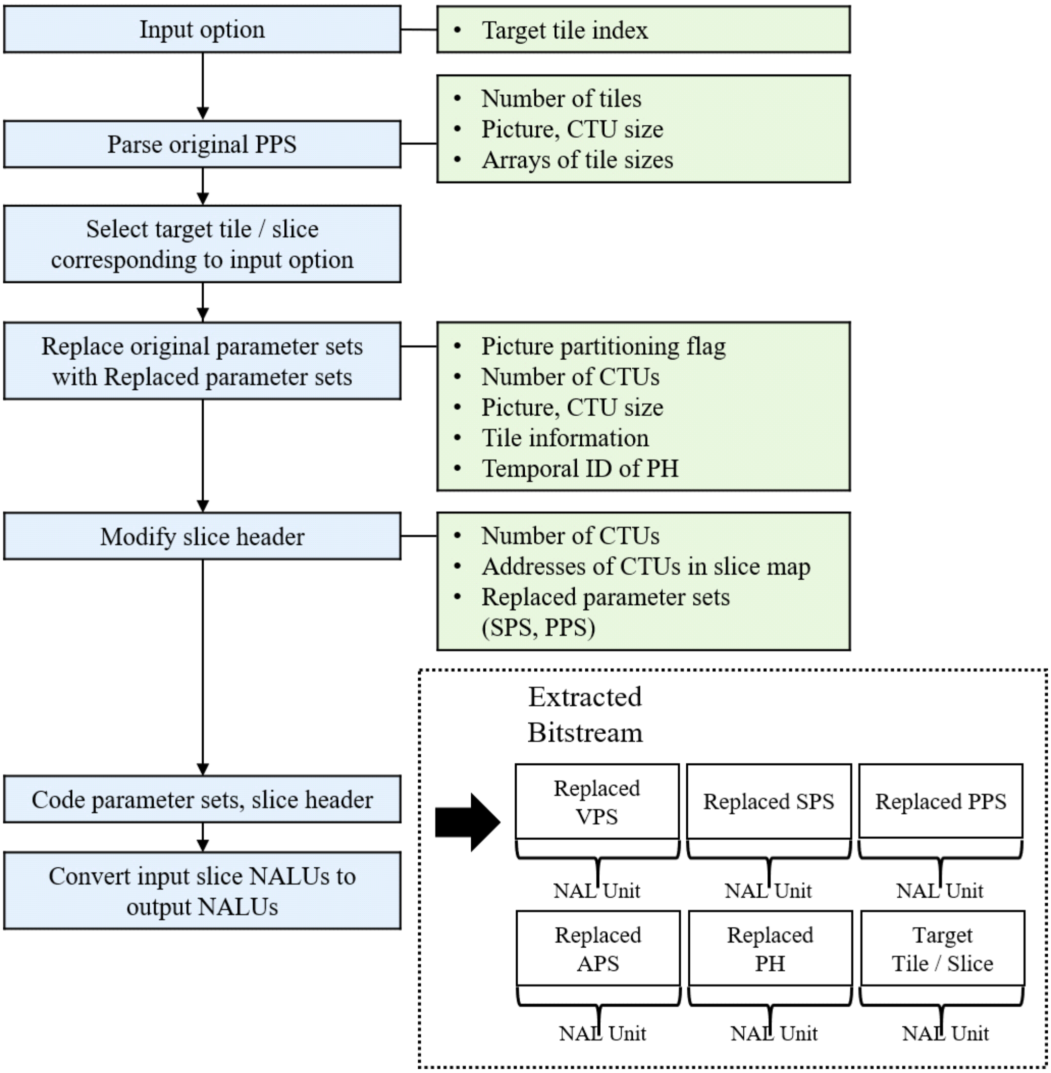


Fig. 3. Functional flow chart of the proposed VVC tile extractor

partitioning flag, the number of CTUs, the size of picture and CTU, the tile information, and the temporal ID of the PH. In VVC, when parameter sets such as the SPS and PPS are changed, they should also be applied to the slice header. Therefore, the proposed extractor replaces the parameter set of the slice header with the modified parameter sets. The number of CTUs and the addresses of the CTUs in the slice map are also modified. Then, the proposed extractor codes the parameter sets and slice header and converts the input slice NAL unit to the output NAL unit. Finally, the output bitstream, which is compatible with the VVC decoder, is generated.

Ⅳ. Experimental Results

This section introduces the experimental environments and the analyses of the results. One server was used for this experiment; it had two Intel Xeon E5-2687w v4 CPUs and 128GB of memory, and Ubuntu 18.04 was installed on it. The experiment was conducted following the JVET CTC for 360-degree video[2]. The CTC defines the test sequences, softwares, and parameters for a fair evaluation. Table 1 describes the properties of the test sequences used here: AerialCity, DrivingInCity, and DrivingInCountry. These sequences have 4K resolution and represent an omnidirectional view. For the projection format, an equirectangular projection (ERP) was used. Table 2 lists the experimental settings based on the CTC: tie first four quantization parameter (QP) values were used to encode the high-quality tile layer, and the last QP value was used to encode the low-quality base layer. The frame rate was set to 30 fps, and random access was applied as the encoding configuration. The FoV for the viewport generation was 90°×90°. The test sequences were divided into 3×6 grid tiles because [9] conducted tiled streaming experiment on 18-, 12-, and 8-grid tiling and the first partitioning scheme, 3×6, showed the highest BD-rate saving. Also, different viewport orientations were applied for each test sequence. Therefore, in this experiment, six, nine, and six tiles were extracted for AerialCity, DrivingInCity, and DrivingIn Country, respectively.

To encode the test sequences, VTM 7.3 was used in this experiment. The HEVC test model (HM) 16.20 was used as a baseline. Both non-tiled streaming and tiled streaming were conducted on HM and VTM. In non-tiled streaming, a non-MCTS high-quality bitstream is generated and transmitted to the client side. In tiled streaming, a base layer which containing a non-MCTS low-quality bitstream and a tile layer that contains a MCTS high-quality viewport tile bitstream are generated. By simulcasting both the tile and base layers, the proposed method can compensate when the user moves unexpectedly: the low-quality base layer is briefly displayed until the high-quality tile layer can be transmitted and rendered. For the quality evaluation, the peak signal-to-noise ratio (PSNR), video multimethod assessment fusion (VMAF)[12], multi-scale structural similarity (MS-SSIM)[13], and immersive video PSNR (IV-PSNR)[14] were used.

Table 2. Experiment settings

Items	Experimental values
QPs	22, 27, 32, 37, 42
Framerate	30 fps
Encoding configutration	Random access
FoV	90° × 90°
Tiling	3 × 6

Table 3 presents the Y-PSNR BD-rate savings of the HEVC tiled streaming, the VVC anchor, and the proposed extractor-based streaming method compared to the HEVC anchor. The HEVC tiled streaming produced a BD-rate gain of 30.20%, while the VVC anchor provided a gain of 32.34%. The proposed method yielded the highest gain at 48.00%. Figure 4 shows the rate distortion (RD) curves of the test sequences. As evident in the Figure, the proposed method outperforms the comparison methods. For the HEVC tiled streaming and the VVC anchor, the performance varied: at low bandwidth, the VVC anchor had better results than the HEVC tiled streaming, whereas at high bandwidth, the HEVC tiled streaming outperformed the VVC anchor.

Table 1. Properties of the test sequences

Sequence name	Resolution	Format	Frame count	Bit depth
AerialCity	3840 × 1920	Omnidirectional ERP	300	8
DrivingInCity	3840 × 1920	Omnidirectional ERP	300	8
DrivingInCountry	3840 × 1920	Omnidirectional ERP	300	8

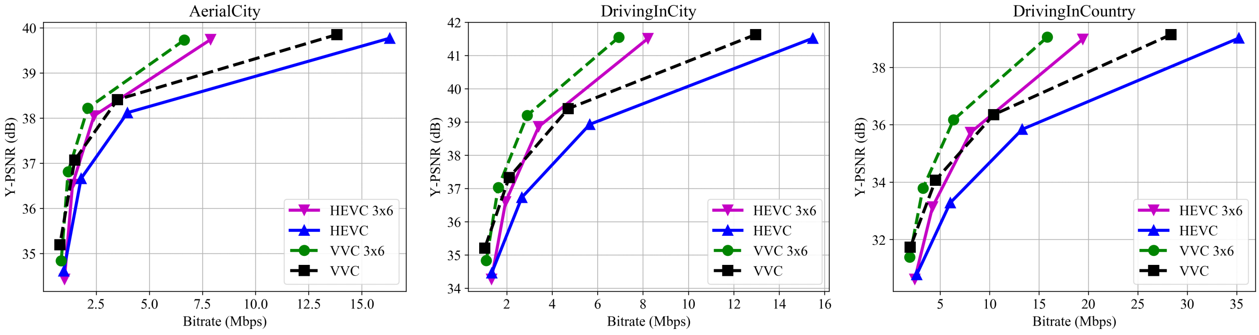
Table 3. Y-PSNR BD-rate savings of the HEVC tiled streaming, VVC anchor, and VVC tiled streaming compared to HEVC anchor

Sequence name	HEVC	HEVC 3×6 tiling	VVC	VVC 3×6 tiling
AerialCity	0.00%	-27.03%	-28.77%	-42.21%
DrivingInCity	0.00%	-31.37%	-31.94%	-47.90%
DrivingInCountry	0.00%	-32.20%	-36.31%	-53.90%
Average	0.00%	-30.20%	-32.34%	-48.00%

Table 4. Y-PSNR, VMAF, MS-SSIM, and IV-PSNR BD-rate savings of the proposed method compared to VVC anchor

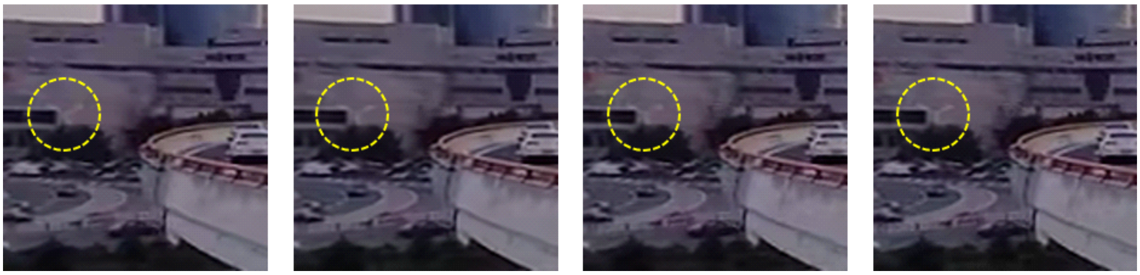
Sequence name	Y-PSNR	VMAF	MS-SSIM	IV-PSNR
AerialCity	-21.01%	-3.07%	-15.61%	-32.04%
DrivingInCity	-24.98%	-16.70%	-18.13%	-31.18%
DrivingInCountry	-28.45%	-18.79%	-21.88%	-29.37%
Average	-24.81%	-12.85%	-18.54%	-30.86%

Table 4 lists the BD-rate savings of Y-PSNR, VMAF, MS-SSIM, and IV-PSNR for the proposed method as compared to the VVC anchor. The proposed method produced average BD-rate savings of 24.81%, 12.85%, 18.54%, and 30.86% in terms of Y-PSNR, VMAF, MS-SSIM, and IV-PSNR, respectively. Because IV-PSNR was designed to evaluate immersive video, this result indicates that the proposed method is advantageous for human quality assessment. Figure 5 shows the enlarged noticeable sections of the generated viewport in HEVC anchor, VVC anchor,



(a) (b) (c)

Fig. 4. RD-Curves, HEVC anchor vs HEVC 3×6 tiling, VVC anchor, and VVC 3×6 tiling on (a) AerialCity, (b) DrivingInCity, (c) DrivingInCountry



(a) (b) (c) (d)

Fig. 5. Generated viewport comparison with enlarged noticeable sections in DrivingInCity: (a) HEVC anchor, 3.103Mbps@37.21dB, (b) VVC anchor, 2.856Mbps@38.16dB, (c) HEVC 3×6 tiling, 2.990Mbps@38.39dB, (d) VVC 3×6 tiling, 2.903Mbps@39.19dB

HEVC tiled streaming, and VVC tiled streaming, each of which used QP values 31, 30, 28, 27 to match the bitrate. As shown in the figure, the HEVC anchor provides several noticeable visual artifacts, e.g., a street lamp in a yellow dotted circle. However, these artifacts are not shown in the VVC tiled streaming. Thus, the advantage of the proposed extractor-based streaming method is verified.

Ⅴ. Conclusion

This paper proposes a VVC-compliant tile extractor and VVC tiled streaming system. When the client determines the viewport tiles and transmits them to the server, the server extracts high-quality viewport tiles and transmits them with the low-quality 360-degree video to the client simultaneously. Subsequently, the client decodes the bitstreams and generates the viewport. The proposed method showed a 24.81% Y-PSNR BD-rate gain compared to the VVC anchor. In the future, a bitstream extractor and merger (BEAMer) will be needed to implement a real-time tiled streaming system.

References

[1] M. -L. Champel, T. Stockhammer, T. Fautier, E. Thomas, R. Koenen. “Quality Requirements for VR”, 116th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG 116/m39532, 2016.

[2] J. Boyce, E. Alshina, A. Abbas, Y. Ye, “JVET common test conditions and evaluation procedures for 360° video”, 118th MPEG meeting of ISO/IECJTC1/SC29/WG11, MPEG118/ n16891, 2017.

[3] J. B. Jeong, D. Jang, J. Son, E. -S. Ryu, “3DoF+ 360 Video Location-Based Asymmetric Down-Sampling for View Synthesis to Immersive VR Video Streaming”, Sensors 18(9), 3148, 2018.

[4] J. -B. Jeong, S. Lee, D. Jang, E. -S. Ryu, “Towards 3DoF+ 360 Video Streaming System for Immersive Media”, IEEE Access, 7, pp. 136399-136408, 2019.

[5] J. Son, D. Jang, E. -S. Ryu, “Implementing 360 video tiled streaming system”, In Proceedings of the 9th ACM Multimedia Systems Conference, pp. 521-524, 2018.

[6] J. Son, E. -S. Ryu, “Tile-based 360-degree video streaming for mobile virtual reality in cyber physical system”, Computers & Electrical Engineering, 72, 361-368, 2018.

[7] S. Lee, D. Jang, J. -B. Jeong, E. -S. Ryu, “Motion-constrained tile set based 360-degree video streaming using saliency map prediction”, In Proceedings of the 29th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video, pp. 20-24, 2019.

[8] J. -B. Jeong, S. Lee, I. -W. Ryu, T. T. Le, E. -S. Ryu, “Towards Viewport-dependent 6DoF 360 Video Tiled Streaming for Virtual Reality Systems”, In Proceedings of the 28th ACM International Conference on Multimedia, pp. 3687-3695, 2020.

[9] A. Zare, A. Aminlou, M. Hannuksela, M. Gabbouj, “HEVC-compliant tile-based streaming of panoramic video for virtual reality applications”, In Proceedings of the 24th ACM international conference on Multimedia, pp. 601-605, 2016.

[10] M. Coban, V.Seregin, M. Karczewicz, “AHG12: On rectangular slices”, 15th JVET meeting, JVET-O0199, 2019.

[11] L. Chen, C. -Y. Chen, Y. -W. Huang, S. -M. Lei, “AHG17: [SYS-VVC] 14th JVET meeting of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-B1001, 2019.

[12] C. Bampis, A. Bovik, Z. Li, “A Simple Prediction Fusion Improves Data-driven Full-Reference Video Quality Assessment Models”, In 2018 Picture Coding Symposium (PCS), pp. 298-302, 2018.

[13] Z. Wang, E. Simoncelli, A. Bovik, “Multiscale structural similarity for image quality assessment”, In The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Vol. 2, pp. 1398-1402, 2003.

[14] A. Dziembowski, “Software manual of IV-PSNR for Immersive Video”, 128th MPEG meeting of ISO/IEC JTC1/SC29/ WG11, MPEG127/n18709, 2019.

Introduction Authors

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 Ph.D. 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

- 2020. 2. : Received B.S. degree in Department of Computer Engineering from Gachon University
- 2020. 3. ~ Current : Pursuing M.S. degree in Department of Computer Education from Sungkyunkwan University (SKKU)
- ORCID : https://orcid.org/0000-0002-8951-0335
- Research of Interest : Multimedia communication and system, video compression standard

Inae Kim

- 2013. 8. : Received B.S. degree in Department of Nutrition and Foodservice Management from Pai Chai University
- 2019. 9. ~ Current : Pursuing M.S. degree in Department of Computer Education from Sungkyunkwan University (SKKU)
- ORCID : https://orcid.org/0000-0003-4263-6448
- 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

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. : Assistant professor in Department of Computer Engineering from Gachon University
- 2019. 9. ~ Current : Assistant 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