Delta QP Allocation for MPEG Immersive Video

Jong-Beom Jeong, Soonbin Lee, Eun-Seok Ryu
Department of Computer Science Education, Sungkyunkwan University (SKKU), Seoul, Republic of Korea
E-mail: {uof4949, soonbinlee, esryu}@skku.edu

***Abstract*—One way to experience six degrees of freedom (6DoF) immersive video in virtual reality (VR) based on natural content is by using a moving picture experts group (MPEG) immersive video (MIV). To accomplish outstanding bjøntegaard delta (BD)-rate performance, the MIV coding standard removes the inter-view correlation, detects the residuals, and outputs them into videos called atlases. This paper presents a delta quantization parameter (QP) allocation method that considers the characteristics of each atlas to achieve higher BD-rate gains. The experimental results showed that the proposed delta QP allocation method achieved higher BD-rate gains than the existing experimental configurations of the MIV.**

***Index Terms*—Metaverse, Virtual reality, Six degrees of freedom, MPEG immersive video, Versatile video coding**

I. Introduction

In recent times, the extensive increase in metaverse users owing to the COVID-19 pandemic has emphasized the importance of efficient immersive video streaming methods. By deploying head-mounted displays (HMDs) such as Oculus Quest series from Meta and Vive series from HTC, more realistic and immersive content became available. However, because HMDs display only a portion of the entire video, the technical requirements for 360-degree immersive video playback are challenging [1], [2]. [3] defines the technical requirements for virtual reality (VR) video streaming, claiming that immersive videos that provide 90 frames per second (FPS) and contain larger than 12K resolution of pictures are required. Therefore, selective and asymmetric streaming approaches for high-quality immersive video streaming have been proposed, using asymmetric down-sampling [4] and viewport-dependent selective streaming based on the motion-constrained tile set (MCTS) [5]–[8].

One way to provide motion parallax for a better quality of experience (QoE) is to provide multiple videos that simultaneously capture a space. By providing multiple videos, HMD users can move in the VR space, and the HMD must provide the corresponding videos, even though the location that the user is watching was captured in a different position. The moving picture experts group (MPEG) is responsible for the standardization of immersive video compression; therefore, the MPEG established the MPEG immersive video (MIV) coding standard [10] to improve the coding efficiency of immersive videos, each of which consists of texture (color) and geometry (depth) videos, are input to the MIV encoder. The MIV encoder removes the inter-view correlation, aggregates the residuals, and extracts rectangular patches, and packs them into output videos; these are defined as atlases. When encoding atlases, asymmetric bit allocation is worth to discussing because the characteristics of the texture and geometry are different. [11] reported that assigning delta quantization parameter (QP) to the geometry resulted in 31.6% bjøntegaard delta (BD)-rate gains. Therefore, the current MIV applies different QPs to texture and geometry atlases. In addition, further discussion on delta QP is required, considering patch arrangements within atlases.

This work was supported by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2020-0-00231-003, Development of Low Latency VR·AR Streaming Technology based on 5G edge cloud).

(a) (b)

Fig. 1: Examples of texture atlases generated by the MIV encoder: sequence *Fan* [9], (a) atlas index 0 that mainly contains basic views, (b) atlas index 1 that contains patches from additional views.

This paper presents a delta QP allocation method for MIV. Fig. 1 shows two atlases generated by the MIV encoder using test sequence *Fan* proposed by InterDigital [9]. Fig. 1a shows multiple immersive videos where the inter-view redundancy is not removed. Fig. 1b shows residuals that have similar

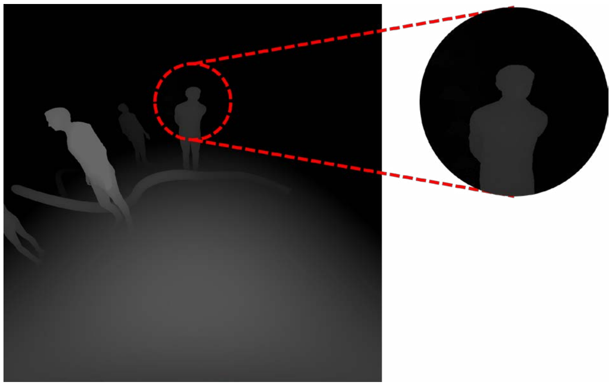


Fig. 2: Example of geometry video containing constant values within an object and sharp edges.

characteristics to those of geometry videos (for example, sharp edges and constant background values). The proposed method allocates delta QP to the atlases shown in Fig. 1b to achieve a higher BD-rate performance.

The remainder of this paper is constructed as follows. Section II presents the background focusing on the MIV, and Section III explains the proposed delta QP allocation method. Section III-A describes the experimental conditions, and section III-B explains the experimental results. Finally, Section IV gives the conclusion and insights toward rate-distortion optimization (RDO) for the MIV.

II. Background

Simulcasting multiple videos using existing video coding standards, such as high-efficiency video coding (HEVC) [12], have been widely discussed [13], [14]. Multi-view HEVC (MV-HEVC), an extension of HEVC, was developed to increase the BD-rate gain for multi-view videos using inter-view coding structures [15]. However, as an extension of HEVC, MV-HEVC has limitations because it is not compatible with the next generation video compression standard, versatile video coding (VVC) [16]. Further, inter-view redundancy compression is conducted under MV-HEVC, leading to difficulties in real-time implementation based on the existing acceleration methods of HEVC. The MPEG issued a call for proposals (CfP) for three degrees of freedom plus (3DoF+) to support a limited 6DoF in 2019 [17], where the proposed methods had to be agnostic to the videos to ensure compatibility with 2-D video compression tools. Five responses on the CfP were submitted, and they proposed pre- and post-processing methods outside video compression tools [18]–[22]. Based on the responses, the test model for immersive video (TMIV) was developed as a reference software for MIV [23].

The MIV encoder distinguishes the input immersive videos into basic and additional views, and the inter-view correlation is removed from the additional views using the view synthesis method [24]; this is defined as *pruning*. Because the basic views are representative of the input immersive videos, pruning is not conducted for the basic views. Then, the MIV encoder distinguishes between valid and invalid areas using masks [25], and the masks are aggregated within the intraperiod range (typically 32 frames) to fix the location of the patches, contributing to the coding efficiency [26]. Based on the aggregated masks, the MIV encoder conducts *packing*, where patches that represent valid areas as rectangles are extracted and packed into the output videos that are called *atlases*. A video encoder such as HEVC or VVC encodes the atlases that are then transmitted to the client device. After video decoding, the MIV decoder performs *unpacking*, *reconstruction* and *view synthesis* to provide intermediate views.

At the beginning of MIV, the same QP was used to encode the texture and geometry altases. However, as [27] proposed a delta QP for the geometry, that proved to be efficient in terms of the BD-rate, the MIV applies lower QPs to the geometry atlases. Fig. 2 shows the enlarged geometry videos, emphasizing two characteristics of the geometry: i) constant values within an object, and ii) discrete values on the edges of the object. Therefore, geometry videos require fewer bits than texture videos using the same QP. Furthermore, because MIV adpots depth-based image rendering (DIBR), preserving the geometry is important; this leads to the use of lowered delta QPs.

III. Delta QP Allocation for MPEG Immersive Video

This section introduces the proposed delta QP allocation method for MIV, experimental conditions, results, and insights for efficient 6DoF immersive video streaming. As introduced in Section II, the current MIV recommends using lower QPs for the geometry than for the texture. Similarly, delta QPs on the additional view atlases shown in Fig. 1b must be discussed because they contain sharp edges and constant background values similar to the geometry. Fig. 3 shows the proposed delta QP allocation method for additional view atlases. The MIV encoder generates two atlases, where atlas #0 contains basic views, and atlas #1 contains additional view patches. Five predefined QPs exist to satisfy the specific bitrate, from QP_1 to QP_5, where QP_1 represents the lowest QP (highest bitrate). To measure the BD-rate, four QPs must be applied to each atlas [28], where QP_1 to QP_4 are used to encode atlas #0, and QP_2 to QP_5 are applied to atlas#1. After QP allocation, the atlases are encoded by the video encoder and transmitted to the client side. This case assigns higher QPs (lower bitrates) to the additional view atlases. This paper also presents experimental results for assigning lower QPs (higher bitrates) to additional view atlases. The remainder of this section is organized as follows. Section III-A describes the experimental conditions, and Section III-B discusses the experimental results.

A. Experimental Conditions

This section introduces the experimental conditions. The common test conditions (CTC) of the MIV define the experimental conditions, including the recommended software, test sequences, QP values, and evaluation procedures [29]. These experiments followed the recent version of the MIV CTC

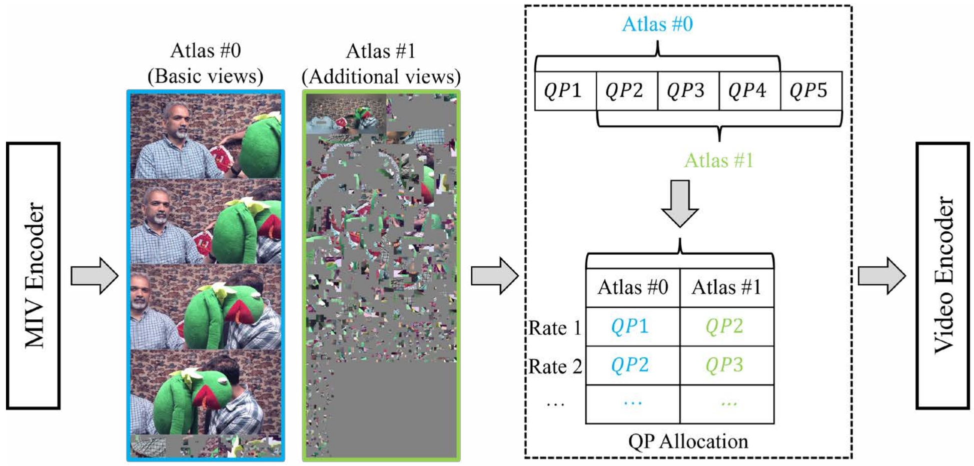


Fig. 3: Proposed delta QP allocation method for additional view atlases.

TABLE I: Softwares used in the experiment.

Software	Version
TMIV	v13.0
VVenC	v0.3.1.0
VVdeC	v1.0.1
IV-PSNR	v4.0

TABLE II: Definitions of the test conditions. Abbreviations: A17, MIV mode for 17 frames; BR, bitrate.

Condition	Basic view atlases	Additional view atlases
A17	QP1 to QP5	QP1 to QP5
DeltaQP_high	QP1 to QP4 (high BR)	QP2 to QP5 (low BR)
DeltaQP_low	QP2 to QP5 (low BR)	QP1 to QP4 (high BR)

released on July 30th, 2022 (document number n00232). Table I lists the softwares used in these experiments based on the MIV CTC. TMIV v13.0 was used to conduct MIV encoding and decoding. For video encoding and decoding, versatile video encoder (VVenC) and versatile video decoder (VVdeC) were used, which are implementations of the VVC, developed by Fraunhofer HHI [30]. Intermediate view synthesis at source view positions was conducted during MIV decoding, and an end-to-end quality assessment was recommended to measure the BD-rate. The immersive video peak signal-to-noise ratio (IV-PSNR), developed to reflect the subjective quality of synthesized videos, was used in these experiments [31].

Table II lists the test conditions used in these experiments. In the table, A17 represents the MIV mode for 17 frames. QP_1 to QP_5 were tested. The MIV mode, which is defined in the MIV CTC, performs both *pruning* and *packing* processes in the MIV encoder. The MIV view mode does not conduct *pruning* but packs the basic views into atlases. According to [32], the MIV mode showed a 13.88% IV-PSNR BD-rate gain compared with the MIV view mode; therefore, the experiments were conducted in the MIV mode. Furthermore, because the TMIV has not been optimized, the MIV CTC recommends conducting experiments for 17 frames, and these experiments

TABLE III: Texture QPs for the experiment. Target bitrate: 50Mbps (QP1), 28Mbps (QP2), 16Mbps (QP3), 9Mbps (QP4), and 5Mbps (QP5).

Name	QP1	QP2	QP3	QP4	QP5
ClassroomVideo	26	28	33	41	51
Museum	29	38	44	48	51
Fan	30	35	41	46	51
Kitchen	17	24	29	35	41
Painter	21	29	38	45	51
Frog	28	32	37	42	46
Carpark	22	26	33	40	47
Chess	17	25	31	37	45
Group	22	28	34	39	46

also followed the recommendation. The proposed delta QP allocation method is two-fold:

- *DeltaQP_high* used QP_1 to QP_4 for basic view atlases and QP_2 to QP_5 for additional view atlases (lower bitrate using higher QPs).
- *DeltaQP_low* used QP_2 to QP_5 for basic view atlases and QP_1 to QP_4 for additional view atlases (higher bitrate using lower QPs).

To conduct experiments on the test conditions listed in Table II, pre-defined QP values were required, which are listed in the MIV CTC. Table III lists the texture QPs from QP_1 to QP_5. The geometry QPs were calculated using Eq. 1:

$$Q_g = max(1, [-14.2 + 0.8 \cdot Q_t]) \tag{1}$$

where Q_g represents the geometry QP, and Q_t denotes the texture QP. Table IV presents the specifications of the test sequences used in the experiments. The test sequences are mandatory datasets that the MIV CTC recommends to conduct experiments on them. Note that the prefix *CG* denotes that the test sequence was computer-generated, and *NC* represents the natural content. The camera array type denotes whether the cameras form a 1-D or 2-D structures or are unstructured.

TABLE IV: Specifications of the MIV test sequences.

Name	Class	Camera array type	Resolution (Source view) (Atlas)	No. of source views
ClassroomVideo	CG-A	Unstructured	4096×2048 4096×2176	15
Museum	CG-B	Unstructured	2048×2048 2048×4352	24
Fan	CG-O	Structured (5×3)	1920×1080 1920×4640	15
Kitchen	CG-J	Structured (5×5)	1920×1080 1920×4640	25
Painter	NC-D	Structured (4×4)	2048×1088 2048×4352	16
Frog	NC-E	Structured (13×1)	1920×1080 1920×4640	13
Carpark	NC-P	Structured (9×1)	1920×1088 1920×4640	9
Chess	CG-N	Unstructured	2048×2048 2048×4352	10
Group	CG-R	Unstructured	1920×1080 1920×4640	21

In these experiments, four sequences had unstructured camera arrays that generally increased the number of additional view patches owing to lower inter-view redundancy.

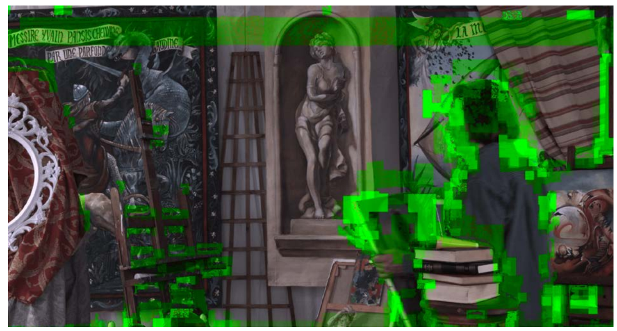
B. Experimental Results

This section explains the experimental results and provides insights into efficient 6DoF immersive video streaming. The three test conditions listed in Table II were tested, and A17 was selected as the anchor. Y-PSNR and IV-PSNR were chosen as the quality assessment methods, and the BD-rate was measured. Table V presents the BD-rate performance of DeltaQP_high on intermediate view synthesis. A high BD-rate represented the BD-rate using QP_1 to QP_4, whereas a low BD-rate was calculated using QP_2 to QP_5, which were defined in the MIV CTC. Because QP_1 to QP_4 were used for the basic view atlases in DeltaQP_high, Table V presents only the high BD-rates. Note that negative BD-rate values represent the bitrate gain of the proposed method. On average, the proposed method showed a 0.11% Y-PSNR BD-rate loss but an IV-PSNR BD-rate gain of 2.03%. Given that the IV-PSNR was designed to measure the quality of the 6DoF immersive video, the IV-PSNR BD-rate gain verified the efficiency of the proposed method. Specifically, the proposed DeltaQP_high method outperformed the anchor in sequences *ClassroomVideo*, *Painter*, and *Carpark* that represent high-quality objects and reflection of light. For the *Museum* and *Kitchen* sequences, the proposed method showed a trivial increase in the bitrate, whereas BD-rate gains were observed in the other sequences, except for sequence *Group*. This was because the number of additional view patches were higher than that of the other sequences, thereby lowering the quality of the additional view atlases, leading to BD-rate loss.

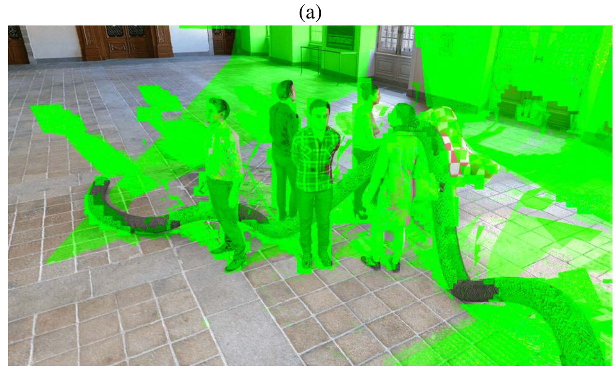
Fig. 4 shows the synthesized intermediate views, where the green areas represent the pixels from additional views. In sequence *Painter*, most of the synthesized intermediate views contained a small portion of the additional view pixels, as shown in Fig. 4a. Fig. 4b shows the synthesized intermediate view *v02* in sequence *Group*, and a higher portion of additional view pixels across most of the views was observed. Thus, it was verified that lowering the QPs to the additional view atlases was efficient when the basic view atlases mainly contributed to the view synthesis; however, in some cases, the proposed method was inefficient when most of the pixels in the synthesized views were from additional views. Given that 1-D or 2-D camera arrays are often used for immersive

TABLE V: BD-rate performance of DeltaQP_high on intermediate view synthesis (negative values indicate that the proposed method has better performance).

Sequence	High Y-PSNR BD-rate	Low Y-PSNR BD-rate	High IV-PSNR BD-rate	Low IV-PSNR BD-rate
ClassroomVideo	-5.32%	-	-9.47%	-
Museum	-0.32%	-	0.11%	-
Fan	-2.11%	-	-1.88%	-
Kitchen	3.89%	-	2.84%	-
Painter	-10.36%	-	-9.61%	-
Frog	-1.57%	-	-1.71%	-
Carpark	-1.65%	-	-4.28%	-
Chess	2.58%	-	-2.31%	-
Group	15.89%	-	8.04%	-
Average	**0.11%**	**-**	**-2.03%**	**-**



(a)



(b)

Fig. 4: Synthesized intermediate views where green areas represent the pixels from additional views, (a) sequence *Painter*, view *v2*, (b) sequence *Group*, view *v02*.

TABLE VI: BD-rate performance of DeltaQP_low on intermediate view synthesis (negative values indicate that the proposed method has better performance).

Sequence	High Y-PSNR BD-rate	Low Y-PSNR BD-rate	High IV-PSNR BD-rate	Low IV-PSNR BD-rate
ClassroomVideo	-	19.24%	-	21.03%
Museum	-	17.23%	-	20.31%
Fan	-	1.37%	-	-0.13%
Kitchen	-	4.87%	-	4.85%
Painter	-	30.89%	-	30.89%
Frog	-	7.11%	-	7.77%
Carpark	-	0.87%	-	2.27%
Chess	-	14.45%	-	19.40%
Group	-	3.76%	-	11.51%
Average	**-**	**11.09%**	**-**	**13.10%**

video acquisition, the proposed method can simply increase the coding efficiency because there are small number of additional view patches in these types of materials containing natural contents. Nevertheless, a delta QP allocation method can be discussed on unstructured camera arrays, which was verified by the *Museum* and *Chess* sequences.

These experiments also tested DeltaQP_low that assigned lower QPs (high quality) to additional view atlases. Table VI presents the BD-rate performance of DeltaQP_low on the Y-PSNR and IV-PSNR. On average, the proposed DeltaQP_low method showed Y-PSNR and IV-PSNR BD-rate losses. Sequence *Group*, which showed the highest BD-rate losses in DeltaQP_high, showed BD-rate losses when lower QPs (higher quality) were assigned to the additional view atlases. These results suggested that a slight modification in delta QP or uniform QP allocation on basic and additional view atlases was efficient for sequences similar to *Group*. As noted from Tables V and VI, assigning lower delta QPs to the additional view atlases was efficient in terms of the BD-rate.

IV. Conclusion

This paper proposed a delta QP allocation method for a 6DoF immersive video and verified its efficiency on the MIV coding standard. Specifically, when multiple immersive videos were input to the MIV encoder, it distinguished basic and additional views, removed the inter-view redundancy from the additional views, and packed them into videos called atlases. Because the additional view atlases had characteristics different from those of typical 2-D videos, the proposed method assigned delta QPs to the additional view atlases. Assigning lower QPs to the additional view atlases showed an IV-PSNR BD-rate gain of 2.03%. In the future, extensive experiments will be conducted to determine the optimal delta QPs for additional view atlases.

References

[1] K. K. Sreedhar, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, “Viewport-adaptive encoding and streaming of 360-degree video for virtual reality applications,” in *2016 IEEE International Symposium on Multimedia (ISM)*. IEEE, 2016, pp. 583–586.

[2] J.-B. Jeong, S. Lee, I. Kim, and E.-S. Ryu, “Performance Analysis of Viewport-dependent Tiled Streaming on 16K Ultra High-quality 360-degree Video,” *Journal of Internet Computing and Services*, vol. 22, no. 3, pp. 1–8, 2021.

[3] M.-L. Champel, T. Stockhammer, T. Fautier, E. Thomas, and R. Koenen, “Quality Requirements for VR.” 116th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG116/m39532., 2016.

[4] J. Jeong, D. Jang, J. Son, and E.-S. Ryu, “3DoF+ 360 video location-based asymmetric down-sampling for view synthesis to immersive VR video streaming,” *Sensors*, vol. 18, no. 9, p. 3148, 2018.

[5] T. Thanh Le, J.-B. Jeong, S. Lee, J. Kim, and E.-S. Ryu, “An Efficient Viewport-Dependent 360 VR System Based on Adaptive Tiled Streaming,” 2021.

[6] J.-B. Jeong, S. Lee, I.-W. Ryu, T. T. Le, and 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*, 2020, pp. 3687–3695.

[7] J.-B. Jeong, S. Lee, I. Kim, S. Lee, and E.-S. Ryu, “Implementing VVC Tile Extractor for 360-degree Video Streaming Using Motion-Constrained Tile Set,” *Journal of Broadcast Engineering*, vol. 25, no. 7, pp. 1073–1080, 2020.

[8] S. Lee, D. Jang, J. Jeong, and 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*. ACM, 2019, pp. 20–24.

[9] R. Doré, G. Briand, and F. Thudor, “[MPEG-I Visual] InterdigitalFan0 content proposal for MIV.” 131th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2020/m54732, 2020.

[10] J. M. Boyce, R. Doré, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital, and L. Yu, “MPEG Immersive Video Coding Standard,” *Proceedings of the IEEE*, 2021.

[11] B. Wang, L. Yu, B. Kroon, and J. Jung, “[MPEG-I Visual] Results on depth QPs in CTC of 3DoF+ Video.” 124th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2018/m44688, 2018.

[12] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[13] J. Jeong, D. Jang, J.-W. Son, and E.-S. Ryu, “Bitrate efficient 3DoF+ 360 video view synthesis for immersive VR video streaming,” in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2018, pp. 581–586.

[14] J. Stankowski, M. Domanski, O. Stankiewicz, J. Konieczny, J. Siast, and K. Wegner, “Extensions of the HEVC technology for efficient multiview video coding,” in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 225–228.

[15] M. M. Hannuksela, Y. Yan, X. Huang, and H. Li, “Overview of the multiview high efficiency video coding (MV-HEVC) standard,” in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 2154–2158.

[16] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, “Overview of the versatile video coding (VVC) standard and its applications,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3736–3764, 2021.

[17] I. JTC1/SC29/WG11, “Call for Proposals on 3DoF+ Visual.” 125th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG/n18145, 2019.

[18] B. Kroon and B. Sonneveldt, “Philips response to 3DoF+ Visual CfP.” 126th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2019/m47179, 2019.

[19] J. Fleureau, F. Thudor, R. Dore, B. Salahieh, M. Dmytrychenko, and J. Boyce, “echnicolor-Intel Response to 3DoF+ CfP.” 126th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2019/m47445, 2019.

[20] M. Domanski, A. Dziembowski, D. Mieloch, O. Stankiewicz, J. Stankowski, A. Grzelka, G. Lee, and J. Seo, “Technical description of proposal for Call for Proposals on 3DoF+ Visual prepared by Poznan University of Technology (PUT) and Electronics and Telecommunications Research Institute (ETRI).” 126th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2019/m47407, 2019.

[21] V. K. M. Vadakital, K. Roimela, L. Ilola, J. Keränen, M. Pesonen, S. Schwarz, J. lainema, and M. Hannuksela, “Description of Nokia’s response to CFP for 3DOF+ visual.” 126th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2019/m47372, 2019.

[22] B. Wang, Y. Sun, and L. Yu, "Description of Zhejiang University's response to 3DoF+ Visual CfP." 126th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2019/m47684, 2019.

[23] B. Salahieh, B. Kroon, J. Jung, and A. Dziembowski, "Test model 14 for immersive video." 139th MPEG meeting of ISO/IEC JTC1/SC29/WG4, MPEG/n00242., 2022.

[24] J.-B. Jeong, S. Lee, D. Jang, I.-W. Ryu, T. T. Le, J. Ryu, and E.-S. Ryu, "Implementing Multi-view 360 Video Compression System for Immersive Media," in *Proceedings of the Korean Society of Broadcast Engineers Conference*. The Korean Institute of Broadcast and Media Engineers, 2019, pp. 140–143.

[25] H.-C. Shin, J.-Y. Jeong, G. Lee, M. U. Kakli, J. Yun, and J. Seo, "Enhanced pruning algorithm for improving visual quality in MPEG immersive video," *ETRI Journal*, vol. 44, no. 1, pp. 73–84, 2022.

[26] J.-B. Jeong, S. Lee, D. Jang, and E.-S. Ryu, "Towards 3DoF+ 360 Video Streaming System for Immersive Media," *IEEE Access*, vol. 7, pp. 136 399–136 408, 2019.

[27] B. Wang, Y. Sun, and L. Yu, "[MPEG-I Visual] On Depth Delta_qp in Common Test Condition of 3DoF+ Video." 123th MPEG meeting of ISO/IEC JTC1/SC29/WG11, MPEG2018/m43801, 2018.

[28] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *VCEG-M33*, 2001.

[29] J. Jung and B. Kroon, "Common Test Conditions for MPEG Immersive Video." Standard ISO/IEC JTC1/SC29/WG4, MPEG/n00232, 2022.

[30] A. Wieckowski, J. Brandenburg, T. Hinz, C. Bartnik, V. George, G. Hege, C. Helmrich, A. Henkel, C. Lehmann, C. Stoffers *et al.*, "VVenC: An open and optimized VVC encoder implementation," in *2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2021, pp. 1–2.

[31] A. Dziembowski, D. Mieloch, J. Stankowski, and A. Grzelka, "IV-PSNR–the objective quality metric for immersive video applications," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

[32] J.-B. Jeong, S. Lee, and E.-S. Ryu, "Rethinking Fatigue-Aware 6DoF Video Streaming: Focusing on MPEG Immersive Video," in *2022 International Conference on Information Networking (ICOIN)*. IEEE, 2022, pp. 304–309.