Rethinking Fatigue-Aware 6DoF Video Streaming: Focusing on 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*—The moving picture experts group (MPEG) immersive video (MIV) coding standard enables the implementation of a practical six degrees of freedom (6DoF) virtual reality (VR) video streaming system by two approaches: 1) removing the redundancy between multi-view videos or 2) selecting and transmitting representative views among multi-view videos using traditional 2-D video compression standards. This paper presents a strategy for providing immersive video considering two aspects: 1) minimizing the fatigue in the synthesized and rendered viewport, and 2) providing rate-distortion optimization (RDO) to yield high-quality videos at a low bitrate. The results of the two approaches in the MIV coding standard using high-efficiency video coding (HEVC) are discussed, and the advantages and drawbacks of each method for 2-D display and head-mounted display are detailed.**

***Index Terms*—Virtual reality, MIV, 6DoF, HEVC, Immersive video, Metaverse**

I. Introduction

Recently, virtual reality (VR) has gained enormous popularity with the release of immersive head-mounted displays (HMDs) such as Samsung Gear VR, Oculus Quest 2, Valve Index, and HTC Vive. For an improved immersion, these HMDs need to provide a larger resolution and extended freedom for the video. For example, Champel et al. [1] stated that a video with a resolution of 12K, 90 frames per second (FPS) for framerate, and 20ms for motion-to-photon (MTP) latency is required to minimize motion sickness by providing an appropriate quality of immersive video to users. Several researches have been proposed to overcome these challenges. For instance, viewport-dependent streaming approaches have been proposed to save bandwidth by dividing a high-resolution 360-degree video into rectangular tiles using the motion-constrained tile set (MCTS), extracting viewport tiles and transmitting them [2]–[6]. To provide six degrees of freedom (6DoF) in VR, which allows users to move without any restrictions, multi-view videos captured simultaneously need to be streamed. Therefore, multi-view efficient compression approaches that exploit the inter-view redundancy have been proposed [7]–[9]. Moreover, the moving picture experts group (MPEG) established the MPEG-Immersive group to build an immersive video coding standard that removes the inter-view redundancy in the pixel domain; this aimed to obtain compatibility with the existing coding standard, for example, high-efficiency video coding (HEVC) [10], as well as the next generation coding standard, for example, versatile video coding (VVC) [11]. As a result, the MPEG immersive video (MIV) coding standard was established [12], and the test model for immersive video (TMIV) as a reference software for MIV has been developed to improve performance. Figure 1 shows the immersive video compression process of the MIV encoder. Each immersive video consists of a texture (color) and geometry (depth) pair to project the 2-D multi-view videos to the 3-D space, and the views are divided into basic and additional views. Most of the scene information can be represented by basic views; therefore, the inter-view redundancy is removed from additional views, whereas the basic views preserve the scene information. Redundancy removal is conducted in the MIV mode, and the residuals are merged into a video defined as *atlas*. Furthermore, the MIV provides the MIV view mode which selects only the basic views and stores them into atlases.

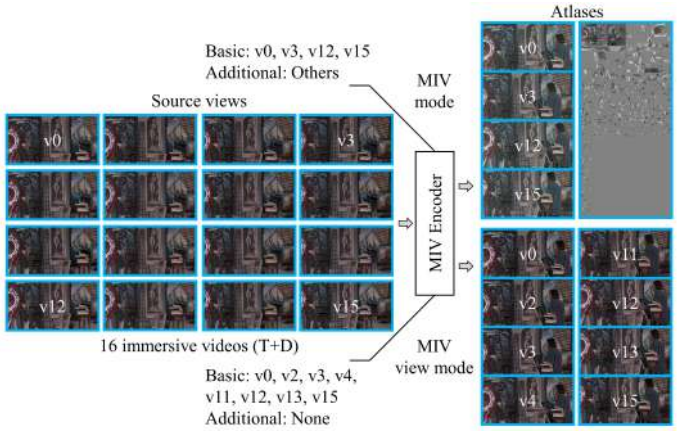

Fig. 1: Generation of the atlases from the source views using the MIV encoder

This paper proposes a fatigue-aware 6DoF video streaming approach for the MIV, considering the advantages and limitations of the two modes for different test sequences and use-cases. The atlases generated by the MIV encoder were encoded and decoded by the HEVC. The MIV decoder uses the decoded atlases to synthesize two types of videos: 1) virtual views that have the same field of view (FoV) with the source views, and 2) pose traces that simulate the user's movement using the HMD. The synthesized videos were evaluated using the full-reference objective quality metrics. Further, subjective quality comparisons were provided considering the compression and synthesis artifacts including holes because of

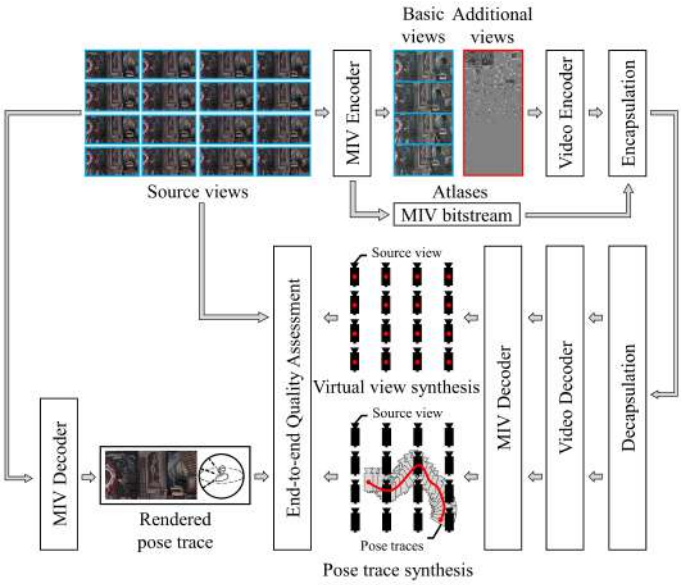

Fig. 2: General architecture of the MIV-compliant 6DoF immersive video streaming system

the occluded areas from the atlases.

The remainder of this paper is organized as follows. Section II explains the background, and Section III describes the experimental conditions, experimental results, and analysis of the results. Finally, Section IV presents the conclusions and provides insights toward high-quality 6DoF video streaming.

II. Background

Multi-view video compression is not a new subject for signal processing. Researches exploiting inter-view redundancy have been conducted. For example, multi-view video coding (MVC) [13] and multi-view high-efficiency video coding (MV-HEVC) [14] have been proposed as extensions of advanced video coding (AVC), and HEVC coding standard, respectively. However, despite the efficiency of the MVC and the MV-HEVC, they are not widely used because dedicated hardware is required for real-time service. Furthermore, because the inter-view redundancy removal process in MVC and MV-HEVC exists at the codec level, these multi-view coding-based systems are not compatible with new coding standards such as VVC. Moreover, there were many technical difficulties when providing 6DoF from the computer-generated and natural contents. Therefore, the MPEG-Immersive group requested a call for proposals (CfP) for three degrees of freedom plus (3DoF+) [15]; although 3DoF+ provides a limited 6DoF experience to users, there is conceptual difference between them. Five responses for the CfP were submitted [16]–[20], based on which draft versions of the MIV and the TMIV were created. Figure 2 shows the general architecture of the MIV-compliant 6DoF immersive video streaming system. In the Figure, MIV mode encoding is conducted. The MIV encoder divides the source views into basic views and additional views, and the inter-view redundancy is removed from additional views. The MIV encoder extracts the residuals as rectangles which are defined as *patches*, and merges them into atlases. Each basic view is considered as a big patch; therefore, the basic views are also included in the atlases. In the MIV view mode, there are no additional views; only the basic views are selected from the source views and pushed into the atlases. Information about the atlases such as resolution, location of the patches, and number of frames are stored as a MIV bitstream, which follows the visual volumetric-video coding (V3C)/video-based point cloud compression (V-PCC) specification [21]. In the current MIV specification, only two texture and two geometry atlases are allowed to be generated, and the geometry atlases are down-sampled by a factor of 2×2 [22]. This is because the MIV aims to be used in high- to low-end devices; therefore, the number of videos needs to be constrained. After the atlases are generated, the video encoder encodes the atlases and the bitstreams are transmitted after encapsulation. After decapsulation, the bitstreams are decoded, and the MIV encoder synthesizes the virtual views. According to the current common test conditions (CTC) for the MIV, virtual views at the source view positions are synthesized and the quality assessment is conducted using the uncompressed source views as the reference [23]. Further, to investigate the practical performances considering the user's movement, a simulated user's movement data that is used to generate the pose trace video. After the source views are acquired, they are used to synthesize reference pose traces. Finally, an end-to-end quality assessment is conducted to evaluate the system. Note that the MPEG-Immersive group measured the objective quality only for the virtual views, and the subjective quality was evaluated only for the pose traces.

III. Fatigue-Aware 6DoF Immersive Video Quality Assessment

This section explains the experimental conditions, experimental results and analysis of the fatigue-aware 6DoF immersive video quality assessment. As mentioned in Section II, the MPEG-Immersive group measured the objective quality of the virtual views, and the subjective quality of the pose traces was evaluated. In this paper, reference pose traces were generated to measure the objective quality of the pose traces for the following reasons:

- Synthesizing the virtual views at the source view positions is not a fair evaluation for the MIV mode because the MIV view mode contains more basic views; therefore, the MIV view mode atlases will carry more information about the source views, which leads to an increase in the quality of the synthesized views.
- When using an HMD in 6DoF VR, the user can move without restrictions; therefore, the user's viewport may not be identical to the source views. As a result, pose traces can be used for the practical evaluation of HMD-based use-cases.

The remainder of this section explains the experimental conditions based on the MIV CTC, experimental results in

TABLE I: Specifications of the MIV test sequences. Abbreviations: A17, MIV mode for 17 frames; V17, MIV view mode for 17 frames

Sequence	Class	Projection	Resolution (Source view) (Atlas)	No. of source views	No. of basic views A17/V17
ClassroomVideo	CG-A	ERP	4096×2048 4096×2176	15	1 / 2
Museum	CG-B	ERP	2048×2048 2048×4352	24	2 / 4
Painter	NC-D	Perspective	2048×1088 2048×4352	16	4 / 8
Chess	CG-N	ERP	2048×2048 2048×4352	10	2 / 4
Hijack	CG-C	ERP	4096×2048 4096×2176	10	1 / 2
ChessPieces	CG-Q	ERP	2048×2048 2048×4352	10	2 / 4

terms of the objective and subjective quality and analyses considering the 2-D display and the HMD use cases.

A. Experimental Conditions

This section explains the experimental conditions based on the MIV CTC [23]. TMIV v8.0 was used to generate atlases and synthesize virtual videos [24]. Although the versatile video encoder (VVenC) and versatile video decoder (VVdeC) were adopted as the reference tools in the latest MIV CTC, the HEVC test model (HM) 16.20 was used for this experiment to save execution time and apply a tiling-based approach in future work, as the current VVenC does not support the functionality for the tiling. Similarly, a frame packing method proposed by Nokia adopted Kvazaar, an open-source HEVC encoder to apply tiling [25]. Three full-reference quality assessment metrics were used: peak signal-to-noise-ratio (PSNR), video multimethod assessment fusion (VMAF) [26], and immersive video PSNR (IV-PSNR) [27]. For the omnidirectional video, the weighted-to-spherically-uniform PSNR (WS-PSNR) was used, as specified in the MIV CTC [28]. Based on the aforementioned quality metrics, the Bjøntegaard delta rate (BD-rate) and rate-distortion (RD) curves were generated to visualize the bitrate savings when providing the same quality. There are five quantization parameters (QPs) for each test sequence; the MIV CTC uses the four higher QPs to represent the high BD-rate, and the low BD-rate is calculated using the lower four QPs. In this experiment, both high and low BD-rates were used.

For a fair evaluation, six immersive video test sequences were used: *ClassroomVideo, Museum, Painter, Chess, Hijack,* and *ChessPieces*. Table I lists the specifications of the MIV test sequences. In the table, A17 represents the experimental conditions with the MIV mode for 17 frames, and V17 means that the MIV view mode was used for 17 frames. The QPs for the MIV and MIV view modes are specified in the CTC, which were followed in this experiment.

B. Experimental Results and Analysis

This section presents and analyzes the experimental results of the MIV and MIV view modes considering the objective and subjective qualities, including fatigue. Figure 3 shows the IV-PSNR for each synthesized virtual view at the source view position for the test sequence *Painter* using the MIV and MIV view modes. For views *v2*, *v4*, *v11*, and *v13*, the MIV view mode outperformed the MIV mode. The metioned views were classified as additional views in the MIV mode, and they were selected as basic views in the MIV view mode. Although the inter-view redundancy at these views was removed from the MIV mode, the MIV view mode preserved the redundancy, which increased the quality of the synthesized virtual views at the source view positions. Indeed, it was proved that the MIV view mode is advantageous when rendering the virtual view which has the same position as the source view, as mentioned in Section III.

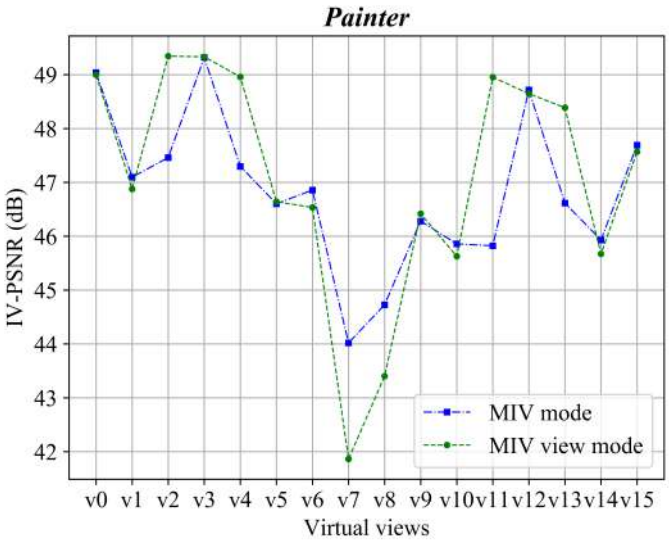

Fig. 3: IV-PSNR comparison of the virtual views at the source view positions for the test sequence *Painter*: MIV mode (58.21Mbps), MIV view mode (57.99Mbps)

Table II compares the BD-rate of the MIV mode with that of the MIV view mode. A negative value indicates that the MIV mode can save the bitrate compared to the MIV view mode while providing the same quality. At low bitrates, the MIV mode showed BD-rate savings of 7.43%, 11.05%, and 13.88% for Y-PSNR, VMAF, and IV-PSNR, respectively. Moreover, Table III shows the pose trace BD-rate of the MIV mode compared to the MIV view mode, which reveals that the MIV mode has BD-rate loss; therefore, the MIV view mode is more efficient than the MIV mode for the pose trace. However, in Tables II and III, there are several 0.00% values which implies that the the BD-rate value is too high to be computed. Therefore, the IV-PSNR RD curves of the MIV and MIV view modes are described, as shown in Figure 5. Figures 5a and 5b provides the average results of the test sequences for the virtual views and pose traces, indicating that the MIV mode is more efficient than the MIV view mode, particularly for pose traces. The MIV mode outperformed the MIV view mode in pose trace synthesis for the test sequences

Fig. 4: Synthesized pose trace comparison with enlarged noticeable sections in the test sequence *Museum*: (a), (d), (g) uncompressed pose trace generated by all source views, (b), (e), (h) pose trace generated by the MIV mode (30.89Mbps@43.52dB of IV-PSNR), (c), (f) (i) pose trace generated by the MIV view mode (30.84Mbps@34.59dB of IV-PSNR)

Museum, *Chess*, *ChessPieces*, and *Hijack*. One of the reasons for this is that the pose traces had occluded areas that were not represented by the basic views in the MIV view mode, whereas the MIV mode synthesized these areas by using the additional view patches that cannot be represented by the basic views. Figure 4 provides examples of the quality degradation of the pose trace in the MIV view mode, which shows the fatigue due to the occluded areas. However, the MIV view mode was advantageous for the test sequence *Painter*, for the following reasons: 1) the number of basic views was eight in the MIV view mode; therfore, the pose traces were synthesized without holes, and 2) the cameras in *Painter* pointed to the same direction, which minimized the occluded areas between the basic views. In contrast, because the test sequences *Museum*, *Chess*, *ChessPieces*, and *Hijack* had distracted camera coordination and a lower number of basic views was selected in the MIV view mode than these of *Painter*, the MIV mode outperformed the MIV view mode when synthesizing the pose trace. Furthermore, the test sequence *ClassroomVideo* had slightly better results in the MIV mode than the MIV view mode. Because *ClassroomVideo*

TABLE II: Virtual view BD-rate performances of the MIV mode compared to the MIV view mode (negative values indicate that the MIV mode has better performance)

Class	High Y-PSNR BD-rate	Low Y-PSNR BD-rate	High VMAF BD-rate	Low VMAF BD-rate	High IV-PSNR BD-rate	Low IV-PSNR BD-rate
CG-A	-11.02%	-5.53%	-65.03%	-19.36%	-27.17%	-14.16%
CG-B	**-54.06%**	**-41.16%**	-19.62%	-14.62%	**-65.64%**	**-49.28%**
NC-D	38.00%	34.20%	19.81%	26.02%	12.26%	20.60%
CG-N	0.00%	0.00%	**-81.47%**	-32.12%	0.00%	0.00%
CG-Q	0.00%	0.00%	0.00%	**-41.50%**	0.00%	0.00%
CG-C	0.00%	-32.11%	-25.10%	15.29%	0.00%	-40.42%
Average	-4.51%	-7.43%	**-28.57%**	-11.05%	-13.43%	-13.88%

TABLE III: Pose trace BD-rate performances of the MIV mode compared to the MIV view mode (negative values indicate that the MIV mode has better performance)

Class	High Y-PSNR BD-rate	Low Y-PSNR BD-rate	High VMAF BD-rate	Low VMAF BD-rate	High IV-PSNR BD-rate	Low IV-PSNR BD-rate
CG-A	63.25%	36.74%	46.46%	24.50%	2.70%	2.51%
CG-B	**0.00%**	0.00%	**-38.92%**	**-26.39%**	0.00%	0.00%
NC-D	63.13%	49.14%	16.32%	26.78%	28.28%	28.13%
CG-N	**0.00%**	**-63.31%**	95.31%	67.40%	**-73.77%**	**-39.48%**
CG-Q	**0.00%**	0.00%	0.00%	183.7%	0.00%	0.00%
CG-C	**0.00%**	0.00%	-37.53%	1.17%	0.00%	0.00%
Average	21.06%	3.76%	13.61%	46.19%	**-7.13%**	-1.47%

contained multiple omnidirectional 360-degree videos and the distance between the cameras were relatively short, the two modes showed similar results. Consequently, to minimize the fatigue for the 6DoF immersive video, the MIV mode that eliminates the inter-view redundancy is advantageous when the directions of the cameras are diverse, and few basic views are included in the atlases. Moreover, the MIV view mode has advantages when there are few differences between multi-view videos. Otherwise, if many basic views are selected, the MIV view mode can be effective. Furthermore, pose trace-based assessment is advantageous for HMD-based use-cases. Evaluation using a virtual view that has the same or larger FoV with the source views is beneficial when the target use-case is based on a 2-D display.

IV. Conclusion

This paper proposes a fatigue-aware 6DoF immersive video streaming approach for the MIV coding standard. Specifically, the MIV has two compression modes that have their advantages and drawbacks: the MIV mode, which eliminates the inter-view redundancy, and the MIV view mode, which exploits the multi-view coding by selecting representative views from the source views. For immersive videos with diverse camera coordination, the MIV mode is desirable in terms of the minimization of fatigue. Although cameras of an immersive video have large differences in orientation, the MIV view mode can minimize fatigue when there are several basic views in the atlases. For quality assessment, pose trace, which mimics the user's movement, is useful for the HMD-based streaming system, whereas the virtual view-based evaluation has its advantages when using a 2-D display. In the future,

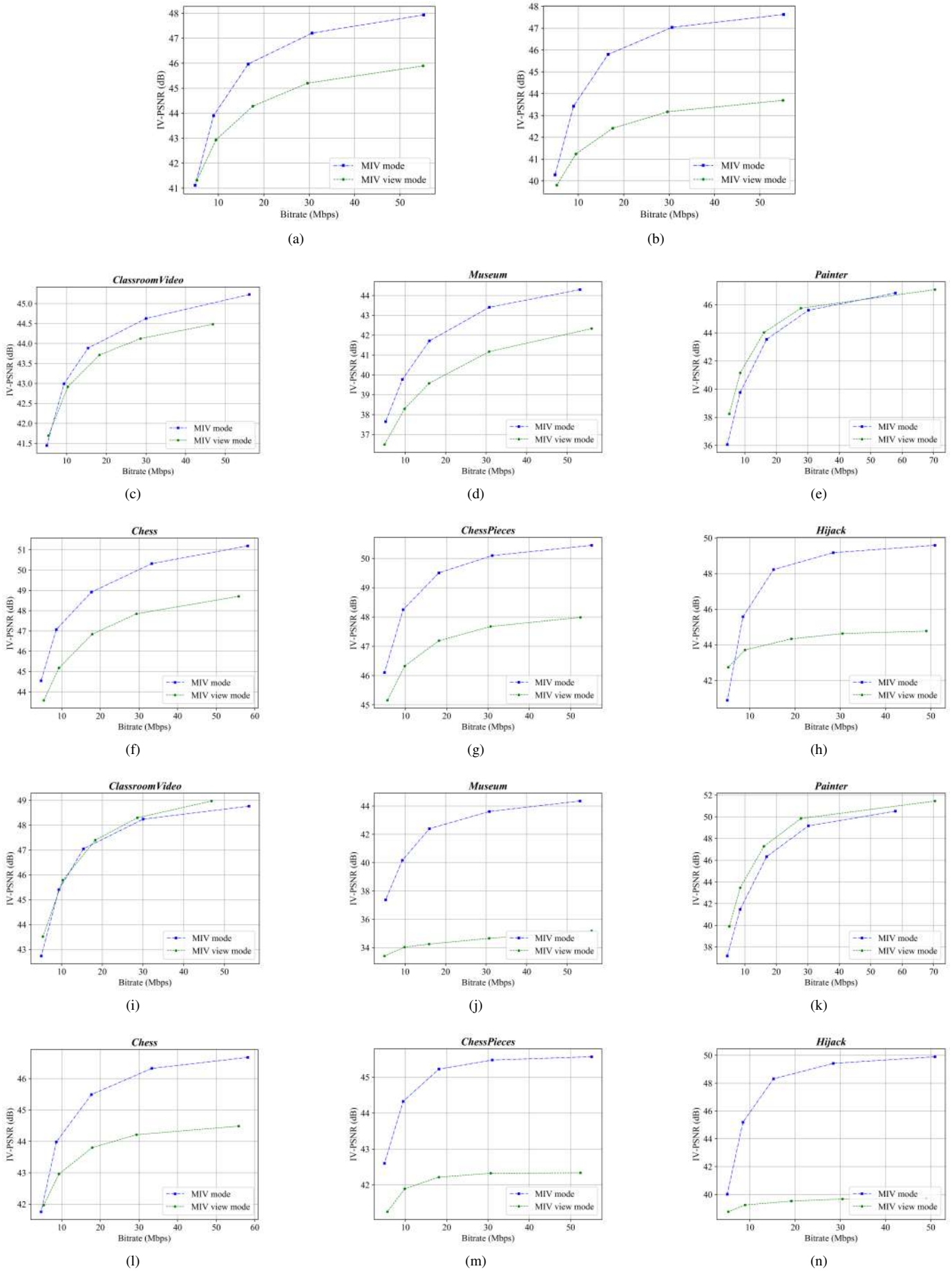

Fig. 5: IV-PSNR RD curves of the synthesized virtual views at the source view position for (a) average, (c) *ClassroomVideo*, (d) *Museum*, (e) *Painter*, (f) *Chess*, (g) *ChessPieces*, and (h) *Hijack*; and IV-PSNR RD curves of the pose traces for (b) average, (i) *ClassroomVideo*, (j) *Museum*, (k) *Painter*, (l) *Chess*, (m) *ChessPieces*, and (n) *Hijack*.

extensive experiments for deriving equations that determine the optimal coding modes for the MIV will be conducted.

ACKNOWLEDGMENT

Following are results of a study on the "Convergence and Open Sharing System" Project, supported by the Ministry of Education and National Research Foundation of Korea. (No. COSS-2021-A1-01, Development of Metaverse-Education Platform Using AI/VR/AR Technologies)

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] D. V. Nguyen, T. T. Le, S. Lee, and E.-S. Ryu, "SHVC tile-based 360-degree video streaming for mobile VR: PC offloading over mmWave," *Sensors*, vol. 18, no. 11, p. 3728, 2018.

[5] J.-B. Jeong, S. Lee, I. Kim, and E.-S. Ryu, "Implementing Viewport Tile Extractor for Viewport-Adaptive 360-Degree Video Tiled Streaming," in *2021 International Conference on Information Networking (ICOIN)*. IEEE, 2021, pp. 8–12.

[6] 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.

[7] 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.

[8] H.-H. Kim, S.-G. Lim, G. Lee, J. Y. Jeong, and J.-G. Kim, "Efficient Patch Merging for Atlas Construction in 3DoF+ Video Coding," *IEICE Transactions on Information and Systems*, vol. 104, no. 3, pp. 477–480, 2021.

[9] H. Park, S.-h. Park, and J.-W. Kang, "Toward 6 Degree-of-Freedom Video Coding Technique and Performance Analysis," *Journal of Broadcast Engineering*, vol. 24, no. 6, pp. 1035–1052, 2019.

[10] 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.

[11] 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.

[12] 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.

[13] Y. Chen, M. M. Hannuksela, T. Suzuki, and S. Hattori, "Overview of the MVC+ D 3D video coding standard," *Journal of Visual Communication and Image Representation*, vol. 25, no. 4, pp. 679–688, 2014.

[14] 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.

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

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

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] B. Choi, Y. Wang, M. Hannuksela, Y. Lim, and A. Murtaza, "Information technology–coded representation of immersive media (MPEG-I)–part 2: Omnidirectional media format," *ISO/IEC*, pp. 23 090–2, 2017.

[22] B. Salahieh, J. Jung, A. Dziembowski, and C. Bachhuber, "Test Model 8 for MPEG Immersive Video." Standard ISO/IEC JTC1/SC29/WG4, MPEG/n0050, 2021.

[23] J. Jung and B. Kroon, "Common Test Conditions for MPEG Immersive Video." Standard ISO/IEC JTC1/SC29/WG4, MPEG/n0051, 2021.

[24] Test model for immersive video (TMIV) v8.0. [Online]. Available: https://gitlab.com/mpeg-i-visual/tmiv/-/tree/v8.0

[25] L. Kondrad, V. K. M. Vadakital, and L. Ilola, "CE-1.3: frame packed video sub-bitstream type." Standard ISO/IEC JTC1/SC29/WG11, MPEG/m54274, 2020.

[26] C. G. Bampis, A. C. Bovik, and Z. Li, "A Simple Prediction Fusion Improves Data-driven Full-Reference Video Quality Assessment Models," in *2018 Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 298–302.

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

[28] Y. Sun, A. Lu, and L. Yu, "Weighted-to-spherically-uniform quality evaluation for omnidirectional video," *IEEE signal processing letters*, vol. 24, no. 9, pp. 1408–1412, 2017.