2020 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB) | 978-1-7281-5784-9/20/$31.00 ©2020 IEEE | DOI: 10.1109/BMSB49480.2020.9379498

Scalable Coding of 360-degree Video for Streaming Adaptation at 5G Network Edges

J. Carreira*†, Sergio M. M. de Faria*†, Luis M. N. Tavora†, Antonio Navarro*‡, Pedro A. A. Assuncao*†
*Instituto de Telecomunicações, Portugal
† Polytechnic of Leiria / ESTG, Portugal
‡Universidade de Aveiro, Portugal
{*jcarreira, sergio.faria*}*@co.it.pt, luis.tavora@ipleiria.pt, antonio.navarro@av.it.pt, amado@co.it.pt*

***Abstract*—The huge amount of data that is necessary to capture the full field-of-view (FoV) in omnidirectional video, i.e., 360°, imposes the use of highly efficient compressed formats as well as adaptive broadcast and streaming mechanisms, such as those foreseen for 5G networks. To cope with the demanding requirements of 360° video streaming over 5G networks, this work proposes a scalable 360º video coding architecture, by enabling adaptation through the Multi-Access Edge Computing (MEC) server in two different domains of the spherical visual content, namely spatial resolution and FoV. In the proposed architecture two-layers are encoded from the input 360° video content: (i) the base-layer (BL), encoding each 360° image as a whole, at a lower spatial resolution; (ii) the enhancement-layer (EL) encoding each spherical image as a set of multiple FoVs with higher spatial resolution. Such arrangement enables flexible stream adaptation for the smart decision algorithms to be implemented at the MEC server, enabling significant reduction of the overall bit rate through the radio interface. The simulation results show that the proposed scalable coding scheme allows a great deal of bit rate savings across the 5G network, achieving 36% of bit rate saving, on average, for a 90° FoV in comparison with conventional single-layer coding.**

***Index Terms*—Scalable video coding & Content adaptation, Future broadcasting services.**

I. Introduction

In recent years, omnidirectional video has been increasingly integrating applications and devices in the consumer market, providing enhanced multimedia experiences and pushing forward the quality requirements towards Ultra-High Definition (UHD), e.g., 4K or 8K at 60Hz or 120Hz. The huge amount of data that is necessary to represent the full Field-of-View (FoV), i.e., 360°, of any arbitrary visual scene imposes the use of highly efficient compressed formats, such as the current High Efficiency Video Coding (HEVC) standard [1], as well as adaptive broadcast and streaming infrastructures, such as those foreseen for 5G networks [2], [3]. The forthcoming 5G networks are particularly suited to satisfy the requirements of high bandwidth and fast interaction support through the dense deployment of a wireless infrastructure capable of achieving increased capacity with low-delay and further providing computational resources at network edges through Multi-access Edge Computing (MEC) [4].

This work was supported by Programa Operacional Regional do Centro, project ARoundVision CENTRO-01-0145-FEDER-030652 and by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/EEA/50008/2020, Portugal.

In general, multimedia users of omnidirectional video, e.g., applications using Head Mounted Display (HMD), are able to dynamically change their viewing direction to focus the visual attention on certain specific regions of a spherical scene [5]. This means that only a limited area of a whole spherical scene is required for display at the end user device at any specific time instant. Therefore, it is not absolutely necessary to deliver and decode the whole 360° video content because only a reduced FoV (e.g., a viewport) is displayed at the receiver-side. In this case, scalable video coding schemes can be used to save a great deal of transmission bandwidth and decoding resources by providing flexible coding and delivery mechanisms for reduced FoV, while still allowing dynamic adaptation to users' needs [6], [7]. Previous works have studied the capabilities of 5G networks to deliver UHD [8] and immersive video such as omnidirectional video [9]. Evolution from conventional video broadcasting to 360° video, maintaining backward compatibility was also recently addressed in [10]. However, backward compatible approaches may not provide full FoV access to all users, which could only be achieved by broadcasting a single-layer low-resolution stream of the whole omnidirectional video. Another shortcoming of many existing works using scalable 360° video coding is the performance evaluation in terms of the bit rate required by multiple layers versus that of non-scalable solutions capable of providing the same visual quality.

Fig. 1 shows the context of application of this work, where a 5G broadcasting environment delivers 360° video from a content server through the MEC server, which is typically located close the network edge (i.e., the radio access network). Such edge computing architecture provides computing resources close to users, enabling local and personalised real-time VR and fast content processing functionalities, giving rise to a whole new range of possibilities as discussed in recent studies [11]. The MEC server may be used to provide storage and support for dynamic multi-layer 360° video streaming, by implementing smart decisions in regard to partial content delivery according to user and network parameters related to navigation, interaction, available bandwidth, channel capacity, etc. As presented in [11], this type of architecture enables significant reduction of the overall bit rate through the radio interface while still guaranteeing high Quality of Experience (QoE).

978-1-7281-5784-9/19/$31.00 ©2020 IEEE

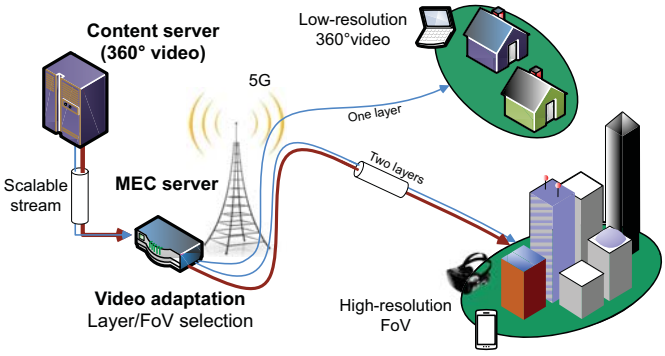

Fig. 1. 360° video delivery over 5G networks for matching low-delay adaptation requirements of FoV / resolution.

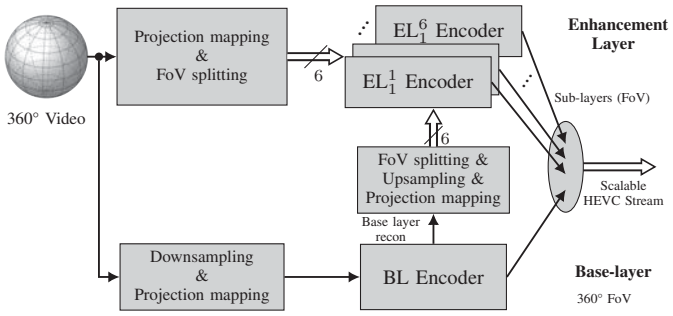

Fig. 2. Proposed scalable 360° encoding scheme.

Different approaches have been proposed to reduce the overall amount of 360° video coded data delivered through networks. For instance, tile-based approaches map the whole 360° video scene into tiles which are independently processed, allowing receivers to select individual tiles according to the viewing area or relevant viewport [12], [13]. In [14] tiles are encoded with different levels of quality and spatial resolution in order to allow fine grain operating levels in terms of bit rates and QoE. While adaptive tile-based 360° video streaming can achieve significant bit rate reduction, it has been shown that the end-to-end delay is a critical factor in the overall QoE.

This work advances one step beyond existing methods by proposing a scalable 360° video coding architecture, for efficient and low-delay delivery over 5G networks by enabling adaptation through the MEC server in two different domains of the spherical visual content, namely spatial resolution and FoV. Two-layers are encoded from the original full resolution 360° video content, with a lower spatial resolution representation in the base-layer (BL) and an Enhancement-Layer (EL) with high resolution differentially encoded, but independently from other FoVs. Such arrangement enables different delivery options for the smart decision algorithms to be implemented at the MEC server. For instance, the MEC adaptation engine may decide to deliver either (i) the BL to users with lower available bandwidth and/or basic service subscriptions, or (ii) the BL plus one or several high resolution FoVs to users with high access bandwidth and/or premium service subscriptions. The proposed scalable coding architecture along with the quality and bit rate tradeoffs of this type of adaptive solution is discussed next.

The remainder of the paper is organised as follows. Section II describes the proposed scalable 360° video coding architecture. Section III presents and discusses the results and performance evaluation and Section IV concludes this paper.

II. Proposed Method

This section describes the proposed scalable 360° video coding architecture for efficient and low-delay delivery over 5G networks by enabling adaptation through the MEC in two different domains of the spherical visual content, namely spatial resolution and FoV. Two-layers are encoded from the original full resolution 360° video content, providing different levels of functionalities. The BL carries the full FoV (i.e. 360°) at a lower spatial resolution using the polyhedron-based Cube-Map Projection (CMP) - note that any other projection can be used, such as Equirectangular Projection (ERP). The EL, carries a higher spatial resolution arranged as a set of different FoVs covering the full 360°. Each FoV in the EL is differentially encoded with respect to the BL, but independently from other FoVs. Such arrangement enables different delivery options for the smart decision algorithms to be implemented at the MEC server.

Fig. 2 shows the scalable coding architecture of the proposed scheme. The input 360° video is downsampled and a projection mapping is applied before encoding the BL as a low-resolution and low-bit rate version of the whole 360° video. In the EL, the higher resolution of the 360° video is split into multiple FoVs, which are then efficiently encoded by using predictions from the BL. To this end, the original high resolution 360° video is mapped into a planar representation by using the CMP, which results in a 2D mapping with six different FoVs. Then, the residual information obtained after predicting each FoV from the corresponding BL counterpart, is encoded as a sub-layer of the EL. In order to decrease the overall bitrate, the reconstructed BL is used for inter-layer prediction, since the BL is always available at the decoder.

A. Enhancement-layer coding and delivery

The proposed scalable 360° coding architecture takes advantage of two layers to provide different levels of quality and reduced bitrate. Moreover, in order to achieve an efficient decoding the polyhedron-based CMP video is independently encoded. Fig. 3-(a) shows the six cube faces of the Cube-Map Projection (CMP) representation, which correspond to six independent FoVs identified by their initials. Conventional coding schemes arrange all cube faces into a sequence of single rectangular matrices (i.e., each frame with 2×3 cube faces) to encode the whole 360° video, as illustrated in Fig. 3-(b). However, since this is a rigid coding approach, it does not allow any kind of flexibility to extract and decode either only one or few FoVs, i.e., the whole 360° video must be always decoded, even when only a smaller region is necessary for display.

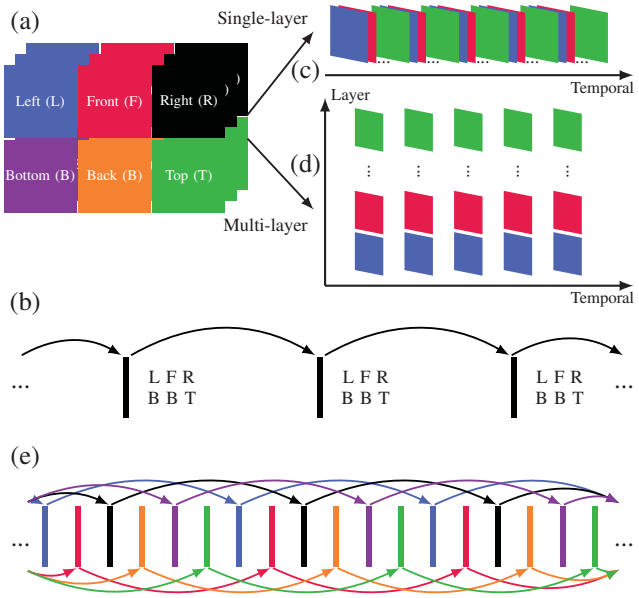

Fig. 3. Coding approaches and respective dependencies.

The flexible solution proposed in this work is to consider each cube face (i.e., each FoV) as a single frame. Then the FoVs can be simply encoded as a unique frame sequence, as shown in Fig. 3-(c). The scalable alternative that is able to provide further flexibility is to encode each FoV as a frame sequence in a different layer, as shown in Fig. 3-(d).

This approach requires constrained temporal predictions to avoid useless inter-FoV prediction between uncorrelated FoVs, i.e., only prediction between frames from the same cube face is allowed, as illustrated by the temporal dependencies in Fig. 3-(e) Therefore, by using such constrained prediction approach it is possible to decode one or more 90° FoVs by simple truncation of the single compressed stream containing the full FoV video. In the proposed coding architecture this is achieved by encoding each FoV into a different layer, where the inter-layer dependency only exists with the BL. Therefore, one can classify these encoded FoVs as sub-layers of the EL. Since there is no dependency between sub-layers, each FoV can be independently decoded at the EL, thus avoiding delivery/decoding of useless information from other FoVs.

Furthermore, switching/navigation between FoVs incurs in lower delay in comparison with other approaches (e.g., tile based single-layer coding), because for the same transmission rate, the amount of data that is necessary to reach the receiver is lower in the case of the proposed scheme. This is achieved by performing the FoV switching at the EL, which represents much lower amount of coded data compared with a single-layer high resolution FoV. In a 5G network, the MEC adaptation engine may decide to deliver either (i) the BL to users with lower available bandwidth and/or basic service subscriptions, or (ii) the BL plus one or several high resolution FoVs of the EL to users with high access bandwidth and/or premium service subscriptions. Moreover, under dynamic bandwidth

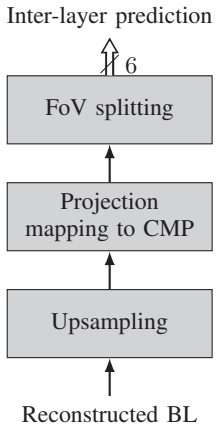

Fig. 4. Processing steps of the proposed inter-layer prediction.

constrains, the proposed EL coding solution allows the MEC server to discard the unnecessary EL sub-streams, saving bandwidth for the remaining ones.

B. Inter-layer prediction

As mentioned before, inter-layer prediction contributes to achieve efficient encoding of in the EL, by using the lower resolution BL as reference. Since the BL encodes the whole 360° FoV in a single stream, it provides predictions for all FoVs encoded in the EL.

Fig. 4 illustrates the various processing steps required to build the inter-layer predictions from the reconstructed BL. As in spatial scalable coding approaches [15], the BL reconstruction is firstly upscaled to match the resolution of the EL. Moreover, as different planar projections are allowed at the BL, such as ERP, OHP, ISP among others, a projection mapping to CMP is necessary to obtain the resulting 360° images represented as six cube faces. Finally, since multiple sub-layers are used in the EL, FoV splitting is used to convert the six cube faces into six different FoV predictions for corresponding EL encoders.

III. Performance Evaluation

The performance of the proposed scalable coding scheme was evaluated from two different perspectives. On the one hand, the rate-distortion performance achieved by the scalable 360° video coding architecture is compared with simulcast using the same two resolutions and also with non-scalable encoding (single-layer). On the other hand, the proposed scalable coding scheme was evaluated for a 5G networking scenario, by simulating two adaptive options to run at the MEC server, both of them capable of providing the same service functionalities, i.e., access to 360° video content with two different resolutions.

The five 360° video sequences presented in Table I were used in the simulations. These are included in the data set of the Common Test Conditions (CTC) [16]. The omnidirectional video content has a spatial resolution of 8192×4096 in the ERP format. This test material covers a wide range of

TABLE I
TEST SEQUENCES USED IN THE EXPERIMENTS.

Sequence	SI	TI	Description
ChairliftRide	31.25	8.50	Chairlift ride with moderate motion and trees in the background
GasLamp	28.93	2.15	Street interception with cars and persons passing through
KiteFlite	51.82	4.25	Persons moving with moderate motion and several kites shaking with a building in background
SkateboardInLot	27.58	13.8	Self-captured skateboard ride with high motion in a parking lot
Trolley	54.36	4.85	Garden with buildings in the background and a train passing through

TABLE II
BJØNTGAARD DELTA-RATE USING SIMULCAST AS REFERENCE.

Sequence	Total bit rate		EL bit rate
	Non-scalable (single-layer)	*Scalable (two layers)*	
ChairliftRide	-28.73	-19.47	-27.37
GasLamp	-31.63	-17.91	-31.86
KiteFlite	-32.40	-15.52	-27.61
SkateboardInLot	-28.73	-20.59	-25.68
Trolley	-33.46	-24.80	-40.35
Average	**-30.99**	**-19.66**	**-30.57**

motion and texture complexities, as shown by the different spatial (SI) and temporal information (TI) [17] presented in Table I. Following the CTC, the original video resolution of 8192×4096 was converted to CMP with resolution of 6144×4096. All methods under comparison use the CMP format to ensure a fair comparison. For encoding both the BL and EL of the proposed architecture, the scalable extension of the HEVC standard (SHVC) [15] was used. An IDR period of 16 frames was used with Random-Access and hierarchical B-frames, using 4 reference frames. The WS-PSNR metric was used to measure quality. This is commonly used to evaluate the end-to-end quality in 360° video delivery systems [18].

A. Rate-Distortion performance

The rate-distortion performance was evaluated by comparing the proposed scalable coding architecture with the following: (i) *simulcast coding* using two independent streams, both of them encoding the full FoV, i.e. 360°, at the same two different resolutions as provided by the scalable scheme; (ii) *non-scalable (single layer) coding* using the full FoV at the highest resolution.

Table II shows the Bjøntgaard Delta-Rate (BD-Rate) results obtained using simulcast as reference. Negative percentages correspond to less bit rate than the reference, at the same average quality, i.e., WS-PSNR. In columns 2 and 3 of Table II, the BD-Rate was computed using the total bit rate of all layers. These results show that the proposed method clearly outperforms the case of simulcast, which is due to the efficiency gains obtained by using inter-layer prediction in the scalable coding scheme. As expected, by comparing the bit rate savings obtained by the scalable and the non-scalable streams, the former achieves lower gains, due to the excess bit rate required to encode the EL. This represents the cost incurred by scalable coding of the 360° video into the two layers. However, it is worthwhile to notice that, while the average bit rate difference between the scalable (proposed) and non-scalable coding of 360° video is about 11% (last line of Table II), the corresponding difference using SHVC with conventional video and the same configuration (i.e., Random Access) is 19%, as found in [15]. Therefore, the proposed scalable 360° video coding scheme achieves better efficiency in the EL than SHVC with conventional video.

The last column of Table II shows the BD-Rate of the EL, using the total bit rate of the simulcast stream with the highest resolution as reference for comparison. These results confirm that both resolutions of 360° video can be delivered with 30.57% less average bit rate by using scalability rather than simulcast. Obviously this has significant impact on the storage capacity required in the MEC server as well as in the transmission bandwidth. Furthermore, by comparing the bit rate savings with the Spatial Information (SI) in Table II, one can notice higher savings for video content with higher spatial complexity (i.e., Trolley sequence) up to 40.35%.

TABLE III
MAXIMUM BIT RATE REQUIRED TO DELIVERY UP TO 90° ×90° FOV.

Sequence	Bit rate (MBps)	
	Non-scalable (single-layer)	*Scalable (two layers)*
ChairliftRide	1.96	1.26 (-36%)
GasLamp	1.02	0.62 (-39%)
KiteFlite	3.21	2.08 (-35%)
SkateboardInLot	3.34	2.07 (-38%)
Trolley	2.33	1.50 (-36%)
Average	**2.37**	**1.50 (-36%)**

B. 5G networking scenario - adaptation through MEC

To demonstrate the benefits provided by the proposed encoding architecture, a 5G networking scenario was simulated, by considering that adaptation is carried out at the MEC server to only deliver a limited FoV (less than 360°). In this simulation a maximum FoV of 90° is considered for extraction from the scalable stream and delivery to users. In order to avoid content biased results, all 90° FoVs that are possible to decode from the EL were used. Table III shows the maximum bit rate that must be delivered by the MEC server, either considering the scalable or the non-scalable stream, both allowing to decode any of such FoVs. Note that this corresponds to the worst possible case for the proposed scalable scheme since it considers the case when 4 sub-layers are decoded to obtain a 90° ×90° FoV. For the non-scalable, the whole bit rate must be delivered because it is not possible to extract individual FoVs. The results presented in the Table reveal that the proposed scalable scheme allows to deliver much less bit rate than a conventional non-scalable system. Although higher overall bit rate is required for scalable coding of the whole 360° video (see Table II), the use of scalability allows a great deal of bit rate savings across the 5G network

because only the necessary FoV is delivered. On average, for delivering a 90° FoV, 36% of bit rate saving is obtained with the proposed scheme in comparison with conventional single-layer coding. In addition, as mentioned before, the proposed scheme has the ability to deal with FoV switching and navigation under reduced delay. As final remark, it should be emphasised that all bit rate comparisons presented and discussed above are done under fair and valid conditions, i.e., at the same 360° video resolution and same average quality, measured as WS-PSNR.

IV. Conclusions

In this paper a flexible scalable coding approach for 360° video delivery over 5G networks was proposed, to enable edge computing adaptation in two different domains of the spherical visual content, namely spatial resolution and FoV. The proposed two-layer architecture encodes a BL to carry the full FoV at a lower spatial resolution and an EL carrying a higher spatial resolution, which in turn is split into a set of different FoVs encoded as multiple sub-layers. The simulation results have shown that constraining sub-layers to only use intra-FoV prediction, enables flexible delivery options for smart decision algorithms to be implemented at 5G MEC servers, also allowing up to 39% reduction of the overall bit rate through the radio interface. Summarising, the proposed scalable coding solution is seen as an efficient approach to benefit from the edge computing capabilities of 5G networks.

References

[1] M. Wien, J. M. Boyce, T. Stockhammer, and W. Peng, "Guest editorial immersive video coding and transmission," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, p. 1–4, Mar. 2019.

[2] C. Colman-Meixner, P. Diogo, M. S. Siddiqui, A. Albanese, H. Khalili, A. Mavromatis, L. Luca, A. Ulisses, J. Colom, R. Nejabati, and D. Simeonidou, "5G city: A novel 5G-enabled architecture for ultra-high definition and immersive media on city infrastructure," in *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun. 2018, p. 1–5.

[3] C. Ge, N. Wang, I. Selinis, J. Cahill, M. Kavanagh, K. Liolis, C. Politis, J. Nunes, B. Evans, Y. Rahulan, N. Nouvel, M. Boutin, J. Desmauts, F. Arnal, S. Watts, and G. Poziopoulou, "QoE-assured live streaming via satellite backhaul in 5G networks," *IEEE Transactions on Broadcasting*, vol. 65, no. 2, p. 381–391, Jun. 2019.

[4] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, p. 1657–1681, May 2017.

[5] W. Mason. (2015, Aug.) VR HMD Roundup: Technical Specs. [Online]. Available: https://uploadvr.com/vr-hmd-specs

[6] M. Zink, R. Sitaraman, and K. Nahrstedt, "Scalable 360° video stream delivery: Challenges, solutions, and opportunities," *Proceedings of the IEEE*, vol. 107, no. 4, p. 639–650, Apr. 2019.

[7] D. Liu, P. An, R. Ma, W. Zhan, and L. Ai, "Scalable omnidirectional video coding for real-time virtual reality applications," *IEEE Access*, vol. 6, p. 56323–56332, Oct. 2018.

[8] P. Salva-Garcia, J. M. Alcaraz-Calero, R. M. Alaez, E. Chirivella-Perez, J. Nightingale, and Q. Wang, "5G-UHD: Design, prototyping and empirical evaluation of adaptive ultra-high-definition video streaming based on scalable H.265 in virtualised 5G networks," *Computer Communications*, vol. 118, p. 171–184, 2018.

[9] L. Sun, F. Duanmu, Y. Liu, Y. Wang, Y. Ye, H. Shi, and D. Dai, "Multi-path multi-tier 360-degree video streaming in 5G networks," in *Proceedings of the 9th ACM Multimedia Systems Conference*, ser. MMSys '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 162–173. [Online]. Available: https://doi.org/10.1145/3204949.3204978

[10] T. Biatek, W. Hamidouche, P. Cabarat, J. Travers, and O. Déforges, "Scalable video coding for backward-compatible 360° video delivery over broadcast networks," *IEEE Transactions on Broadcasting*, p. 1–11, 2019.

[11] X. Hu, W. Quan, T. Guo, Y. Liu, and L. Zhang, "Mobile edge assisted live streaming system for omnidirectional video," *Mobile Information Systems*, p. 1–15, May 2019.

[12] Y. Hu, S. Xie, Y. Xu, and J. Sun, "Dynamic VR live streaming over MMT," in *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Jun. 2017, p. 1–4.

[13] T. C. Nguyen and J. Yun, "Predictive tile selection for 360-degree VR video streaming in bandwidth-limited networks," *IEEE Communications Letters*, vol. 22, no. 9, p. 1858–1861, Sep. 2018.

[14] A. TaghaviNasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using layered video coding," in *IEEE Virtual Reality (VR)*. IEEE, Mar. 2017, p. 347–348.

[15] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of SHVC: Scalable extensions of the high efficiency video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 20–34, Jan. 2016.

[16] P. Hanhart, J. Boyce, and K. Choi, "Algorithm descriptions of projection format conversion and video quality metrics in 360lib (version 9), document JVET-M1004," Joint Video Experts Team (JVET), Macau, CN, Tech. Rep., Jan. 2019.

[17] ITU-T, "Recommendation P.910, subjective video quality assessment methods for multimedia applications," Apr. 2008.

[18] X. Xiu, Y. He, Y. Ye, and B. Vishwanath, "An evaluation framework for 360-degree video compression," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, Dec. 2017, p. 1–4.