On Subpicture-based Viewport-dependent 360-degree Video Streaming using VVC

Maryam Homayouni, Alireza Aminlou, Miska M. Hannuksela
Nokia Technologies, Tampere, Finland
Email: firstname.lastname@nokia.com

***Abstract*—Virtual reality applications create an immersive experience using 360° video with high resolution and frame rate. However, since the user only views a portion of 360° video according to his/her current viewport, streaming the whole content with high resolution causes bandwidth wastage. To address this issue, viewport-dependent approaches have been proposed such that only the part of the video which falls within user's current viewport is transmitted in high quality while the rest of the content is transmitted in lower quality. The selection of high- and low-quality parts is constantly adapted according to the user's head motion, which requires frequent intra coded frames at switching points, leading to an increment in the overall streaming bitrate. In this paper a viewport-adaptive streaming scheme is introduced, which avoids intra frames at switching points by introducing long intra period for non-changing parts of the content during head motion. This scheme has been realized taking advantage of mixed Video Coding Layer (VCL) Network Abstraction Layer (NAL) unit feature of Versatile Video Coding (VVC) standard. This method reduces bitrate significantly, especially for the sequences with either no or only slow camera motion, which is common for 360° video capturing.**

Keywords—Virtual Reality, Versatile Video Coding

I. Introduction

Advances in virtual reality (VR) equipment and related standards have boosted the use of 360° video. An immersive experience in VR requires high resolution 360° video to be transmitted to the viewer. Conventional transmission of the whole 360° content, at the highest resolution and quality, not only consumes an unnecessary high network bandwidth, but also requires high computational complexity to decode the whole content. However, in practice, at any time instance, only a portion of a 360° video, limited to the field of view (FOV) of the head-mounted display (HMDs) in use, is watched by the viewer. Hence, different viewport-adaptive streaming (VAS) schemes have been developed to reduce the bandwidth consumption, in which only the current viewport of the client is transmitted in high-quality (HQ). Furthermore, due to latency of the codec and transmission system, the rest of the content (i.e. non-viewport part) is also sent in low-quality (LQ) to provide a fallback in the case of quick head movement.

In particular, tile-based VAS using the motion-constrained tile set (MCTS) technique has been suggested (e.g., [1], [2]). In tile-based VAS, the client selects at which quality and resolution each tile is received. In tile-based VAS, the term tile is used to refer to an isolated region [3] that depends only on the collocated isolated region in reference pictures and does not depend on any other picture regions. The received isolated regions are merged into a video bitstream, which is decoded by a single decoder instance. The MCTS-based VAS was compared with other tile-based VAS methods (based on Region of Interest scalability and simulcast layers at different random access periods) [4] and found to be competitive in streaming rate-distortion performance while it only requires single-layer decoding with a single decoder instance as opposed to the alternative methods requiring multi-layer decoding or multiple decoder instances. Moreover, MCTS-based VAS was found to be more efficient in streaming rate utilization and storage requirement than viewport-specific streams achieved by viewport-specific projection or viewport-specific region-wise resampling [5], [6]. Hence, tile-based VAS for High Efficiency Video Coding (HEVC) is supported in Omnidirectional Media Format (OMAF) [7], [8], developed by the Moving Picture Experts Group (MPEG).

There are several suboptimalities with VAS based on MCTSs of HEVC which have been addressed in the Versatile Video Coding (VVC) standard [9]. First, in MCTS, motion vectors are limited to point inside each tile group; this increases the bitrate about 3 to 5% and 11 to 15% for coarse and fine tiling grid, respectively [10]. VVC includes a picture partitioning unit called subpicture, which can be independent from other subpictures and hence serves as an isolated region. The rate-distortion penalty of MCTSs has been alleviated by enabling the boundaries of subpictures in the reference frames to be padded for the purpose of inter prediction. Hence, motion vectors can point outside of the subpicture area.

Second, in HEVC, merging MCTSs requires rewriting of slice headers, either assisted by instructions provided through extractor tracks [8] or as pure client-side process. Slice headers of VVC include a subpicture identifier, and the mapping of subpicture identifiers to a subpicture layout is indicated in a sequence or picture parameter set. Thus, no slice header rewriting is required for merging subpictures. Consequently, similarly to MCTSs, subpictures can be selectively merged into a bitstream for achieving VAS with a single decoder instance.

Third, HEVC requires all slices to have the same Network Abstraction Layer (NAL) unit type and consequently a viewport change is typically responded with an entire intra random access point (IRAP) picture. Hence in HEVC-based VAS, frequent intra-coded frames (e.g. every one second) are needed when there is head motion and hence switching in viewport tiles. This also increases the overall streaming bitrate. This problem can be been avoided by allowing subpictures of different video coding layer (VCL) NAL unit types in the same coded picture, as first proposed by one of the authors in [11]. In other words, some subpictures in a coded picture of the merged bitstream may originate from an

2020 IEEE International Symposium on Multimedia (ISM) | 978-1-7281-8697-9/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISM.2020.00021

978-1-7281-8697-9/20/$31.00 ©2020 IEEE
DOI 10.1109/ISM.2020.00021

IRAP picture while other subpictures of the same coded picture can have a non-IRAP type.

We noticed that when there is viewport switch due to head motion, it is typical that the new viewport covers only a few additional subpictures compared to the previous one. Consequently, most of the subpictures can be kept with the same quality and only a few of them are needed to be switched from low- to high-quality version and vice versa. In this paper, given this observation and taking the benefit of the mixed VCL NAL unit feature of VVC, we introduce subpicture-based VAS with mixed IRAP periods. In this method, two versions of every subpicture sequence are coded and stored, using short and long IRAP periods. Then during streaming, the long IRAP period version is transmitted for the subpictures for which the quality is not changed by a viewport change, while the short IRAP period version is applied for the remaining subpictures. This method effectively reduces the number of intra frames, and hence reduces streaming bitrate in both cases, with and without head motion.

The rest of the paper is organized as follows. Viewport adaptive streaming related background is reviewed in section II. Section III illustrates the proposed method. Quality evaluation, experimental conditions and results are presented in section IV and finally section V concludes the paper.

II. BACKGROUND

A. Tile and MCTS concepts in HEVC

In HEVC standard, the video pictures can be spatially split horizontally and vertically, along a grid of columns and rows, into non-overlapping rectangular shaped partitions, called tiles [11]. Since intra prediction and entropy coding do not cross tile boundaries, tiles are spatially independent, therefore they are independently decodable. However, in inter prediction, motion vectors can use any predictor beyond the co-located tile boundaries in the reference pictures. This is while in MCTS technique, first suggested in [2], an extra encoding constraint has been added, which restricts the motion vectors to point only to a set of tiles, including the co-located tiles in the reference pictures. In order to make MCTSs independently decodable, each MCTS is usually carried in a specific slice.

B. NAL unit types in HEVC and VVC

The high-level syntax of HEVC and VVC bitstreams consists of NAL units, which provide inherent support for packet-oriented network interfacing. The HEVC and VVC standards include a large variety of NAL unit types, indicated in the NAL unit headers. The NAL unit types can be categorized into VCL NAL units and non-VCL NAL units. In HEVC, all VCL NAL unit of a coded picture have the same type, which therefore also indicates the picture type. In VVC, it is indicated whether a coded picture contains subpictures with mixed VCL NAL unit types. The VCL NAL unit types can be categorized in intra random access point (IRAP) and non-IRAP NAL unit types. Pictures having a particular IRAP NAL unit type can be used for random accessing, i.e., decoding can start from an IRAP picture. An instantaneous decoding refresh (IDR) picture is such an IRAP picture which is followed, in decoding order, by pictures that can be correctly decoded when the decoding starts from the IDR picture.

C. Tile-based viewport-adaptive streaming

In adaptive streaming over HTTP, representations are time-wise partitioned into segments and the server announces the available representations in a manifest, which enables the client to derive a Universal Resource Locator (URL) for each segment of each available representation. The client then requests the content by issuing HTTP GET requests with the URLs it has selected from the manifest. Content generation for adaptive bitrate streaming includes generation of different representations of a video, which might be different in spatial resolution, quality or prediction structure. Each sequence of isolated regions can be then extracted using bitstream extractor mechanisms into a separate representation. The isolated regions with different quality can be merged at the client side to generate a decodable bitstream. So not only the client is able to request the required bitrate and resolution from the server, it should be able to concatenate, decode and play the different representations and achieve seamless playout with one single decoder instance. This capability is used in tile-based VAS and the isolated regions covering the viewport are sent in high quality while the rest of the isolated regions are sent in low quality.

D. Viewport switching in MCTS based streaming

The viewport switching can happen in each segment starting with an IRAP picture in both high and low resolution bitstream. In the other words, when the client needs to switch between MCTSs of different resolutions/quality, all received segments must start with MCTSs having the same IRAP NAL unit type at the switch point. Having frequent IRAP pictures, however, increases the required bitrate significantly, since inter prediction cannot be utilized.

E. Subpictures in VVC

The VVC standard provides new concepts such as subpictures. A subpicture is a rectangular set of slices that has its own subpicture identifier value. Slice addresses are relative to the subpicture in which they reside, and the subpicture location within the picture is determined by the subpicture layout given in the sequence parameter set (SPS) and by the mapping of subpicture identifier values to the subpicture layout, which is provided in the SPS, in picture parameter set(s), or derived implicitly. Encoders can choose to encode subpicture boundaries like picture boundaries, which makes a subpicture independent of other subpictures. In this case a subpicture is treated like a picture in the decoding process.

III. PROPOSED STREAMING AND EVALUATION METHODS

A viewing orientation change in VAS methods is conventionally handled by refreshing the entire 360° coverage with a random-access picture. However, frequent IRAP pictures in a stream pumps extra bitrate by the intra frame. To reduce the impact of IRAP frames on streaming bitrate, the IRAP picture period can be increased, but this reduces the switching capability. The main idea of this study is to stream the subpicture with longer IRAP period, when there is no head motion, and no need to switch the quality of subpicture. In the case of head motion, only few of subpictures which require quality change are switched from the long IRAP period stream to the short IRAP period stream for a short period of time.

A. Mixed-IRAP VAS

To design a rate-distortion (RD) efficient viewport adaptive streaming system for 360° video, it is important to know the head movement pattern and velocity. According to

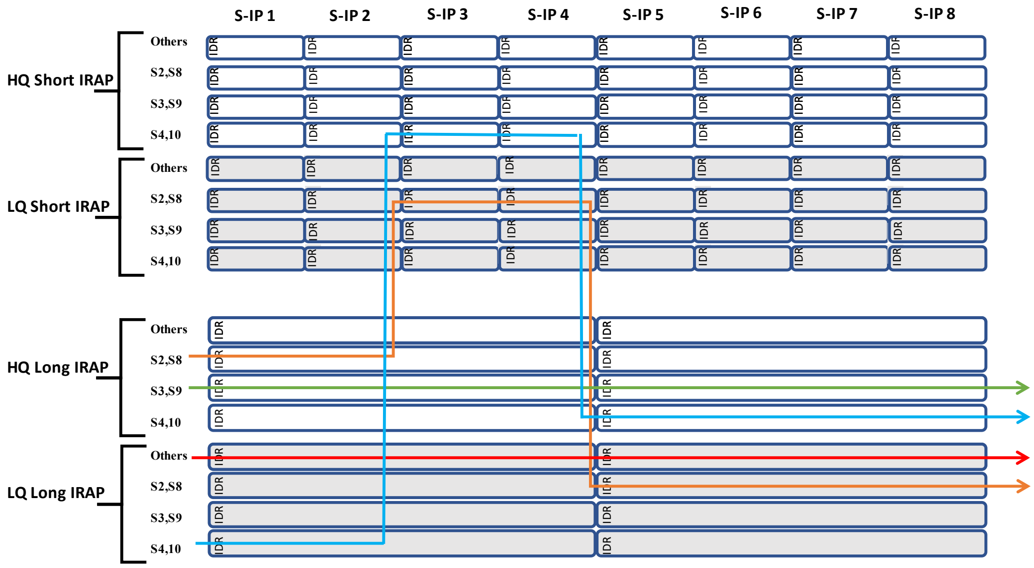

Figure 1:subpicture stream switching after head movement

the studies performed in [13] and [14] and many others, between 80% to 95% of the users move less that 90-degree within a 2 second duration segment. In another study [15] with 10 video data sets and 130 subjects, the mean angular velocity of the head movement was 10.3°/sec. As a conclusion, the isolated regions that are involved in the current viewport overlap with those of the previous one. Therefore, only a few isolated regions should be updated when viewing orientation changes.

In this paper, we propose to take advantage of the capability of utilizing mixed VCL NAL unit types within a picture and use an IRAP NAL unit type only for the subpictures that require refreshing and choose the rest of the subpictures from a stream that is encoded with less frequent IRAP pictures. To realize this method, subpictures with different qualities as well as with both short and long IRAP periods are encoded. In the performed simulations, the short IRAP period was chosen according to the common test conditions (CTC) [16] (i.e. 32 pictures, which is about one IRAP frame per second) and by the long IRAP period was 4 times of that (i.e. one IRAP frame per 4 seconds).

During the normal case, when there is no head motion, the long IRAP period version of the subpictures are streamed. When a head movement happens, the subpictures that do not need quality level change remain to be transmitted from the long IRAP period version of the content. The subpictures that need quality change are sent from the short IRAP interval bitstream for a short time until the next IRAP position in the respective long IRAP period bitstream.

A simple example is shown in Fig. 1, where the current viewport is changed from V1 to V2 (see Fig. 2) just before the start of segment 3, which put s4, s10 subpictures into the new viewport while removes s2 and s8 from the current viewport of the user. Hence, s2 and s8 need to switch from low- to high-quality, while s4 and s10 need to switch from high- to low-quality stream. Thus, segments 3 and 4 of subpictures s2, s4, s8, and s10 are streamed from the short IRAP period version, after which they are again streamed from the long IRAP period version. To keep Fig. 1 as simple as possible, it only shows subpictures in the range of s1 to s4, inclusive.

B. Quality evaluation method

To compare the RD performance of the proposed usage of mixed IRAP periods with that of the short IRAP period, we compare the RD performance of the two methods in two cases, i.e., with and without head motion. When there is no head motion, the viewer watches the same part of the content and does not turn his/her head, so there is no need to switch the viewport and subpicture bitstreams.

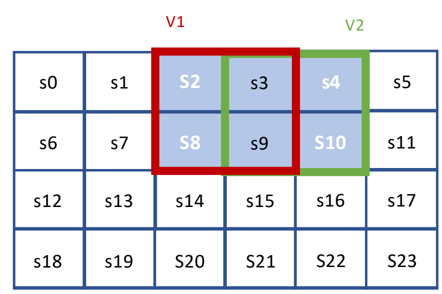

Figure 2: Subpictures involve in viewport change S2, S8 from high to low quality and S4, S10 from low to high quality.

TABLE I. SWITCHING STREAM AND THE WEIGHT FACTORS FOR PSNR CALCULATION IN THE CASE OF HEAD MOTION

Switching IP	Stream	S-IP1 (0-31)	S-IP2 (32-63)	S-IP3 (64-95)	S-IP4 (96-127)
1	S-IRAP	0	0	0	0
	L-IRAP	P	P	P	P
2	S-IRAP	0	M	M	M
	L-IRAP	P	P-M	P-M	P-M
3	S-IRAP	0	0	M	M
	L-IRAP	P	P	P-M	P-M
4	S-IRAP	0	0	0	M
	L-IRAP	P	P	P	P-M
Average	S-IRAP	0	M/4	2M/4	3M/4
	L-IRAP	P	P-M/4	P-2M/4	P-3M/4

In the case of head motion, we analyze the RD performance when the head motion causes only one occurrence of bitstream switching within one IRAP period (IP) of the long IRAP period bitstream. Since the switching effect does not propagate to the next long-IP, the bitrate and peak signal-to-noise ratio (PSNR) is calculated for each long-IP and the average PSNR and bitrate for a sequence is calculated by averaging over different long-IPs.

Let N be the total number of subpictures, P be the number of high-quality subpictures in streaming, Q = N − P be the number of low-quality subpictures, and M be the number of subpictures that are changing in a single head motion. Bitstream switching can happen at the beginning of each short-IP (S-IP) segment. Since in our test setup there are four short-IP in each long-IP, four cases can happen with the same probability, as defined below and summarized in TABLE I:

- S-IP1: If the switching happens at the beginning of the first short IRAP period segment, then all the subpictures are transmitted using the long IRAP period versions and there is no need to stream any short IRAP period versions.
- S-IPn where n is 2, 3, or 4: If the switching happens at the beginning of the n-th short IRAP period segment, for the first n×32 frames all the streams are from the long IRAP period versions, while the remaining segments of subpictures that need a quality switch are from the short IRAP period versions.

Note that since different cases have equal probability, the average PSNR can be calculated as an average of above cases. TABLE I shows the weight factors that can be used to calculate bitrate and PSNR.

In order to simplify the PSNR calculation, it has been assumed that the viewport has uniform distribution on the sphere, so the P high-quality subpictures have uniform distribution and their bitrate and PSNR can be averaged to calculate the RD performance of the whole sequence. Moreover, since subpictures have a uniform distribution, we can average their bitrate, MSE, or PSNR values to derive the expected values. The same method is used to calculate the bitrate and PSNR of the anchor.

IV. Experimental Evaluation

A. Experimental setup

The simulations are performed using 300 frames original uncompressed 360° video test sequences, listed in TABLE II. As described in JVET-M0870 [17], the 6K and 8K 360° video test sequences, defined in [18], were first converted to the CMP format with a face size of 1536×1536 and 2048×2048, respectively, using JVET 360Lib-10.0 software [19]. Uncompressed video was then partitioned to individual, non-overlapping subpictures of size 512×512 before encoding. Each subpicture is then coded with an individual encoder. In other words, each 1536×1536 cube face of a 6K sequence was split to 3×3 subpictures, while the 2048×2048 cube faces of 8K sequences were split to 4×4 subpictures, resulting in 54 and 96 individual subpictures, respectively. Fig. 3 shows the system setup. Assuming that half of the subpictures are sent in high-quality, in performance evaluation, method N is the number of subpictures, P = N/2; therefore Q = N/2 and M is set to 10 and 12 for 6K and 8K pictures, respectively. The subpicture sequences were coded with VTM 8.0 [20], using random access (RA) configuration. luma mapping with chroma scaling in SPS, scaling list in SPS, sample adaptive offset and joint chroma coding residual sign flag were disabled for all tests to enable independent coding of subpictures according to our simulation setup. All the subpictures/sequences were encoded with quantization values of 20, 23, 26, and 29. This range of values was selected to obtain acceptable picture quality for viewing on a head-mounted display.

B. Test cases

The performance of three test scenarios have been assessed in this paper, random access CTC (short IRAP), long IRAP and mixed IRAP. Each scheme is described in more details below.

Short IRAP refers to the case where each subpicture is individually coded with random access configuration recommended by [16], with GOP size equal to 16 and IRAP period is equal to 32, meaning every 32 frame, is coded with intra mode. In long IRAP scenario, the stream has GOP size twice the normal size (32) with IRAP interval picture of 128, which makes the bitrate to be used more efficiently. We expect that long IRAP scenario has better RD performance.

In the mixed VCL NAL unit type scenario, presented here, for each quality, there is one short IRAP stream with GOP size of 16, and IRAP interval picture of 32. while the long IRAP version of the stream has GOP size 32 with IRAP interval picture of 128. As mentioned earlier in section III, when viewing orientation changes, some of the subpictures should switch from high-quality to low-quality and vice versa. For those, the client switches the stream from long IRAP to short IRAP and switch back to long IRAP when the subpicture remains in steady state (on the next coming long IRAP IDR)

C. PSNR calculation

PSNR, as the most commonly used objective quality measure in image/video coding communities, is used to evaluate the quality of subpicture coding. Traditionally Mean Square Error (MSE) between the pixels of the coded picture and the original uncompressed one is calculated to measure

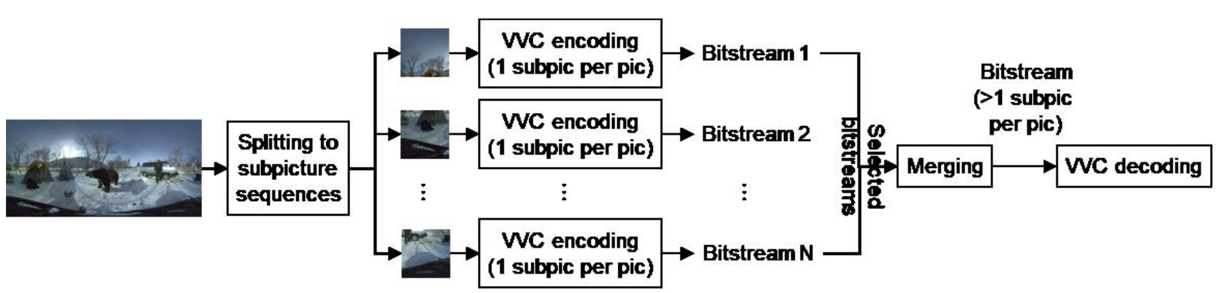

Figure 3:Selective subpicture based operation system [21]

the PSNR of a frame. The PSNR of the whole video is then calculated by averaging over PSNRs of frames. To be aligned with this definition, for a sequence with n frames and m individual subpictures, the PSNR is calculated as follows:

In this study, to enable parallel encoding of subpictures for reducing the simulation time, we split the video sequence to independent subpictures. To calculate the PSNR, aligned with the anchor method described above, first, the PSNR of a frame is calculated by averaging over MSEs of all the subpictures within that frame. Then the sequence PSNR is calculated as the mean value of PSNRs of individual frames.

$$\boldsymbol{PSNR} = \frac{1}{F}\sum_{f=0}^{n} \mathbf{10} * \log_{10}\left(\frac{MAX^2}{\frac{1}{m}\sum_{i=1}^{m}(\mathrm{MSE}_{i,f})}\right) \quad (1)$$

D. Experimental Results

TABLE III shows the RD performance of each of long IRAP and Mixed IRAP scenarios over short IRAP. As shown in this table, both long IRAP and Mixed IRAP outperform short IRAP in terms of BD-rate bitrate reduction. However, there is an obvious difference between the average gain in S1 and S2 sequences. Considering per sequence results, not only 6k sequences (S2) but also the 10-bit 8k sequences (i.e. SkateboardInLot, and ChairliftRide) show minor gain compared to the rest of the sequences. This is because of the camera motion that these sequences have, which causes most of the CTUs at the end of each GOP could not be predicted from the reference frames, hence they are encoded in Intra mode and consequently consumes more bitrate than expected.

V. CONCLUSIONS

In this paper, we took benefit of the capability of allowing mixed VCL NAL unit types within a coded picture in VVC bitstreams and introduced a viewport adaptive streaming scheme based on mixed IRAP period streaming. In this method viewing orientation changes are handled by switching the stream of the subpictures that are affected by viewing change from long IRAP to short IRAP periods and keep non-affected ones in the bitstreams with the long IRAP period. The method was shown to have significant streaming bitrate reduction of average 19.5 % BD-rate for the test sequences with either no or only slow camera motion while showing a minor gain in other test sequences.

TABLE II. TEST SEQUENCES USED IN THIS STUDY

Class	Sequence name	ORG ERP Resolution@FPS	CMP Resolution	Bit depth
S1	SkateboardInLot	8192×4096@30	6144×4096	10
	ChairLift	8192×4096@30	6144×4096	10
	KiteFlite	8192×4096@30	6144×4096	8
	Harbor	8192×4096@30	6144×4096	8
	Trolley	8192×4096@30	6144×4096	8
	GasLamp	8192×4096@30	6144×4096	8
S2	Balboa	6144×3072@60	4608×3072	8
	Broadway	6144×3072@60	4608×3072	8
	Landing2	6144×3072@60	4608×3072	8
	BranCastle2	6144×3072@60	4608×3072	8

TABLE III. RD PERFORMANCE (BD-RATE) OF LONG AND MIXED IRAP SCENARIOS TEST SEQUENCES USED IN THIS STUDY

		Long IRAP	Mixed IRAP
S1	SkateboardInLot	-0.41%	-0.32%
	ChairliftRide	-5.65%	-5.47%
	KiteFlite	-18.84%	-18.47%
	Harbor	-21.61%	-21.18%
	Trolley	-32.21%	-31.63%
	Gaslamp	-40.63%	-39.86%
S2	Balboa	-0.94%	-0.83%
	Broadway	-1.25%	-2.97%
	Landing2	-0.09%	-0.06%
	BranCastle2	-0.08%	-0.05%
Avg.	S1	**-19.89%**	**-19.49%**
	S2	**-0.59%**	**-0.98%**

REFERENCES

[1] A. Zare, A. Aminlou, M. M. Hannuksela and M. Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," In Proceedings of the 2016 ACM on Multimedia Conference, pp. 601-605, Oct. 2016.

[2] Y. Sánchez, R. Skupin, and T. Schierl, "Compressed domain video processing for tile based panoramic streaming using HEVC, " Proc. of IEEE International Conference on Image Processing (ICIP), pp. 2244-2248, Sep. 2015.

[3] M. M. Hannuksela, Y.-K. Wang and M. Gabbouj, "Isolated regions in video coding", IEEE Transactions on Multimedia, vol. 6, no. 2, pp. 259-267, Apr. 2004.

[4] R. Ghaznavi-Youvalari et al., "Comparison of HEVC coding schemes for tile-based viewport-adaptive streaming of omnidirectional video," IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), 2017.

[5] A. Zare, A. Aminlou, and M. M. Hannuksela, "Virtual reality content streaming: viewport-dependent and tile-based techniques," Proc. of IEEE International Conference on Image Processing (ICIP), Sep. 2017.

[6] K. K. Sreedhar, A. Aminlou, M. M. Hannuksela and M. Gabbouj, "Viewport-Adaptive Encoding and Streaming of 360-Degree Video for Virtual Reality Applications," Proc. of IEEE International Symposium on Multimedia (ISM), 2016. doi: 10.1109/ISM.2016.0126.

[7] ISO/IEC 23090-2:2019, Information technology — Coded representation of immersive media — Part 2: Omnidirectional media format

[8] M. M. Hannuksela, Y.-K. Wang, and A. Hourunranta, "An overview of the OMAF standard for 360° video," Proc. of Data Compression Conference (DCC), Mar. 2019.

[9] ITU-T Recommendation H.266, "Versatile video coding," pre-published Aug. 2020. https://www.itu.int/rec/T-REC-H.266

[10] A. Zare, A. Aminlou, M. M. Hannuksela, and M. Gabbouj, "HEVC-compliant tile-based streaming of panoramic video for virtual reality applications," Proc. of the 24th ACM international conference on Multimedia, 2016. DOI:https://doi.org/10.1145/2964284.2967292

[11] M. M. Hannuksela, "On merging of MCTS of viewport-dependent streaming," Joint Video Experts Team (JVET) document JVET-M0388, Jan. 2019. http://phenix.int-evry.fr/jvet/doc_end_user/documents/13_Marrakech/wg11/JVET-M0388-v1.zip.

[12] Schwarz, H., Schierl, T., and Marpe, D. 2014. Block structures and parallelism features in HEVC. In High Efficiency Video Coding (HEVC): Algorithms and Architectures, Springer, pp. 49-90

[13] X. Corbillon, F. De Simone, and G. Simon, "360-Degree Video Head Movement Dataset," Proc. of the 8th ACM on Multimedia Systems Conference, 2017, pp. 199–204.

[14] F. Duanmu, Y. Mao, S. Liu, S. Srinivasan and Y. Wang, "A Subjective Study of Viewer Navigation Behaviors When Watching 360-Degree Videos on Computers," Proc. of IEEE International Conference on Multimedia and Expo (ICME), 2018. doi: 10.1109/ICME.2018.8486537.

[15] F. Qian, B. Han, Q. Xiao, and V. Gopalakrishnan, "Flare: Practical Viewport-Adaptive 360-Degree Video Streaming for Mobile Devices," Proc. of the 24th Annual International Conference on Mobile Computing and Networking, 2018.

[16] A. Segall, E. François, S. Iwamura, D. Rusanovskyy, " JVET common test conditions and evaluation procedures for HDR/WCG video" , Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, 12th Meeting: Macao, CN, 3–12 Oct. 2018

[17] M. Coban, R. Skupin, “AHG12: Proposed JVET common test conditions and evaluation procedures for MCTS and sub-pictures with boundary padding,” JVET-M0870, Marrakech, MA, Jan. 2019

[18] P. Hanhart, J. Boyce, K. Choi, J.-L. Lin, “JVET common test conditions and evaluation procedures for 360°,” JVET-L1012, Oct. 2018.

[19] 360Lib-10.0. [Online]. Available: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-10.0/

[20] VTM8.0, reference software for VVC. [Online]. Available: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tags/VTM-8.0

[21] A. Hallapuro, M. Homayouni, A.Aminlou, M. M. Hannuksela, "Subpicture merging experiment," Joint Video Experts Team (JVET) document JVET-R0148, April 2020.