Viewport-Aware Omnidirectional Video Streaming Using Visual Attention and Dynamic Tiles

Cagri Ozcinar*, Julián Cabrera†, and Aljosa Smolic*
*V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, Ireland.
†Grupo de Tratamiento de Imágenes, Information Processing and Telecommunications Center
and ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain.

***Abstract*—In this paper, we introduce a new adaptive omnidirectional video (ODV) streaming system that uses *visual attention* (VA) maps, providing enhanced virtual reality (VR) video experiences. Our proposed method benefits from *dynamic tiling* and *viewport-aware* bitrate allocation algorithms. Our main contribution is utilizing the VA maps for deciding the tiling structure (*i.e.,* tile scheme) per chunk and distributing a given bitrate budget to each tile in a viewport-aware way. For this, we first estimate viewport-based VA maps using the collected users' viewport trajectories. Then, an optimal pair of tiling scheme and unequal bitrate allocation for each tile of a given content is determined per chunk by calculating the expected viewport quality using our proposed VA-weighted objective quality measurement (*OmniVA*). We evaluate the proposed method performance with varying bandwidth conditions and viewport trajectories from different users. The results show that the proposed method significantly outperforms the existing tiled-based method in terms of viewport-PSNR.**

***Index Terms*—omnidirectional video, visual attention, tiling, adaptive streaming, virtual reality**

I. Introduction

Significant industrial investment can be observed these days regarding immersive *virtual reality* (VR) applications using *omnidirectional video* (ODV), also known as 360° video. ODV is typically captured by multiple cameras that cover 360° of the scene, rendered through head-mounted displays (HMDs) which allow the viewers to look around a scene from a central point of view in VR. This emerging representation achieves more immersive experience than consuming traditional video sequences.

The delivery of perceptually acceptable quality level, however, is a challenging task for ODV because of the limitations of the present Internet, processing and decoding constraints on the available client devices [1]. Existing HMDs have a viewable field of view and use only a fraction of the given ODV at a given time, namely *viewport*. Therefore, transmission of ultra-high resolution of ODV (*e.g.,* $\geq$ 8K×4K) is needed to obtain a decent VR video quality level. In this context, transmission of a region-of-interest at a time [2]–[4], or viewport-dependent solutions [5]–[7], can reduce the required transmission bandwidth of ODV. However, in scenarios with delay-prone communication pipelines [8] and rapid head orientation activities [9], such solutions are merely inefficient to comply with the motion-to-photon latency requirement, thus penalizing the quality of experience (QoE) [2].

To provide seamless video playback, videos are delivered over the Internet using the MPEG-Dynamic adaptive streaming over HTTP (DASH) [10], [11] standard. In DASH, each video has a set of DASH representations that contain its different *bitrate levels*. Each DASH representation consists of multiple self-decodable time segments, namely chunks, which can be individually requested and decoded by DASH players.

Given its interactive look-around nature with its very large resolution requirement, transmission of ODV demands effective compression and streaming solutions to meet both network and device constraints. To this end, tile-based encoding [12] with the spatial representation description feature of DASH [10] can manage the transmission of high-resolution videos. Tiles are self-decodable spatial regions that allow the client to select which portions have to be extracted from a given bitstream [3], [4], [10], [12], [13]. In addition, tile-based encoding introduces several opportunities for usage in the video. For example, partial decoding [10], [14]–[16], cost-effective video coding [6], [17]–[19], and utilizing *visual attention* (VA) maps [20] in video streaming can be possible. VA maps describe how the users consume a given video at a given time [9].

Although tile-based encoding can provide several new benefits, the selected *tiling scheme* impacts the ODV streaming efficiency [7], [21]. More clearly, the tiling scheme represents the spatial partitioning structure [16] that contains a set of non overlapping tiles. Larger tiles can increase the coding gain for some content by exploiting a large number of redundant pixels, but less flexibility of exploiting the redundant pixels outside the viewport region [7]. In contrast, using small tiles can decrease the coding efficiency because of exploiting fewer spatial redundancies. Therefore, it is necessary to find an appropriate tiled representation for a given ODV; thus, a smart delivery strategy can save network bandwidth and improve the overall QoE [22].

This paper proposes a novel adaptive streaming solution to obtain a decent VR video quality through delay-prone networks. The objective is to provide high video quality by finding the most appropriate tile sizes and target encoding bitrate level for each tile of a given ODV by consulting VA maps. For this, we developed a novel VA-driven quality metric for ODV, namely, *OmniVA*, to determine the optimal tiling scheme for each chunk and to allocate an appropriate target encoding bitrate for each tile from a given bitrate budget. Users' viewport trajectories [23] are transformed to the viewport-based VA maps, and the most appropriate tiling scheme with the required encoding bitrate for each tile is then determined using the developed OmniVA metric on a chunk basis. The proposed design does not require any modification of the existing DASH players, being entirely transparent to them. As such, we expect that our work will provide beneficial input for the streaming industry considering the transmission of dynamic tiled ODV for each chunk and varying bitrate for each tile with the help of VA maps. To verify our method, we recorded viewport trajectories from participants in disjointed subjective viewing sessions using an HMD, estimated the viewport objective quality scores and compared our proposed method with the reference solution, which is based on a naive tiled-based adaptive streaming approach.

The remainder of this paper is organized as follows. In Sec. II, we summarize the related work on adaptive ODV streaming. We then describe the proposed system in Sec. III. We present evaluation results in Sec. IV. Finally, we conclude the paper in Sec. V.

II. Related Work

Recent research works focus on adaptive ODV streaming using fixed-sized tiles. Various practical adaptation strategies, for example, were discussed in [2] using the tile-based encoding. In addition, navigation-aware transmission strategy is proposed in [21], where the viewport quality is enhanced by optimizing the downloading rate of each tile. Visual attention model was consider in [24] to navigate throughout omnidirectional image. Furthermore, Xie *et al.* minimized the total expected distortion of the prefetched tiles using the probabilistic model in [25]. Although their described strategy is ideal for saving bandwidth, it requires extremely low end-to-end delay [2] to predict accurate viewports for each client and content. Hence, the overall QoE might well be deteriorated by the inaccurate predictions.

Viewport-dependent techniques are ideal to save bandwidth in ODV streaming. Two different encoded versions of the same content, for example, were delivered in [6] to reduce the transmission rate of a given ODV. In the work, the tiles that were overlapped with the current viewport were streamed in high resolution while the rest of the tiles were streamed in low resolution. Similarly, a viewport-adaptive video delivery system was developed in [26] that uses tiles and different DASH representations that differ by their bitrate and different scene regions. In addition, an optimal DASH representation for each tile was requested in a viewport aware manner in [7]. Several version of DASH representations were generated for different viewport positions in [27], where the opposite areas of the defined viewport were set to black to reduce the encoding bitrate. Similarly, various viewport-dependent cube map projection-based tiles were prepared in [5], and streamed to clients based on their viewport positions. However, none of the described works consider dynamic tiling and viewport-based VA maps, which are the major contributions of our work.

III. Proposed System

We consider an end-to-end adaptive streaming system for VR to deliver very high resolution of ODVs over the Internet as depicted in Fig. 1. The proposed system enables DASH-VR clients to navigate through a delivered ODV and creates an immersive VR experience with HMDs. The server side of the proposed system contains tiling, encoding, estimation of VA map, optimization of tile scheme, and packing for adaptive streaming.

Each captured ODV is first mapped onto a 2D plane using the projection techniques for backward-compatibility purpose with the existing video coding standards. Because the equirectangular projection (ERP) [28] is the most widely deployed ODV format currently [29], we consider the use of ERP as the input ODV format, which contains full panoramic 360° horizontal and 180° vertical views of the captured scene.

A given ERP content is then divided into a predetermined number of *tiles* which in turn are encoded at various bitrate levels. Each encoded tile is then segmented into chunks and stored in the HTTP server. To this end, a set of several predetermined tiling schemes is available to prepare the DASH representations at each target bitrate budget. For a given tiling scheme, a viewport-based tile bitrate allocation algorithm is applied with the objective of reinforcing the quality of the selected viewport, which is estimated using VA maps. Each VA map is generated using the recorded dataset in [23], which contains viewport trajectories.

Finally, the proposed system selects the optimal tiling schemes for each chunk of a given DASH representation based on the expected viewport quality of a given content using the generated VA maps. The selection of optimal tiling schemes is carried out on a chunk basis. Thus, for each chunk within each representation, a decision about the most appropriate tiling scheme is made. Hence, full 360° tiled chunks are delivered to the clients upon their HTTP requests. This way, the QoE for ODV is not negatively affected by inevitable end-to-end streaming latency.

A. Estimation of visual attention map

Given the collection of viewport trajectories in [23][1], a viewport-based VA map is estimated for each chunk of an ODV. In this work, each VA map serves as an image of the same size of a given content, and its pixel values represent the viewing attention that the clients have paid to the analogous pixels in the ODV. The higher the pixel value is, the more times the pixel at that position in the ODV has been watched. For this, from each recorded viewport trajectory (a pair of u and v texture coordinates), the viewport area is estimated, and an attention mask, $V_{\theta,\Phi}$, is created. In the mask, pixels within the viewport are one and pixels outside are zero. As there is a gradually decreasing visual acuity from the center of the viewport toward edges of the viewport [30], isotropic Gaussian filtering is then applied in the viewport domain to

[1]https://github.com/cozcinar/omniAttention

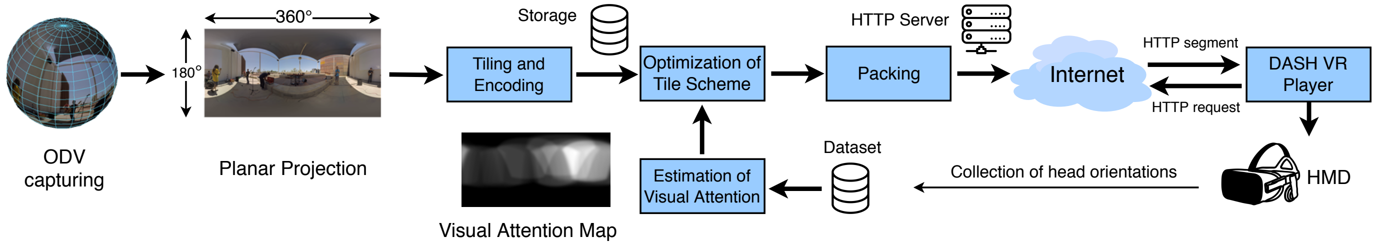


Fig. 1: Schematic diagram of the proposed adaptive ODV streaming system for VR.

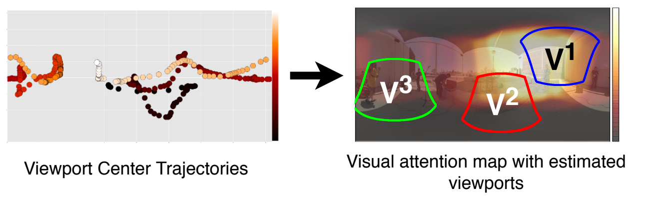


Fig. 2: Estimation of visual attention map and most expected viewports.

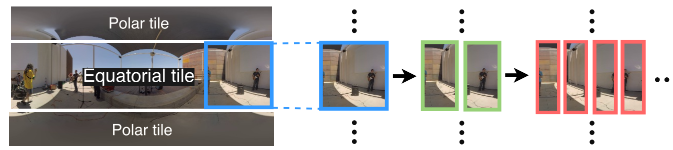


Fig. 3: The used tiling scheme with its different structures.

lower the contribution of the pixels as their distance to the center of the viewport grows.

Given a particular client, $c \in \mathcal{C}$, for the k-th chunk all the generated filtered masks are fused together resulting in the chunk, and viewport-based VA map for c, denoted as $a_{k,c}$. Then, each client contribution is added again resulting in the final viewport-based VA map, A_k:

$$A_k = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} a_{k,c}, \quad (1)$$

where $a_{k,c} = \sum G(V_{\theta,\Phi}, \sigma)$, and $G(V_{\theta,\Phi}, \sigma)$ is the Gaussian filter applied to the attention mask. The filter strength is controlled by the constant of σ, which is chosen as 15 as suggested in [30].

B. Optimization of tiling scheme

The proposed system works by dividing a given ODV into tiles using the tiling scheme, paying attention to the typically lower importance and low-motion characteristics of the poles and the dominant viewing adjacency of the equatorial region [9], [31]. As the poles occupy the largest regions of redundant pixels [32], in those areas, larger tile resolution size is used to compress them using a lower bitrate. In addition, as the equator is associated with the most dominant viewing adjacency, it is further divided horizontally into several tiles in a dynamic way to achieve higher coding gain. Fig. 3 illustrates the used tiling scheme in this paper.

Each given ODV consists of a set of tiles, $\mathcal{T}$, of various sizes and each i-th tile, $t_i \in \mathcal{T}$, has m different target encoding bitrates $\{R^1_{t_i} \dots R^m_{t_i}\}$. Thus, to build the set of DASH representations, an optimal pair of tiling scheme and bitrate allocation for its tiles has to be determined.

A tiling scheme, T_s, can then be defined as a selection of a subset of tiles in $\mathcal{T}$ that cover without overlapping the whole 360° area. We consider that there exists a predefined set of n_t possible tiling schemes, $\mathcal{T}_\mathcal{S}$, that is evaluated in the optimization process. Therefore, let R_r be the target bitrate of the r-th DASH representation. For each k chunk, the optimal tiling scheme $T^*_{s,k}$ can then be calculated as follows:

$$T^*_{s,k} = \max_{T_s \in \mathcal{T}_\mathcal{S}} Q_S(T_s, A_k, B(R_r, .)), \quad (2)$$

where $Q_S(\cdot)$ represents the expected quality of T_s applied to the content, taking into account the VA map A_k and following a bit allocation function, $B(R_r, \cdot)$, subject to a target bitrate R_r.

1) Viewport-based bitrate allocation for tiles: For any tiling scheme, T_s, to be considered in the optimization process, we propose to adopt the bitrate allocation algorithm for tile bitrate management at the client side presented in [7] to the conditions of our approach on the server side. The idea is to distribute the target bitrate R_r among the tiles reinforcing those that are placed within a viewport with high VA values, V^r. Thus, our proposed bit allocation function will depend on the target bitrate and the selected viewport: $B(R_r, V^r)$.

Let $\mathcal{S}^{in}$ and $\mathcal{S}^{out}$ be the sets of tiles inside the viewport or overlapping with it, and outside the viewport respectively. The bitrate assigned to the i-th tile in $\mathcal{S}^{in}$ is the following:

$$\widetilde{R_{t_i}} = (\gamma R_r)\omega_i \quad t_i \in \mathcal{S}^{in}, \quad (3)$$

where γ defines the percentage of the bitrate that is assigned to the tiles in $\mathcal{S}^{in}$ and ω_i is the weight of the i-th tile. In this work, γ is selected as 0.8 empirically, and ω_i is calculated as:

$$\omega_i = \frac{\#\text{ of pixels in } (V^r \cap \mathcal{S}^{in})}{\#\text{ of pixels in } V^r}. \quad (4)$$

To gradually distribute the remaining bitrate among the outside-viewport tiles, the Euclidean distance, δ_i, is calculated between the center of V^r, and the center of each tile in $\mathcal{S}^{out}$. The bitrate estimation for the i-th outside tile is then estimated being inversely proportional to δ_i:

$$\widetilde{R_{t_i}} = \widehat{\kappa_i}((1-\gamma)R_r) \quad t_i \in \mathcal{S}^{out}, \quad (5)$$

where $\widehat{\kappa_i}$ is calculated as $\widehat{\kappa_i} = \frac{\kappa_i}{\Sigma_i \kappa_i}$ and $\kappa_i = \frac{\max\{\delta_i\}}{\delta_i}$.

Finally, to accommodate the DASH representations, from the set of encoded versions of each tile, the ones with the bitrate closest to the computed values are selected, on the condition that the total bitrate is not higher than the target bitrate, R_r.

2) Estimation of the quality of a tiling scheme: For a given chunk, a choice of tiling scheme and viewport input for the bitrate allocation algorithm, (T_s, V^r) is used. Therefore, the bitrate allocation of the tiling scheme will reinforce the quality of the input viewport at the expense of the rest of the ODV area. However, during the duration of a chunk, the user may look at different parts of the 360° scene. Thus, to measure the performance of a given pair (T_s, V^r) considering this fact, we propose to compute the tiling scheme objective quality PSNR in a number of nonoverlapping viewports. These viewports will be those most-probably watched according to the chunk VA map. Fig. 2 illustrates a VA map with its most-likely selected viewport. We refer to our proposed quality measurement as Omnidirectional VA-weighted PSNR (OmniVA-PSNR).

Let $\{V_i\}$ represent the set of the N_V most-probably watched and nonoverlapping viewports according to the visual map. Thus, the quality of a pair (T_s, V^r), can be estimated as:

$$Q_S(T_s, A_k, B(V^r, R_r)) = \sum_{i=1}^{N_v} P_{A_k}(V_i) Q_m(V_i, B(V^r, R_r)) + P_{A_k}(O) Q_m(O, B(V^r, R_r)) \quad (6)$$

where $P_{A_k}(.)$ is the probability that a viewport be watched according to the VA map, $Q_m(.)$ represents the PSNR value of the encoded representation of the viewport according to the tiling scheme and bitrate allocation done, and O represents the remainder of the ODV area that is not covered by any $\{V_i\}$.

IV. Experiments

A. Setup

We use the following two ODVs (8K×4K ERP) from the JVET and MPEG video coding exploration experiments: $\mathcal{V}$ = $\{$*Train*, *Basketball*$\}$ [33]. The HEVC standard [34] was utilized to encode each tile of a given ODV. For this, we used the FFmpeg software (*ver.* N-85291) [35] for encoding purposes with two-pass and 200 percent constrained variable bitrate configurations. In addition, we used a set of *target bitrates* $\mathcal{B}$ = {1, 1.20, 1.44, 1.73, 2.07, 2.49, 2.10, 3.58, 4.30, 5.16, 6.19, 7.43, 8.92, 10.70, 12.84, 15.41, 18.49, 22.19, 26.62, 31.95, 38.34, 46.01} (in terms of *Mbps*) to encode each ODV content. To enable selectively choosing various tile sizes and bitrate level combinations per chunk, each ODV frame was divided into N tiles, encoded to be independently decodable. Each bitstream was divided into 2 *sec.* chunks. Two tiles were used for the poles, and $N-2$ tiles were used for the equator. We examined our *proposed method* with the reference naive tiled-based adaptive streaming approach that used N fixed-sized tiles, namely *fixed-sized N tiles*. Therefore, the encoded bitrate for each tile of a given video is equally distributed for the reference methods by dividing the *target bitrate* to a given N tiles.

To estimate VA maps and viewing trajectories for viewport-based quality measurements, the previously developed test-beds [9] were extended to gather subjective user data for the given videos. We organized subjective experiments under task-free conditions. A total of 17 participants (13 males and four females) took part in the test. Participants were split into two groups for (*i*) modeling of VA data and (*ii*) validation of the proposed approach, consisting of 12 and 5 participants, respectively. In our subjective tests, we used the Oculus Rift consumer version [36] as an HMD and Firefox Nightly as a web browser. Each participant was seated in a rotatable chair and allowed to turn freely.

B. Performance evaluation

To verify and assess the expected coding gains, we have compared our *proposed* method with fixed-sized tiling schemes with the number of tiles N=$\{6, 10, 18\}$. We computed rate-distortion curves for all the schemes using our OmniVA-PSNR measurement at the bitrates of the target DASH representations $\{2, 5, 10, 15, 20\}$ *Mbps*. Fig 4 (a,b) shows how our approach outperforms by a significant margin the fixed tiled schemes. As was expected, our dynamic tiling scheme is able to reinforce those areas of the ODV scene that are more likely to be watched.

Sequence	**Fixed-size tiles**		
	$N=6$	$N=10$	$N=18$
Train	1.06	0.93	1.10
Basketball	0.69	0.92	1.19

TABLE I: Quality gain of the proposed method in terms of BD quality (*dB*) saving.

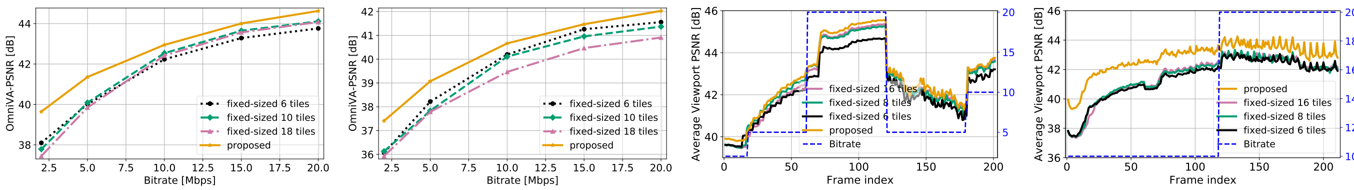
Moreover, this gain has been characterized in terms of the Bjøntegaard metric [37] in Table I. It can be observed that for both sequences, quality gains ranging from 0.69 *dB* to 1.19 *dB* have been obtained with our approach.

Sequence	**Number of tiles**		
	N=6	N=10	N=18
Train	75%	25%	0%
Basketball	40%	60%	0%

TABLE II: Selected tiling schemes (in terms of %) for the proposed method.

Table II shows the distribution of the decisions made by the algorithm. It can be observed that the tiling scheme choices are different according to the VA maps of each sequence. In addition, it is also worth noting that for these two test sequences, the $N=18$ tiling scheme was not suitable due to the characteristics of their VA maps.

As a further assessment, we computed the PSNR of the actual viewports observed by the users at each frame of the sequence. Here, we simulate the varying bandwidth, which is shown in blue dashed line. Results have been averaged for the trajectories of the five users left for validation. Fig. 4 (c,d)



(a) *Train* (b) *Basketball* (c) *Train* (d) *Basketball*

Fig. 4: (a,b) Performance comparison using the RD curves computed with the average OmniVA-PSNR. (c,d) viewport-PSNR quality over frame between the proposed and the reference methods on *varying bandwidth (Mbps)*.

shows how our approach is able to optimize the DASH representations based on the VA map of the sequence. As can be observed, for almost all the frames of the sequences and given different bitrates, the quality of the viewports that the user is actually watching is higher than that of the fixed schemes.

V. Conclusion

This paper introduced a new adaptive omnidirectional video (ODV) streaming system, utilizing visual attention (VA) maps. Our proposed method does not need any modification at the client side, being entirely transparent to the existing DASH players. The developed system aimed at an enhanced quality of ODV streaming viewed in head-mounted displays. For this, the proposed method utilized dynamic-sized tiles per chunk and varying bitrate allocation per tile by consulting the estimated VA maps. The performance of the proposed method was verified in experimental evaluations. The results showed that our proposed method achieves significant quality gain compared with the fixed-sized tiling methods which are a naive approach and used by most existing tiling based ODV streaming solutions. Experimental results showed the effectiveness of the proposed method in adaptive ODV streaming. As future work, we plan to extend the proposed system by considering additional tile schemes and investigating the effect of viewport quality using comprehensive user datasets and ODV sequences.

Acknowledgment

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776. This work has also been partially supported by the Ministerio de Economía, Industria y Competitividad (AEI/FEDER) of the Spanish Government under project TEC2016-75981 (IVME).

References

[1] D. Podborski, E. Thomas, M. M. Hannuksela, S. Oh, T. Stockhammer, and S. Pham, "Virtual reality and dash," in *International Broadcasting Convention, Ibc*, 2017.

[2] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over http: Design, implementation, and evaluation," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, ser. MMSys'17. New York, NY, USA: ACM, 2017, pp. 261–271. [Online]. Available: http://doi.acm.org/10.1145/3083187.3084016

[3] S. Heymann, A. Smolic, K. Mueller, Y. Guo, J. Rurainsky, P. Eisert, and T. Wiegand, "Representation, coding and interactive rendering of high-resolution panoramic images and video using MPEG-4," in *Panoramic Photogrammetry Workshop*, Berlin, Germany, Feb. 2005, pp. 24–25.

[4] C. Grunheit, A. Smolic, and T. Wiegand, "Efficient representation and interactive streaming of high-resolution panoramic views," in *2002 International Conference on Image Processing (ICIP)*, vol. 3, Sep. 2002, pp. III–209–III–212 vol.3.

[5] R. Skupin, Y. Sanchez, C. Hellge, and T. Schierl, "Tile based HEVC video for head mounted displays," in *IEEE International Symposium on Multimedia (ISM)*, San Jose, CA, USA, Dec 2016, accessed: 2017-1-16.

[6] 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*, ser. MM '16. New York, NY, USA: ACM, 2016, pp. 601–605. [Online]. Available: http://doi.acm.org/10.1145/2964284.2967292

[7] C. Ozcinar, A. De Abreu, and A. Smolic, "Viewport-aware adaptive 360 video streaming using tiles for virtual reality," in *2017 International Conference on Image Processing (ICIP)*, Sep 2017.

[8] G. Cheung, Z. Liu, Z. Ma, and J. Z. G. Tan, "Multi-Stream switching for interactive virtual reality video streaming," *arXiv:1703.09090*, 27 Mar. 2017.

[9] A. De Abreu, C. Ozcinar, and A. Smolic, "Look around you: Saliency maps for omnidirectional images in VR applications," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017, pp. 1–6.

[10] O. A. Niamut, E. Thomas, L. D'Acunto, C. Concolato, F. Denoual, and S. Y. Lim, "MPEG DASH SRD: Spatial relationship description," in *7th International Conference on Multimedia Systems*, ser. MMSys '16. ACM, 2016, pp. 5:1–5:8.

[11] ISO/IEC 23009-1, "Information technology — dynamic adaptive streaming over HTTP (DASH) — part 1: Media presentation description and segment formats," ISO/IEC JTC1/SC29/WG11, Tech. Rep., 2014.

[12] K. Misra, A. Segall, M. Horowitz, S. Xu, A. Fuldseth, and M. Zhou, "An Overview of Tiles in HEVC," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 969–977, Dec 2013.

[13] A. Smolic and P. Kauff, "Interactive 3-D video representation and coding technologies," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 98–110, Jan 2005.

[14] P. Lungaro, R. Sjöberg, A. J. F. Valero, A. Mittal, and K. Tollmar, "Gaze-aware streaming solutions for the next generation of mobile +VR experiences," *IEEE transactions on visualization and computer graphics*, vol. PP, no. 99, pp. 1–1, 2018.

[15] A. Patney, M. Salvi, J. Kim, A. Kaplanyan, C. Wyman, N. Benty, D. Luebke, and A. Lefohn, "Towards foveated rendering for gaze-tracked virtual reality," *ACM transactions on graphics*, vol. 35, no. 6, p. 179, 11 Nov. 2016.

[16] C. Concolato, J. L. Feuvre, F. Denoual, E. Nassor, N. Ouedraogo, and J. Taquet, "Adaptive streaming of hevc tiled videos using mpeg-dash," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2017.

[17] C. Ozcinar, A. De Abreu, S. Knorr, and A. Smolic, "Estimation of optimal encoding ladders for tiled 360 VR video in adaptive streaming systems," in *The 19th IEEE International Symposium on Multimedia (ISM 2017)*, Taichung, Taiwan, Nov. 2017.

[18] Y. Sanchez, R. Skupin, and T. Schierl, "Compressed domain video processing for tile based panoramic streaming using hevc," in *2015 IEEE*

International Conference on Image Processing (ICIP), Sept 2015, pp. 2244–2248.
[19] M. Meddeb, M. Cagnazzo, and B. Pesquet-Popescu, "ROI-based rate control using tiles for an HEVC encoded video stream over a lossy network," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 1389–1393.
[20] Y. Rai, P. L. Callet, and G. Cheung, "Role of hevc coding artifacts on gaze prediction in interactive video streaming systems," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sept 2016, pp. 3344–3348.
[21] S. Rossi and L. Toni, "Navigation-Aware adaptive streaming strategies for omnidirectional video," in *IEEE International Workshop on Multimedia Signal Processing (MMSP 2017)*, 2017.
[22] M. N. Garcia, F. D. Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnstrm, and A. Raake, "Quality of experience and http adaptive streaming: A review of subjective studies," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, Sept 2014, pp. 141–146.
[23] C. Ozcinar and A. Smolic, "Visual attention in omnidirectional video for virtual reality applications," in *10th International Conference on Quality of Multimedia Experience (QoMEX 2018)*, 2018.
[24] T. Maugey, O. Le Meur, and Z. Liu, "Saliency-based navigation in omnidirectional image," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct 2017, pp. 1–6.
[25] L. Xie, Z. Xu, Y. Ban, X. Zhang, and Z. Guo, "360probdash: Improving qoe of 360 video streaming using tile-based http adaptive streaming," in *Proceedings of the 2017 ACM on Multimedia Conference*, ser. MM '17. New York, NY, USA: ACM, 2017, pp. 315–323. [Online]. Available: http://doi.acm.org/10.1145/3123266.3123291
[26] X. Corbillon, A. Devlic, G. Simon, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," *arXiv:cs.MM 1609.08042*, vol. cs.MM, no. 1609.08042, pp. 1–7, May. 2017.
[27] C. Diaz, J. Cabrera, M. Orduna, L. Munoz, P. Perez, J. Ruiz, and N. Garcia, "Viability analysis of content preparation configurations to deliver 360vr video via mpeg-dash technology," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, Jan 2018.
[28] J. P. Snyder, *Flattening the Earth: Two Thousand Years of Map Projections*. University of Chicago Press, 1993.
[29] B. Choi, Y.-K. Wang, M. M. Hannuksela, Y. Lim, and A. Murtaza, "Study of ISO/IEC DIS 23000-20 omnidirectional media format," ISO/IEC JTC1/SC29/WG11, Torino, Italy, Tech. Rep. N16950, Jul 2017.
[30] E. Upenik and T. Ebrahimi, "A simple method to obtain visual attention data in head mounted virtual reality," in *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2017, pp. 73–78.
[31] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *2015 IEEE International Symposium on Mixed and Augmented Reality*, 2015, pp. 31–36.
[32] M. Budagavi, J. Furton, G. Jin, A. Saxena, J. Wilkinson, and A. Dickerson, "360 degrees video coding using region adaptive smoothing," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 750–754.
[33] A. Abbas and B. Adsumilli, "Ahg8: New gopro test sequences for virtual reality video coding," JTC1/SC29/WG11, ISO/IEC, Chengdu, China, Tech. Rep. JVET-D0026, Oct 2016.
[34] J.-R. Ohm and G. Sullivan, "Vision, applications and requirements for high efficiency video coding (HEVC)," ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rep. MPEG2011/N11891, March 2011.
[35] "x265 HEVC Encoder / H.265 Video Codec," http://x265.org/, Jan 2018.
[36] O. VR, "Oculus rift," http://www.oculusvr.com, Jan 2018.
[37] G. Bjøtegaard, "Calculation of average PSNR differences between RD-curves (vceg-m33)," VCEG Meeting (ITU-T SG16 Q.6), Austin, Texas, USA,, Tech. Rep. M16090, Apr 2001.