

Viewport-Driven Rate-Distortion Optimized 360° Video Streaming

Jacob Chakareski, Ridvan Aksu, Xavier Corbillon, Gwendal Simon, and Viswanathan Swaminathan

Abstract—The growing popularity of virtual and augmented reality communications and 360° video streaming is moving video communication systems into much more dynamic and resource-limited operating settings. The enormous data volume of 360° videos requires an efficient use of network bandwidth to maintain the desired quality of experience for the end user. To this end, we propose a framework for viewport-driven rate-distortion optimized 360° video streaming that integrates the user view navigation pattern and the spatiotemporal rate-distortion characteristics of the 360° video content to maximize the delivered user quality of experience for the given network/system resources. The framework comprises a methodology for constructing dynamic heat maps that capture the likelihood of navigating different spatial segments of a 360° video over time by the user, an analysis and characterization of its spatiotemporal rate-distortion characteristics that leverage preprocessed spatial tilling of the 360° view sphere, and an optimization problem formulation that characterizes the delivered user quality of experience given the user navigation patterns, 360° video encoding decisions, and the available system/network resources. Our experimental results demonstrate the advantages of our framework over the conventional approach of streaming a monolithic uniformly-encoded 360° video and a state-of-the-art reference method. Considerable video quality gains of 4 - 5 dB are demonstrated in the case of two popular 4K 360° videos.

I. INTRODUCTION

Emerging virtual and augmented reality (VR/AR) technologies are helping introduce novel immersive digital experiences. It is anticipated that VR/AR technologies will represent a \$108 billion market in the near future [1]. Gaming, entertainment, education and training, and 360° video are the main application domains of VR/AR technologies at present, with a broader set of societal applications spanning remote sensing, the environmental and weather sciences, disaster relief, and transportation anticipated in the future [2].

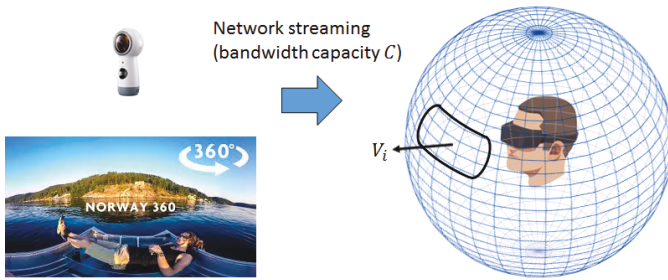


Fig. 1: 360° streaming: Viewport V_i on the 360° sphere.

360° video is an emerging video format captured by an omnidirectional camera that records incoming light rays from every direction. It enables a 360° look-around of the surrounding scene for a remote user, virtually placed at the camera location, on his VR device, as illustrated in Figure 1. Presently, the entire monolithic 360° view panorama

is streamed to the user, who, however, can only experience a small portion of it denoted as viewport V_i , at any time, as also illustrated in Figure 1. However, this results in a huge network overhead/bottleneck and unnecessary computational/bandwidth loading of the device, which, in turn, considerably penalize the user quality of experience. Moreover, to apply conventional video coding, the 360° view sphere is first mapped to a planar shape: equirectangle, pyramid, cube, or dodecahedron. The latter three have been considered since around 30% pixel replication is introduced when the sphere is mapped to an equirectangle [3, 4]. However, they have their own deficiencies, e.g., introduction of projection distortions around the planar shape's edges. Here, we only consider the equirectangular mapping, as the most widely used.

The growing popularity of VR/AR technologies stimulates an equivalent increasing demand for 360° video content, which today can be accessed through over-the-top online providers such as YouTube/Facebook 360 [5, 6]. However, present 360° streaming practices necessitate excessive data rates that even anticipated broadband network access technologies would not be able to support [7, 8], due to the heuristic design shortcomings of the former outlined above. Thus, a broader/faster adoption of emerging 360° technologies that can eventually dominate the market is precluded. On the other hand, delivering the entire 360° view sphere is necessary to avoid simulator/motion sickness [9] that would degrade the quality of experience, as the *intuitive approach of sending only V_i* using traditional server-client delivery architectures, where the server responds to client updates, would preclude application interactivity, due to the inherent network round-trip latency.

To overcome this apparent impasse between 360° application requirements and technology capabilities/design, which essentially stems from the direct application to the 360° domain of existing video coding/streaming technologies that treat 360° content as conventional videos, recent studies have considered uneven spatial quality encoding of 360° videos, to minimize the data rate assigned to 360° regions not navigated by the user presently, thereby considerably reducing the induced network overhead. This is the strategy we also follow, making the following contributions in this context.

We formulate a framework for viewport-driven rate optimized 360° video streaming that integrates the user view navigation pattern and the spatiotemporal rate-distortion characteristics of the 360° video content to maximize the delivered user quality of experience for the given network/system resources. It comprises (i) a methodology for constructing dynamic heat maps that capture the user likelihood of navigating different spatial segments of a 360° video over time, (ii) analysis and characterization of its spatiotemporal rate-distortion characteristics that leverage a preprocessed spatial tilling of the 360° view sphere, and (iii) an optimization problem formulation that

characterizes the delivered user quality of experience given the user navigation patterns, 360° video encoding decisions, and the available system/network resources.

The rest of the paper is organized as follows. In Section II, we first review related work. Subsequently, we present the building components of our system framework in Section III. The problem formulation that aims to maximize the delivered 360° user quality of experience given the user navigation patterns, 360° video encoding decisions, and the available system/network resources, is presented in Section IV. Experimental analysis of the performance of our framework and validation of our system models is carried out in Section V. Finally, concluding remarks and a summary of envisioned future work are provided in Section VI.

II. RELATED WORK

Due to the emerging nature of 360° technologies, only a small body of related work has appeared to date. The study in [10] carried out an empirical characterization of 360° videos highlighting their main features, e.g., their lower temporal rate variability compared to conventional videos. A small number of studies have considered splitting the 360° video into spatial tiles as part of the encoding process, leveraging the tiling feature of the latest High Efficiency Video Coding (HEVC) standard [11]. The encoding data rate of each tile can then be controlled independently to reduce the overall bandwidth usage [12–15]. Transmission aspects of 360° HTTP streaming have been explored in [16, 17]. Applications of scalable video coding to 360° streaming have been studied in [18, 19]. Head movement prediction and the impact of navigation uncertainty have been investigated in [3, 14]. The former study also carries out an empirical analysis of the performance efficiency of the four sphere-to-planar shape projections and investigates the benefits of streaming adaptively multiple 360° representations featuring different quality-emphasized spatial regions [3].

In contrast to the few studies cited above that consider HEVC 360° tiling, we employ preprocessed spatial tiles of the 360° view panorama, which has several advantages in the form of lower complexity at multiple critical aspects of a server-client 360° streaming architecture [3]. Moreover, a formal analysis of the spatiotemporal rate-distortion characteristics of 360° tiling that integrates the user navigation patterns and the available network/system resources has not been carried out towards optimal selection of 360° encoding and streaming decisions. The framework of our paper aims to fill this gap.

III. SYSTEM MODELS

A. Overview

Our 360° networked system architecture comprises several major component blocs that were introduced earlier and is illustrated in Figure 2. We describe each one in detail here.

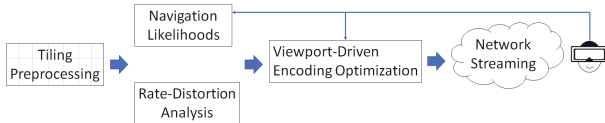


Fig. 2: 360° networked system architecture.

B. Tiling preprocessing

We partition a 360° video into a set of $N \times M$ spatial tiles. In particular, we partition the raw 360° video frames into spatial tiles and consider the collection of thereby constructed (smaller) video frames for each tile as separate videos. The tiles are then separately encoded and streamed to the user, according to our analysis and optimization. As explained earlier, carrying out tiling as a preprocessing step has several advantages over tiling the video as part of the encoding process, as enabled by the tiling feature of the latest video coding standard HEVC. In our experiments, we used two popular 4K 360° videos [20, 21] that we preprocessed into 6×4 spatial times, as illustrated in Figure 3, where the first and second dimension refer to the horizontal and vertical pixel resolution of the video. Each tile is indexed in a raster fashion, top-to-bottom and left-to-right.



Fig. 3: 360° video panorama 6×4 spatial tiling.

We selected this specific tiling based on empirical analysis, as a reasonable choice between the complexity and compression efficiency introduced by a given tiling. In a follow-up study, we plan to integrate the selection of tiling as part of the 360° end-to-end analysis and optimization.

C. 360° VR head movement data

We collected head-movement data that describes how a user navigates a 360° video over time. In particular, a VR device outputs the direction of the current viewpoint of the user V_i on the 360° view sphere up to 250 times per second, with the user considered to be placed at the sphere center, as described earlier. Precisely, this is the surface normal of V_i on the 360° sphere that is uniquely described by the spherical coordinates azimuth and polar angles $\varphi \in [0^\circ, 360^\circ]$ and $\theta \in [0^\circ, 180^\circ]$ it spans on the sphere, in a spherical coordinate system with the 360° sphere center as its origin, as illustrated in Figure 4 (right). These two angles are equivalently denoted as yaw and pitch in the VR community, captured as rotation angles around the Z and Y axes, as denoted in Figure 4 (left). We collected the pairs (φ_j, θ_j) that coincided with the discrete temporal instances t_j of subsequent 360° video frames j displayed to the user as he navigates the content. They are the navigation data points relevant for our analysis.

D. Navigation likelihoods

For various head-mounted displays (HMD) used in VR applications, the viewport size experienced by the user varies.

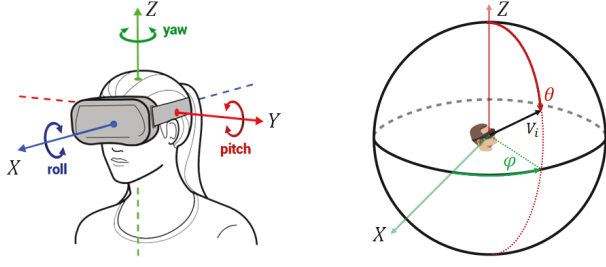


Fig. 4: 360° head movement navigation data of current viewport V_i . Left: Rotation angles yaw, pitch, and roll around the three coordinate axis. Right: Azimuthal and polar angles (φ, θ) in spherical coordinates.

In this paper, we assume a viewport of 110° horizontal and 90° vertical fields of view. For every navigation trace for a given 360° video, we compute the fraction of the surface area of tile k occupied by the user viewport V_i at time instance j , denoted as $w_{k,j}$. To account for the unequal surface area occupied by different viewports, when mapped to a 2D rectangle used to encode the data, depending on their latitude (polar angle θ) on the 360° view sphere, each tile k is assigned a normalized weight $\bar{w}_{k,j}$, computed as $\bar{w}_{k,j} = w_{k,j} / \sum_k w_{k,j}$. We can then aggregate these weights over different time durations, to compute the likelihoods of navigating different tiles of the respective 360° video during those time periods. In our analysis, we are interested in exploiting these navigation likelihoods over the duration of individual Groups Of Pictures (GOPs) comprising the encoded 360° content.

For illustration, Figure 5 shows the average (over the entire video) navigation likelihoods of different tiles comprising the selected 6×4 tiling applied to the 360° video Roller Coaster used in our experiments. We can see that corner tiles appear rarely in a viewport navigated by the user, as their navigation likelihoods are close to zero. Conversely, it appears that the user often navigated through tiles 9, 10, 15, and 16, for instance, as they have much higher navigation likelihoods.

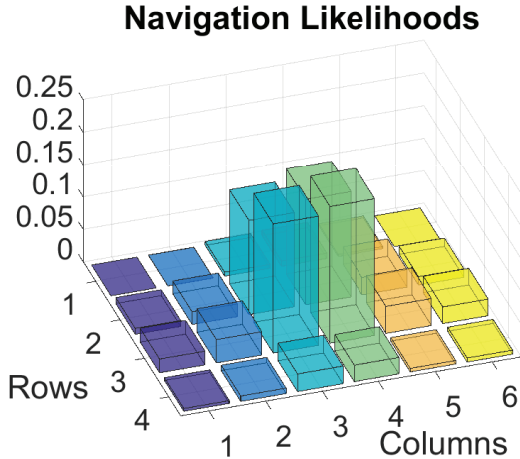


Fig. 5: Navigation likelihoods of tiles for Roller Coaster.

Figure 6 shows the corresponding tile navigation likelihoods for the second 360° video, Wingsuit, used in our experiments. It appears that in this case the viewport navigated by the user is mostly closer to the south pole, as the corresponding tiles have much higher likelihoods now, due to the specific nature of

this video (more interesting content is spatially located there).

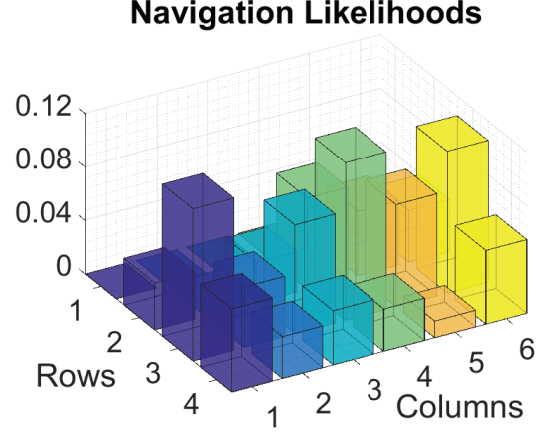


Fig. 6: Navigation likelihoods of tiles for Wingsuit.

A visualization of two representative viewports is shown in Figures 7 and 8. Since mapping a 3D shape (sphere) to 2D causes distortion, the shape of a viewport also changes. In equatorial regions, a viewport is smaller and more compact (Figure 7) while in polar regions a viewport is spread over all polar tiles (Figure 8). Figure 3 can be referenced to understand the spatial locations of these two viewports relative to the underlying tiling of the respective 360° video.

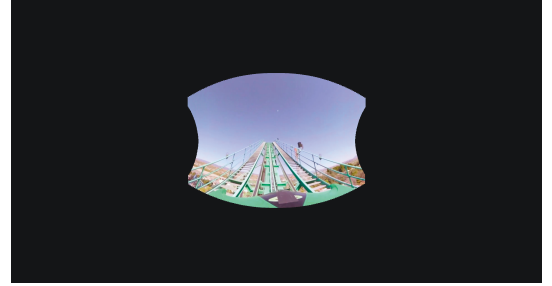


Fig. 7: Viewport at $(\varphi, \theta) = (0^\circ, 0^\circ)$.



Fig. 8: Viewport at $(\varphi, \theta) = (120^\circ, -60^\circ)$.

E. Rate-Distortion models

Changing the quality of tiles is a useful method to control the bitrate of a 360° video. The quantization parameter QP employed by the HEVC (H.265) codec is a convenient tool for tile quality adaptation. We explore two prospective characterizations of the dependency between the parameter QP and the resulting bitrate R of the encoded tile. That is, we

investigate modeling $R = f(\text{QP})$ via an exponential or power law function for f as follows

$$R = a_1 e^{-b_1 \text{QP}} \quad \text{or} \quad R = a_2 \text{QP}^{b_2}. \quad (1)$$

We will validate these relationships by comparing the bitrate and QP for an encoded 360° tile in Section V-B. Since we have a function between the bitrate and QP, we can define bounds for our optimization problem with the highest and lowest QP values that can be selected. And after calculating the optimal bandwidth, going back to QP value and encoding the tiles accordingly can be done easily in the server side.

Similarly, we investigate two prospective characterizations of the dependency between the encoded tile bitrate R and the induced reconstruction error or distortion D for a tile, where the latter can be calculated as the mean-square error (MSE) between the encoded tile video data and the corresponding raw video data for the tile. In essence, the distortion D captures the average deviation of encoded tile pixels from their raw data counterparts. In a raw 360° YUV 4:2:0 video, for every pixel sample of the color (chrominance) components U and V there are 4 pixel samples of the (monochromatic) intensity (luminance) component Y. Thus, the luminance distortion dominates the encoding distortion for the two color components. Therefore, we used the luminance component distortion as the representative of the encoding distortion D for a tile, measured for every 360° tile luminance video frame. We investigate modeling the dependency $D = f(R)$ via an exponential or power law function for f as follows

$$D = c_1 e^{-d_1 R} \quad \text{or} \quad D = c_2 R^{d_2}. \quad (2)$$

We validate these relationships by comparing the encoding bitrate and distortion for a encoded 360° tile in Section V-B. The characterizations $R = f_1(\text{QP})$ and $D = f_2(R)$ will allow us to formulate the aggregate 360° video encoding quality and pursue related optimizations, as explained next.

IV. OPTIMIZATION FRAMEWORK

Given the analytical modeling of the relevant problem variables, we now set out to find the optimal bitrate for each tile. There are constraints that we integrate into the problem formulation. These are the aggregate available network bandwidth C and the allowed QP range per tile.

A. Problem Setup

Given the limited network bandwidth, tiles should be transmitted at data rates according to their navigation likelihoods and rate-distortion characteristics such that we can maximize the delivered aggregate quality of the respective 360° video. Let $R_i(\text{QP}_i)$ denote the bitrate of the i^{th} tile where QP is the encoding quantization parameter, as introduced earlier. This gives us the following inequality to maintain:

$$\sum_i R_i(\text{QP}_i) \leq C, \quad i = 1, \dots, M \times N. \quad (3)$$

For practical reasons, for every tile i we set a range of QP values that can be considered, defined by the upper and lower

bounds $\text{QP}_{\min}$ and $\text{QP}_{\max}$. This therefore induces constraints on the minimum and maximum data rates that can be assigned to a tile, given the monotonic relationship between QP and R_i , as captured by the function $R_i(\text{QP})$. Formally, these two constraints can be written as

$$R_i(\text{QP}_{\max}) \leq R_i(\text{QP}_i) \leq R_i(\text{QP}_{\min}). \quad (4)$$

Finally, we formulate the expected 360° quality of experience that a user observes while navigating the scene, as the navigation likelihood weighted sum of video qualities of all tiles comprising the 360° video content streamed to the user. This can be formally written as $\sum_i p(i|v) D_i(R_i)$, where $p(i|v)$ denotes the navigation likelihood of tile i given that viewport v is requested initially. To be precise, note that we formulated our objective as the expected 360° video distortion, due to the one-to-one correspondence between video quality and reconstruction error (distortion). Therefore, we aim to minimize our objective function, as it will lead to the same goal (maximum 360° quality of experience).

B. Optimization Formulation

Leveraging the problem setup described earlier, we can now formulate the optimization problem of interest as

$$\begin{aligned} \min_{\{R_i\}} \quad & \sum_i p(i|v) D_i(R_i), \\ \text{subject to:} \quad & \sum_i R_i(\text{QP}_i) \leq C, \quad i = 1, \dots, M \times N, \\ & R_i(\text{QP}_{\max}) \leq R_i(\text{QP}_i) \leq R_i(\text{QP}_{\min}), \quad \forall i. \end{aligned} \quad (5)$$

Note that (5) represents a convex optimization problem, due to the nature of the constraints involved and the objective function under consideration. Therefore, it can be efficiently solved using fast convex optimization methods [22]. In our experiments, we carry out the optimization in (5) for every GOP, facilitating the dynamic weight assignment described in Section III-D to compute the navigation likelihoods $p(i|v)$. In particular, after the optimization completes, the QP vs. R dependency for each tile i in a GOP is used to obtain the explicit optimal QP_i^* value that corresponds to the optimal data rate R_i^* produced by (5). Note that for illustration we included the average navigation likelihoods $p(i|v)$ across the applied 360° video tiling for the duration of the entire video in Figures 5 and 6, for the 360° content used in our experiments.

We recall that the analytical dependencies between R and D , and between R and QP are not explicitly denoted in (5). As explained earlier, we explore two models for each dependency $D = f(R)$ and $R = f(\text{QP})$, an exponential one and a power-law one. And the parameters that comprise each model are extracted uniquely for each tile, before we carry out the optimization in (5). In our experiments, we first validate each of these models, for each dependency, and select the one that is more accurate, to carry out the remaining performance evaluation analysis.

V. EXPERIMENTATION

A. System Setup

We used two popular 4K 360° videos from Youtube, Roller Coaster [20] and Wingsuit [21], to evaluate the performance of our framework. VR users watched these videos with HMD devices and their head movements have been tracked using the OpenTrack software [23]. For performance evaluation, we have used one head movement trace per user per video. Based on the collected traces, the navigation likelihoods of each tile in a GOP are calculated, as discussed in Section III-D.

Each 360° video is preprocessed into 6×4 tiles, as explained in Section III-B. Each tile is encoded into Groups of Pictures (GOPs) of size 32 frames using HEVC. There are 60 GOPs in each tile video, corresponding to 1920 frames and 64 seconds of duration of time, assuming a frame rate of 30 fps. Each GOP is encoded using 5 QP values (22, 27, 32, 37, 42). Using the encoded tiles of those QP values, $R-D$ and $QP-R$ parameters are extracted for all tiles and GOPs, to explore the proposed rate-distortion modeling from Section III-E.

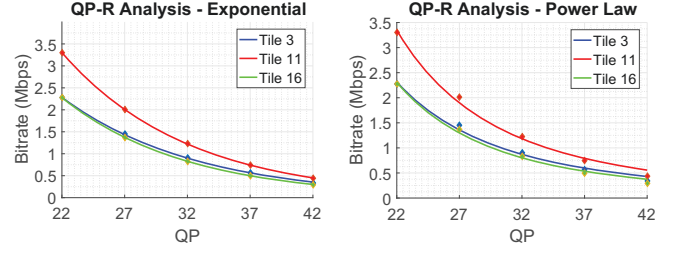
Two reference methods are examined to compare against. First, an entire monolithic 360° video is encoded using the following 5 QP values (32, 34, 36, 39, 42). In each case, the induced average data rates for every GOP are used as the network bandwidth constraint C in our own optimization in Section IV-A. Similarly, we also implemented a state-of-the-art method proposed by Petrangeli et al. [13]. It predicts future viewports accessed by the user, based on the speed and the position of the HMD. Tiles within the current/future predicted viewports in a GOP are encoded with the highest possible QP value. The remaining tiles are encoded with the lowest possible QP value given the remaining bandwidth budget.

B. Rate-distortion model validation

We formulated two prospective models for the dependencies $D = f(R)$ and $R = f(QP)$, described in Section III-E. Here, we explore their accuracy in characterizing the encoded 360° video content we considered in our experiments.

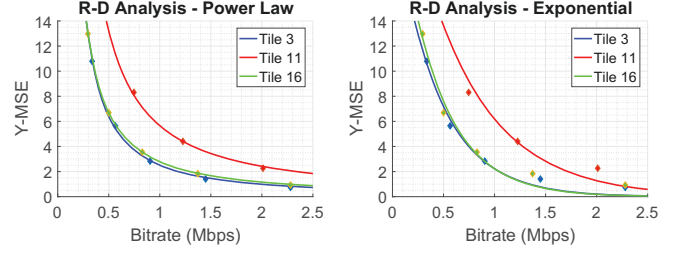
Examination of the employed QP versus induced bitrate relationship for different tiles shows that exponential model fits better the actual data points. In Figure 9, we examine these data points, shown as markers, and the fitted analytical dependencies according to the two formulated models, for three representative tiles, with diverse rate-distortion characteristics, from the Roller Coaster video. Referencing the tile indexing from Figure 3, we can see that while tiles 3 and 16 show lower bitrate requirements due to their relatively static nature, tile 11 requires a higher bitrate as it corresponds to a more dynamic 360° region.

Figure 10 shows the advantage of the power law model in describing the observed D versus R dependency, denoted with markers, across the 360° video tiles. In particular, for lower bitrates, the impact of higher distortion dominates for tiles with more dynamic content (Tile 11), while for higher bitrates the difference across different tiles in this regard becomes smaller, as seen from Figure 10.



(a) QP vs. bitrate dependency using an exponential model. (b) QP vs. bitrate dependency using a power law model.

Fig. 9: QP vs. bitrate dependency for different tiles. Actual data points shown as markers.

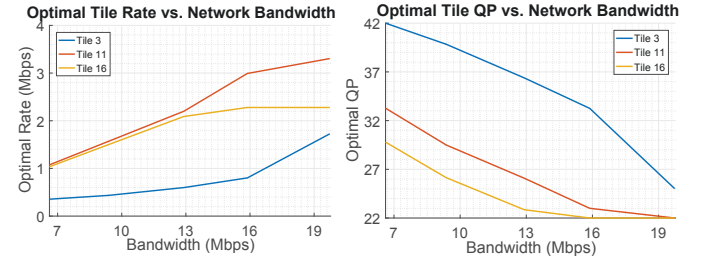


(a) Bitrate vs. distortion dependency using a power law model. (b) Bitrate vs. distortion dependency using an exponential model.

Fig. 10: Bitrate vs. distortion dependency for different tiles. Actual data points shown as markers.

C. Optimal tile QP and data rates vs. available bandwidth

We examine how the optimal data rates R_i and the corresponding QP_i values, produced by the optimization in (5) for every tile i , vary, as the available network bandwidth C is varied. Figure 11a shows the optimal rates produced by (5) for three tiles from the Roller Coaster video, for the GOP number 57 in the 360° video, selected as a representative example. For this GOP, tile 3 has a small navigation likelihood, while tile 16 has the highest among the three tiles considered. Still, it is interesting to note that although tile 16 has a higher navigation likelihood relative to tile 11 and is assigned a smaller QP earlier (as seen from Figure 11b right), encoding tile 11 leads to a higher data rate in the second half of the graph in Figure 11a, due to its more dynamic content, which makes encoding it more challenging.



(a) Tile rate vs. bandwidth. (b) Tile QP vs. bandwidth.

Fig. 11: Optimal tile rate/QP values vs. network bandwidth C .

Figure 12 shows the temporal evolution of the optimal QP and bitrate values for these three tiles over the GOPs comprising the 360° content. We can see that tile 16 typically has a lower QP value relative to the other two tiles, due to its frequently accessed spatial location, while tile 3 is often

navigated only for a brief period of time towards the end of the video. Discontinuities in Figure 12a indicate that a tile has not been assigned any rate (skip encoding mode) by the optimization in (5), as indicated by the corresponding graphs in Figure 12b.

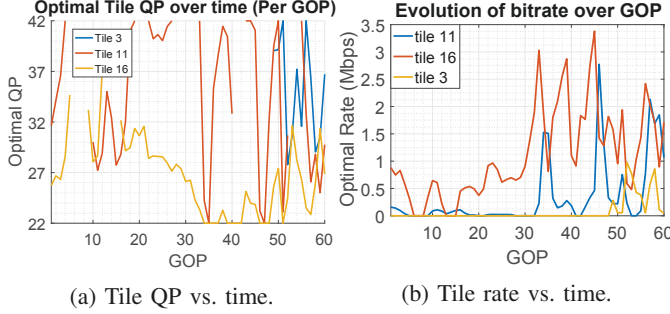


Fig. 12: Optimal tile rate/QP values vs. GOP index.

D. Expected 360° video quality

In the analysis here, *Proposed* denotes our optimization framework, while *Speed-based* and *Monolithic* denote the two references methods introduced earlier. For all three 360° video streaming systems under comparison, we measured the video quality per viewport experienced by a user navigating the 360° content, as the luminance PSNR of the MSE of the pixels displayed in that viewport. Figure 13 shows the viewport PSNR

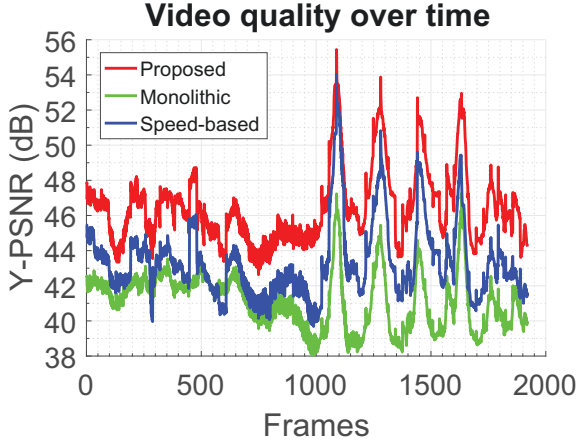


Fig. 13: 360° viewport video quality: Roller Coaster.

over time for the three competing systems in the case of the Roller Coaster video. Here, we encoded the monolithic 360° video (*Monolithic*) with a fixed QP value of 36 and recorded the resulting data rate per GOP to use it as the corresponding rate constraint in (5) for our optimization (*Proposed*) and similarly for the other reference method (*Speed-based*). We can see from Figure 13 that all three systems exhibit the same temporal pattern in viewport PSNR variations, as the dynamic 360° content evolves, with our framework outperforming the two reference methods consistently and considerably. We also observed that *Speed-based* offers an improved performance over *Monolithic*, when viewport prediction succeeds. Though there are minor variations for some frames, we observed that on average *Proposed* provides a 5 dB gain over *Monolithic* and a 3 dB gain over *Speed-based*.

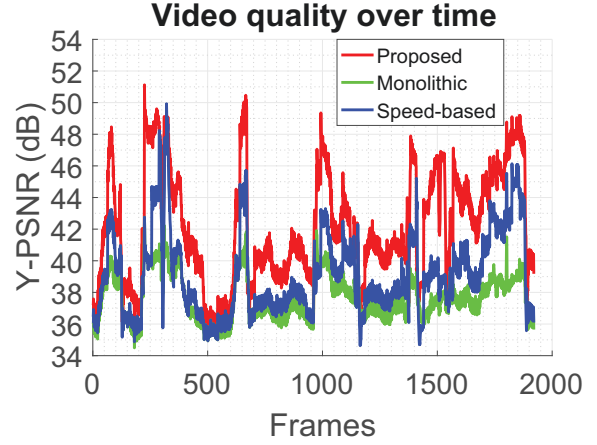


Fig. 14: 360° viewport video quality: Wingsuit.

We observe a different viewport PSNR pattern for the Wingsuit video, as seen from Figure 14. As noted in Figure 6 earlier, the navigation likelihoods over the applied tiling are more uniform in this case, which means the user viewport varies more over time. This negatively impacts the performance of *Speed-based*, which now appears closer to that of *Monolithic*. Still, due to its rate-distortion foundation *Proposed* outperforms again these two reference methods, enabling an average gain of 3 dB and 4 dB relative to *Speed-based* and *Monolithic*, respectively.

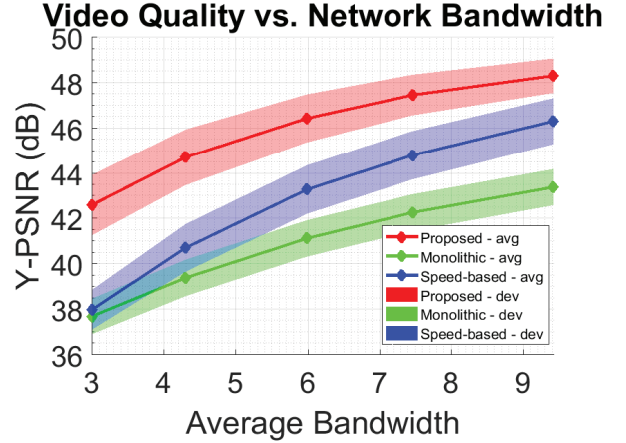


Fig. 15: Average 360° viewport video quality: Roller Coaster.

Next, we examine the average (over time) viewport 360° video quality (Y-PSNR) delivered by the three competing systems, as the available network bandwidth C is varied. Figure 15 show these results in the case of Roller Coaster, together with the corresponding video quality standard deviation exhibited by each system. We can see that again *Proposed* outperforms *Speed-based* and *Monolithic*, with a consistent gain of up to 4-5 dB, across the entire range of values examined for C . As expected, *Proposed* and *Speed-based* exhibit a somewhat higher viewport video quality standard deviation relative to *Monolithic*, since the latter encodes all tiles with a uniform QP value (thus video quality). On the other hand, the reconstruction error can vary more spatially across pixels in viewports delivered by *Proposed* and *Speed-based*.

based, due to the applied tiling, especially as the number of tiles that comprise a viewport increases.

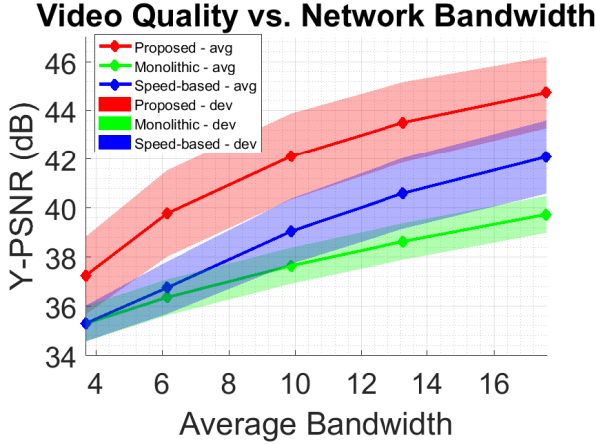


Fig. 16: Average 360° viewport video quality: Wingsuit.

This phenomenon is emphasized even more in the case of Wingsuit, as indicated by the corresponding results in Figure 16, since the navigation likelihoods for tiles closer to the south pole are higher for this 360° video content (see Figure 6), which in turn causes viewports to more often comprise a higher number of tiles.

We measured in our experiments that our framework leads to 42% rate savings relative to the conventional approach *Monolithic*, which is very encouraging.

VI. CONCLUSION

We have formulated a framework for viewport-driven rate optimized 360° video streaming that integrates the user view navigation pattern and the spatiotemporal rate-distortion characteristics of the 360° video content to maximize the delivered user quality of experience for the given network/system resources. It comprises a methodology for constructing dynamic heat maps that capture the user likelihood of navigating different spatial segments of a 360° video over time, analysis and characterization of its spatiotemporal rate-distortion characteristics that leverages preprocessed spatial tiling of the 360° view sphere, and optimization problem formulation that characterizes the delivered user quality of experience given the user navigation patterns, 360° video encoding decisions, and the available system/network resources. Our experimental results demonstrate the advantages of our framework over the conventional approach of streaming a monolithic uniformly-encoded 360° video and a state-of-the-art reference method, enabling considerable video quality of gains of 4 - 5 dB in the case of two popular 4K 360° videos.

There are multiple directions of future work that we consider. In the present framework, we used a given tiling of the 360° view panorama. How the end-to-end performance efficiency varies with the employed 360° tiling is one question we plan to answer. Similarly, will variable-size 360° tiling provide additional gains, and at what cost, is another question we will aim to investigate. Finally, we plan to explore adaptive scalable 360° tiling representations that will account for client and

network heterogeneity intrinsically. Leveraging them towards the design of effective 360° network multicast techniques is another study we plan to carry out in this context.

REFERENCES

- [1] T. Merel. (2017, Jan.) After a mixed year, mobile AR to drive \$108 billion VR/AR market by 2021. [Online]. Available: <http://goo.gl/Lxf4Sy>
- [2] J. G. Apostolopoulos, P. A. Chou, B. Culbertson, T. Kalker, M. D. Trott, and S. Wee, "The road to immersive communication," *Proceedings of the IEEE*, vol. 100, no. 4, pp. 974–990, Apr. 2012.
- [3] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *Proc. IEEE Int'l Conf. Communications (ICC)*, Paris, France, May 2017.
- [4] M. Yu, H. Lakshman, and B. Girod, "A framework to evaluate omnidirectional video coding schemes," in *Proc. IEEE International Symposium on Mixed and Augmented Reality*, Sep. 2015, pp. 31–36.
- [5] "Facebook 360: A stunning and captivating way to share immersive stories, places and experiences." <http://facebook360.fb.com>
- [6] "YouTube: 360° Videos." <https://www.youtube.com/>
- [7] B. Begole, "Why the Internet pipes will burst when virtual reality takes off," *Forbes Magazine*, Feb. 2016.
- [8] E. Knightly, "Scaling Wi-Fi for next generation transformative applications," *Keynote Presentation, IEEE INFOCOM*, Atlanta, GA, May 2017.
- [9] J. D. Moss and E. R. Muth, "Characteristics of headmounted displays and their effects on simulator sickness," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Jun. 2011.
- [10] S. Afzal, J. Chen, and K. K. Ramakrishnan, "Characterization of 360-degree videos," in *Proc. ACM SIGCOMM Workshop Virtual Reality and Augmented Reality Network*, Aug. 2017.
- [11] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits and Systems for Video Technology*, Dec. 2012.
- [12] C. Ozcinar, A. De Abreu, and A. Smolic, "Viewport-aware adaptive 360° video streaming using tiles for virtual reality," in *Proc. IEEE Int'l Conf. Image Processing*, Beijing, China, Sep. 2017.
- [13] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck, "Improving virtual reality streaming using HTTP/2," in *Proc. ACM Multimedia Systems Conference*, Jun. 2017, pp. 225–228.
- [14] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, "Optimizing 360 video delivery over cellular networks," in *Proc. ACM Workshop All Things Cellular: Operations, Applications and Challenges*, Oct. 2016.
- [15] M. Hosseini and V. Swaminathan, "Adaptive 360 VR video streaming: Divide and conquer," in *Proc. IEEE Int'l Symp. Multimedia*, Dec. 2016.
- [16] M. Graf, C. Timmerer, and C. Mueller, "Towards bandwidth efficient adaptive streaming of omnidirectional video over HTTP: Design, implementation, and evaluation," in *Proc. ACM MM Syst. Conf.*, Jun. 2017.
- [17] M. Ben Yahia, Y. Le Louedec, L. Nuyami, and G. Simon, "When HTTP/2 rescues DASH: Video frame multiplexing," in *Proc. IEEE INFOCOM Workshop Communication and Networking Techniques for Contemporary Video*, Atlanta, GA, USA, May 2017.
- [18] A. T. Nasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash, "Adaptive 360-degree video streaming using scalable video coding," in *Proc. ACM Multimedia Conference*, 2017, pp. 1689–1697.
- [19] G. He, J. Hu, H. Jiang, and Y. Li, "Scalable video coding based on user's view for real-time virtual reality applications," *IEEE Communications Letters*, vol. 22, no. 1, pp. 25–28, Jan 2018.
- [20] "Mega coaster: Get ready for the drop (360 video)." [Online]. Available: <https://youtu.be/-xNN-bJQ4vI>
- [21] "Wingsuit 360 degree video over Dubai." [Online]. Available: <https://youtu.be/AX4hWfyHr5g>
- [22] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [23] "Opentrack: Head tracking software." [Online]. Available: <https://github.com/opentrack/opentrack>