Optimal 360° Video Streaming to Head-Mounted Virtual Reality

Ching-Ling Fan
Department of Computer Science, National Tsing Hua University, Taiwan

ABSTRACT

Emerging Virtual Reality (VR) applications become more and more popular because of the increasing availability of 360° videos and commodity Head-Mounted Displays (HMDs). Streaming 360° videos to HMDs is very challenging due to large video sizes, stringent real-time requirements, and complex human visual systems. We propose to study three core problems to optimize 360° videos streaming to HMDs. First, we adopt sensor and content data to design a fixation prediction algorithm. Second, we develop a user study to construct a 360° Quality-of-Experience (QoE) model. Last, we design, implement, and evaluate an optimal bitrate allocation algorithm that leverages both the fixation prediction algorithm and QoE model. Our preliminary results are promising, e.g., our proposed system achieves 92% accuracy in fixation prediction and reduces the bandwidth consumption by about 31%. Several ongoing tasks aim to improve the proposed 360° video streaming system for: (i) higher fixation prediction accuracy, (ii) more comprehensive QoE model, and (iii) optimal bitrate allocation and adaptation. Our research outcomes are also useful to optimize future VR applications, such as social VR, telemedicine, and novel entertainments.

1 INTRODUCTION

360° videos become increasingly popular because they offer free viewing orientation and more immersive experience. It is reported that the annual market potential of live 360° video streaming will reach trillion dollars [3] soon. One of the reasons behind the huge market potential of 360° videos is the emerging Virtual Reality (VR) technologies. In fact, a market study predicts that the VR market will grow at an anual rate of 57.8% between 2016 and 2022 [5]. The growing VR popularity can be attributed to the blooming commodity Head-Mounted Displays (HMDs), such as HTC Vive [2], Samsung Gear VR [4], and Oculus Rift [1]. These HMDs are equipped with sensors to track the head position and orientation for more integrated viewing experience.

Streaming 360° videos, however, is no easy task. First, 360° videos are projected from spheres to flat surfaces, because video codecs only support rectangle videos. There exist different *projection models*, including equirectangular, cube, and rhombic dodecahedron [12]. Individual projection models have pros and cons in diverse aspects like redundancy, shape distortion, and complexity. The projected videos are quite different from the reconstructed videos viewed in HMDs. This in turn makes conventional video tools less effective to 360° videos. For example, computer vision algorithms, trained with conventional videos, may not work well on projected videos; while objective video quality metrics (such as Peak-Signal-to-Noise Ratio, PSNR) may be biased by the redundant pixels in the projected videos. Second, 360° videos contain much more information than conventional ones, and thus 360° videos are in higher resolutions and bitrates. 360° video streaming systems, therefore, are vulnerable to insufficient and unstable bandwidth due to significant cross traffic, or high wireless dynamics, resulting in degraded Quality of Experience (QoE). On the other hand, HMD viewers get to see a small viewable region of each 360° video at any moment. Therefore, streaming whole 360° videos leads to large *unwatched* regions, and thus wastes bandwidth.

In our research project, we optimize 360° video streaming systems in three directions, which are briefly summarized below.

- We propose a new *fixation* prediction algorithm for 360° videos in HMDs, which takes both sensor and content data as inputs and predicts the viewer fixation. The fixation refers to the region that is most likely to be watched.
- We conduct user studies to understand the complex interplay among HMD viewer QoE, streaming system parameters, and video content characteristics. This leads to QoE metrics specifically tailored for 360° videos in HMDs.
- We design an adaptive streaming system to optimize the viewing experience under the bandwidth constraints. The crux of the streaming system is an optimal adaptation algorithm that periodically determines which regions to transmit and at what bitrates.

In the rest of this paper, we report our preliminary findings in these three directions. The expected outcome of our project is *an optimized 360° video streaming system to HMD viewers*[1]. The prototype system built in our project can be used to evaluate the performance of new proposals for optimizing similar systems. The lessons learned in our project can also be leveraged in future 360° video related applications.

2 RELATED WORK

Fixation prediction. Conventional fixation prediction is built on salient object detection, which has been done with content data, such as still images [8]. For videos, Mavlankar and Girod [19] propose to use thumbnails, motion vectors, and navigation trajectories to perform fixation prediction. Emerging machine learning algorithms, such as neural networks, are also adopted for fixation detection [6, 9, 21]. Chaabount et al. [9] perform fixation prediction in videos using deep learning with residual motion features. Alshawi et al. [6] observe the correlation between the eye-fixation maps and the spatial/temporal neighbors to quantify viewer attention on videos. Nguyen et al. [21] point out a close relationship between static and dynamic saliency. Based on the observation, they propose to take both static saliency (in images) and camera motion into considerations for fixation prediction in videos. These studies do not focus on 360° video streaming to HMDs, and the HMD sensor data are ignored.

QoE metrics. Several video quality metrics evaluate the pixel-level and structure-level quality on conventional videos, such

[1]Our proposed solutions are not limited to VR but also applicable for Augmented Reality (AR) applications.

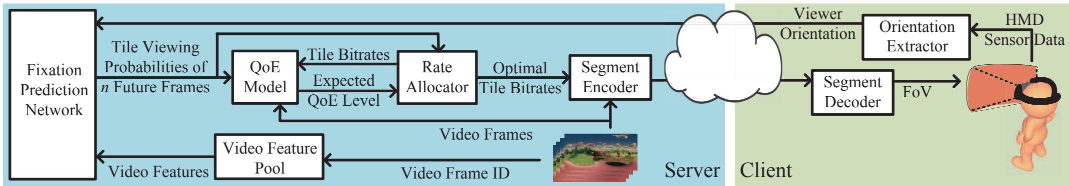


Figure 1: Overview of the proposed 360° video streaming systems.

as Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM). However, these quality metrics fail to evaluate 360° videos, because of the shape distortion and redundant pixels when being projected to flat surfaces. Recently, some quality metrics [30, 31] are proposed for 360° videos. Yu et al. [30] propose to calculate PSNR values of the pixels: (i) normally-sampled on spheres or (ii) in the viewed region. Zakharchenko et al. [31] modify the MSE (Mean Squared Error) by introducing a weighted function of the latitude and longitude for videos under equirectangular projection. Some very recently studies [23, 24, 26] investigate QoE from various aspects for 360° visual content. Upenik et al. [26] conduct user studies for evaluating existing objective quality metrics and their results show that a better quality metric is required for 360° images. Schatz et al. [24] study the negative impact of video frozen on streaming 360° videos. Rai et al. [23] focus on the optimal weighted function for HMD without gaze-tracking functionality. Unfortunately, a promising QoE model designed for 360° videos to HMD has never been proposed.

Adaptive 360° video streaming. 360° video streaming can be done either through real-time transcoding [7, 28] or tile streaming [11, 14, 20, 29]. We consider tile streaming approach for higher scalability. Gaddam et al. [15] propose an interactive panoramic video streaming system with a tile-based encoder. They gradually decrease the quality from the center of the current viewed region. This approach saves bandwidth while providing smooth quality degradation. Spatial Relationship Description (SRD) [22], an extension of DASH (Dynamic Adaptive Streaming over HTTP), enables spatial random access in video streaming. Several studies [11, 14] adopt SRD as a new way to realize tile-based video streaming for higher flexibility in terms of spatial relationship and encoder/decoder supports. Zhou et al. [32] analyze an undocumented projection model, offset cubic projection, implemented in Oculus HMD streaming system. They find that the projection offers comparable video quality, when saves up to 16.4% in bitrate. Graf et al. [17] describe the usage of tiled streaming for 360° videos and conduct experiments to investigate the bitrate overhead and bandwidth requirements under different tiling strategies. These existing studies do not dynamically adapt the 360° video streaming under different HMD viewing patterns.

3 SYSTEM OVERVIEW

Fig. 1 presents the proposed 360° video streaming system. The interactions among the main components are as follows. The video frame IDs are sent to the video feature pool to get the corresponding content features. These features include saliency maps and motion maps, and are pre-extracted when the videos are stored on the server. The sensor features, including viewer orientation, are extracted by the orientation extractor from HMD sensor data. The fixation prediction network predicts the future viewing probabilities of individual tiles using content and sensor features. The predicted probabilities of tiles are sent to the rate allocator for performing bitrate allocation among tiles. To assist the rate allocator, the QoE model estimates the QoE levels based on the predicted probabilities and tentative bitrate allocations. After querying the QoE model several times, the rate allocator chooses an optimal bitrate allocation of video segments of selected tiles. The segment encoder then encodes the video segments according to the assigned tile bitrates. Last, the segment decoder at the client side decodes and displays in the HMDs.

4 RESEARCH PROBLEMS

4.1 Fixation Prediction

The fixation prediction problem aims to estimate the probability of each region of a video to be watched in the *future*. By future, we refer to a few seconds ahead, which is required by the de facto DASH streaming protocol. With the prediction results, 360° video streaming systems can *skip* those video tiles that are unlikely to be watched to save the network bandwidth consumption. Therefore, more bandwidth can be invested on the tiles that will be watched, for better user experience. We adopt neural networks to predict view fixation. We leverage sensor features, which are the viewer orientation read from HMD sensors, and content features, which are saliency and motion detected from video frames, to build our fixation prediction network.

4.2 QoE Modeling

The shape distortion and redundant pixels caused by projection models make the conventional video quality metrics, such as PSNR and SSIM, ineffective for evaluating viewing experience of 360° videos. More importantly, human visual systems are fairly complex to capture. Therefore, modeling the relation among HMD QoE (quantified in Mean Opinion Score, MOS), streaming system parameters, and video content characteristics is time-consuming and difficult. To understand how different factors affect HMD QoE, we plan to design and conduct user studies to exercise diverse projection models, video content types, resolutions, encoding bitrates, and etc. With inputs from HMD viewers, we analyze the correlation between existing quality metrics [30] and QoE. More importantly, we will derive a new QoE model, which to our best knowledge, has never been done in the literature.

4.3 Adaptive Streaming

In the context of 360° video streaming to HMDs, we build a rate control algorithm, to determine the bitrate of each video tile, so as to maximize the HMD QoE. In particular, we leverage the fixation prediction network and QoE model mentioned above. We first decide what are the tiles to transmit using the fixation prediction network, and then allocate the bitrate among all transmitted tiles to maximize the expected QoE across all viewed tiles. In addition to a basic rate allocator for unicast session, we will also build a more comprehensive rate allocator for multiple unicast/multicast sessions in larger-scale 360° video streaming systems.

5 PRELIMINARY SOLUTIONS AND RESULTS

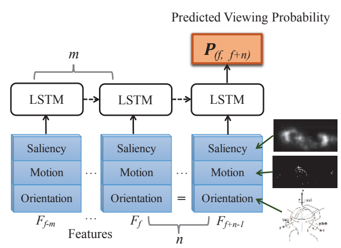


Figure 2: The enhanced fixation prediction network.

5.1 Fixation Prediction

Dataset collection and training. We set up a 360° video streaming testbed to collect our own dataset. We download ten 360° videos from YouTube with 4K resolution and frame rate of 30 Hz. We classify the videos into three categories: (i) Computer-Generated (CG), fast-paced, (ii) Natural Image (NI), fast-paced, and (iii) NI, slow-paced. We recruit 50 viewers and play ten videos to each viewer. The orientation of viewers are logged during watching videos, which are then used to derive the features and ground truth. The fixation prediction network is implemented using Keras and Scikit-Learn. We sample the points within the viewable region of the viewer to the equirectangular projection and the ground truth are calculated according to the tiles overlapped with those sample points.

Enhanced prediction network. In Fan et al. [13], we design fixation prediction networks based on Recurrent Neural Network (RNN), in particular, Long Short Term Memory (LSTM) network to learn long-term characteristics from the video frames. Our proposed networks take features of m past video frames to predict the viewing probabilities of tiles of n future video frames. We enhance our prior proposal in several aspects, as illustrated in Fig. 2. First, we consider the future content features. The future viewer orientation are filled in with the last received viewer orientation. However, including these features may introduce high overhead on training and prediction. Therefore, we sample the features at 1 Hz. This is reasonable because there is almost no change on video content nor on viewer orientation in such a short period of time. Second, the ground truth of the enhanced prediction network is considered as the viewing probability of each tile in next n frames.

Results. We consider the number of neurons in {256, 512, 1024, 2048, 4096, 8192}, the number of layers in {1, 2, 3}, and dropout in {true, false}. We first fix the number of neurons at 2048 and find the optimal combination of the other two parameters. Then, we adjust the number of neurons to find the final optimal parameters. The enhanced network with parameters as 4096 neurons, 2 layers, and dropout performs the best, achieving 91.07% of accuracy and 0.76 of F-score. The detailed numbers are omitted due to the space limitation.

Ongoing work. We have started looking into the negative impacts of 360° video projection models on saliency detection accuracy. Moreover, we will leverage multi-level saliency detection, which considers sampled and overlapped small viewable regions with various sizes. This is to mitigate projection shape distortion, in order to improve the proposed fixation prediction networks. Besides, a viewer-dependent fixation prediction model is also desirable for better prediction accuracy by considering viewer characteristics, such as viewer interests and historical statistics.

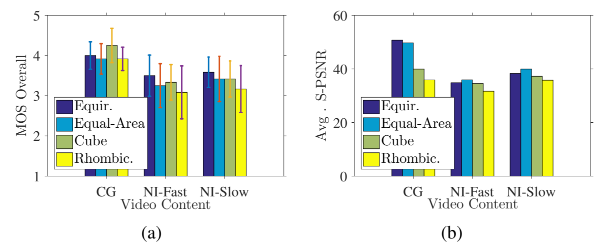


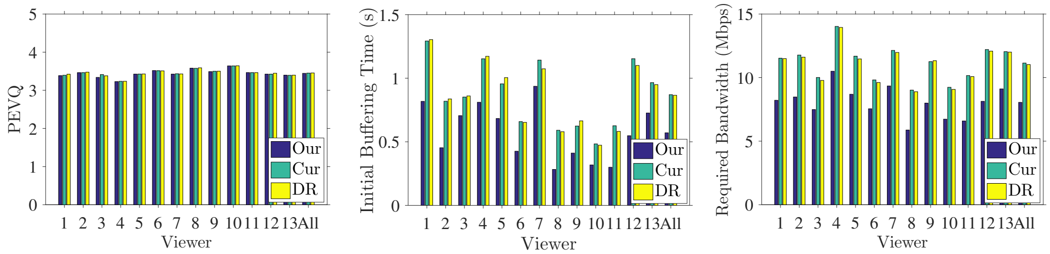
Figure 3: The results of (a) MOS and (b) S-PSNR under different projection models and video contents.

5.2 Quality Metrics

Preliminary user study. We select one 360° video from each category for the user study. We encode each video at 6 Mbps, and consider 4 projection models: (i) equirectangular, (ii) equal-area, (iii) cube, and (iv) rhombic dodecahedron. We set the resolutions of different projection models comparable to each other for fair comparisons. We recruit 12 viewers and randomly play 12 videos (with 3 video contents and 4 projection models) to each viewer. The viewer fills a questionnaire about MOS of their experience, between 1 (lowest) and 5 (highest).

Results. Fig. 3 shows the overall MOS scores and S-PSNR values [30] under different projections and video contents. Both the MOS and the S-PSNR of the CG videos are higher than others since this video is easier to compress. Among different projection models, we expect that the cube may perform better since it avoids encoding redundant pixels closer to poles. However, Fig. 3(a) shows that equirectangular achieves better MOS in general, which is counter-intuitive. This may be caused by the diverse resolutions used by different projection models, which require more investigation. Moreover, Fig. 3 shows the different trends between the QoE and the S-PSNR, which reveals the limitations of S-PSNR [30], which motivates us to develop new quality metrics.

Ongoing work. We have started conducting user studies on 360° video streaming with more controlled factors, such as bitrates and video content. We aim to build a comprehensive QoE model considering various aspects, including bitrate allocation, video types, and viewer characteristics.



(a) (b) (c)

Figure 4: Compared to the existing solutions, our fixation prediction network (a) achieves comparable video quality in PEVQ, while reducing both (b) initial buffering time and (c) required bandwidth.

5.3 Adaptive Streaming

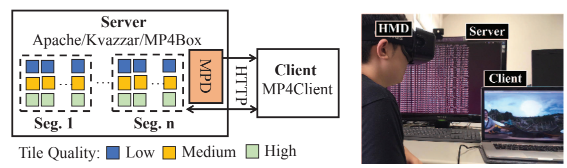
Simulation setup. We have implemented a simulator in Python to quantify the performance of our proposed system. In our simulator, the server supports V concurrent viewers and each viewer randomly selects a 360° video to watch. The total bandwidth is B, which is shared by all viewers. The streaming system has a latency of D secs, and the videos are divided into s-sec segments with a 192x192 tile size. We adopt the enhanced prediction network model for fixation prediction. Each viewer connects to the server and requests for the tiles predicted by the past m sampled frames, where the sample frequency is 1 Hz. The predicted tiles are the tiles predicted to be viewed by the viewers in next n frames, where n is determined by s and D. Each simulation lasts for 1 min. We set the Quantization Parameters (QPs) of all tiles as 28. Only the requested tiles are streamed to the client.

We consider the following performance metrics: (i) missing ratio, which is the number of missed tiles over the number of viewed tiles, (ii) required bandwidth, (iii) initial buffering time, which is the minimum buffering time for smooth playout, and (iv) video quality in PSNR, SSIM, and Perceptual Evaluation of Video Quality (PEVQ) [25]. We compare our solution to three baselines: (i) Current (Cur), which uses the current orientation as the prediction and (ii) Dead Reckoning (DR) [19], which uses the velocity of viewer orientations to predict. We run a preliminary simulation with $V = 13$ viewers, $B = 150$ Mbps, $D = 2$ secs, $m = 1$, $s = 1$ secs, and $\rho = 0.5$. This setting leads to an impractical 30% missing ratio. Thus, we augment our proposed solution and baselines to lower the missing ratio to $< 10\%$ by adjusting ρ and δ, where δ is the number of iterations on adding extra tiles at the edge of predicted tiles to achieve low missing ratio.

Results. Fig. 4(a) plots the perceived quality in PEVQ of each viewer and the average perceived quality over all 13 viewers. This figure shows that the perceived quality of our solution is comparable to other baselines. We also plot the minimum initial buffering time of each solution in Fig. 4(b). This figure shows that our solution leads to shorter initial buffering time, which is inline with the lower required bandwidth of our solution as shown in Fig. 4(c). The superior performance of our solution can be attributed to its higher flexibility and intelligence. In particular, Cur and DR enlarge the streamed region from their current predicted view to lower the missing ratio, while our solution achieves the same goal by streaming the tiles with higher viewing probabilities. Hence, our solution achieves comparable video quality while reducing 31% and 28% in bandwidth consumption and initial buffering time on overage.

Ongoing work. Designing a bitrate allocation algorithm is not an easy task. A possible solution is to employ Lagrangian optimization [18] to design a joint tile selection and bitrate selection algorithm to maximize the expected QoE. Another possibility is the learning-based approach [10] to predict QoE level using viewer and video features.

5.4 Testbed Implementation



(a) (b)

Figure 5: Our 360° streaming testbed: (i) system topology and (b) actual testbed.

We have built an end-to-end testbed using existing opensource projects, as illustrated in Fig. 5. In our testbed, video segments are encoded using Kvazaar [27], which is an open-source H.265/HEVC encoder. The Media Presentation Description (MPD) files, which point to all encoded tile segments, are generated using MP4Box [16]. We adopt MP4Client [16] as the client. Currently, we are enhancing MP4Cleint so that we can: (i) select tiles to request and (ii) select bitrates for individual tiles. The suite of optimization algorithms designed to address the three core problems will be tested in our testbed.

6 CONCLUSION

In our proposal, we study three core problems of optimizing 360° video streaming to HMDs. First, we leverage both sensor and content data to train a neural network for viewer fixation prediction, which is promising but still has room for improvement. Second, we conduct a user study to investigate the impacts of different factors on HMD QoE. We will further extend the user study and analysis for a comprehensive QoE model. Last, we design an adaptive streaming system and conduct trace-driven simulations to understand its practicality and limitations. Based on the simulation results, we plan to design, implement, and evaluate an optimal rate allocation algorithm, which will be the center of the whole streaming system. Our proposed solutions are also applicable to future VR applications, such as social VR, telemedicine, and novel entertainments.

REFERENCES

[1] 2016. Facebook Oculus Rift. (2016). https://www.oculus.com.

[2] 2016. HTC Vive. (2016). https://www.htcvive.com.

[3] 2016. The Market Potential for 360 Video and Virtual Reality Production. (2016). https://goo.gl/o9By8H.

[4] 2016. Samsung Gear VR. (2016). http://www.samsung.com/global/galaxy/gear-vr.

[5] 2016. Virtual Reality Market Worth 33.90 Billion USD by 2022. (2016). http://www.marketsandmarkets.com/PressReleases/ar-market.asp.

[6] T. Alshawi, Z. Long, and G. AlRegib. 2016. Understanding Spatial Correlation in Eye-Fixation Maps for Visual Attention in Videos. In *Proc. of IEEE International Conference on Multimedia and Expo (ICME'16)*. Seattle, WA, 1–6.

[7] R. Aparicio-Pardo, K. Pires, A. Blanc, and G. Simon. 2015. Transcoding Live Adaptive Video Streams at a Massive Scale in the Cloud. In *Proc. of ACM International Conference on Multimedia Systems (MMSys'15)*. Portland, OR, 49–60.

[8] A. Borji, M. Cheng, H. Jiang, and J. Li. 2014. Salient object detection: A survey. *arXiv preprint arXiv:1411.5878* (2014).

[9] S. Chaabouni, J. Benois-Pineau, and C. Amar. 2016. Transfer Learning with Deep Networks for Saliency Prediction in Natural Video. In *Proc. of IEEE International Conference on Image Processing (ICIP'16)*. Phoenix, AZ, 1604–1608.

[10] F. Chiariotti, S. D'Aronco, L. Toni, and P. Frossard. 2016. Online Learning Adaptation Strategy for DASH Clients. In *Proc. of ACM International Conference on Multimedia Systems (MMSys'16)*. Klagenfurt, Austria, 8:1–8:12.

[11] L. D'Acunto, J. Berg, E. Thomas, and O. Niamut. 2016. Using MPEG DASH SRD for Zoomable and Navigable Video. In *Proc. of ACM International Conference on Multimedia Systems (MMSys'16)*. Klagenfurt, Austria, 34:1–34:4.

[12] T. El-Ganainy and M. Hefeeda. 2016. Streaming Virtual Reality Content. *arXiv preprint arXiv:1612.08350* (2016).

[13] C. Fan, J. Lee, W. Lo, C. Huang, K. Chen, and C. Hsu. 2017. Fixation Prediction for 360° Video Streaming in Head-Mounted Virtual Reality. In *Proc. of ACM International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV'17)*. Taipei, Taiwan.

[14] J. Feuvre and C. Concolato. 2016. Tiled-Based Adaptive Streaming Using MPEG-DASH. In *Proc. of ACM International Conference on Multimedia Systems (MMSys'16)*. Klagenfurt, Austria, 41:1–41:3.

[15] V. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen. 2016. Tiling in Interactive Panoramic Video: Approaches and Evaluation. *IEEE Transactions on Multimedia* 18, 9 (2016), 1819–1831.

[16] GPAC–Multimedia Open Source Project 2017. GPAC–Multimedia Open Source Project. (2017). https://gpac.wp.imt.fr/.

[17] M. Graf, C. Timmerer, and C. Mueller. 2017. Towards Bandwidth Efficient Adaptive Streaming of Omnidirectional Video over HTTP. In *Proc. of ACM Conference on Multimedia Systems (MMSys'17)*. Taipei, Taiwan.

[18] C. Kuo, Y. Shih, and S. Yang. 2016. Rate Control via Adjustment of Lagrange Multiplier for Video Coding. *IEEE Transactions on Circuits and Systems for Video Technology* 26, 11 (2016), 2069–2078.

[19] A. Mavlankar and B. Girod. 2010. Video Streaming with Interactive Pan/Tilt/Zoom. In *Signals and Communication Technology*. 431–455.

[20] K. Ngo, R. Guntur, and W. Ooi. 2011. Adaptive Encoding of Zoomable Video Streams Based on User Access Pattern. In *Proc. of ACM Conference on Multimedia Systems (MMSys'11)*. San Jose, CA, 211–222.

[21] T. Nguyen, M. Xu, G. Gao, M. Kankanhalli, Q. Tian, and S. Yan. 2013. Static Saliency vs. Dynamic Saliency: a Comparative Study. In *Proc. of ACM International Conference on Multimedia (MM'13)*. Barcelona, Spain, 987–996.

[22] O. Niamut, E. Thomas, L. D'Acunto, C. Concolato, F. Denoual, and S. Lim. 2016. MPEG DASH SRD: Spatial Relationship Description. In *Proc. of ACM International Conference on Multimedia Systems (MMSys'16)*. Klagenfurt, Austria, 5:1–5:8.

[23] Y. Rai, P. Callet, and P. Guillotel. 2017. Which Saliency Weighting for Omnidirectional Image Quality Assessment?. In *Proc. of IEEE International Conference on Quality of Multimedia Experience (QoMEX'17)*. Erfurt, Germany.

[24] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo. 2017. Towards Subjective Quality of Experience Assessment for Omnidirectional Video Streaming. In *Proc. of IEEE International Conference on Quality of Multimedia Experience (QoMEX'17)*. Erfurt, Germany.

[25] K. Skarseth, H. Bjørlo, P. Halvorsen, M. Riegler, and C. Griwodz. 2016. OpenVQ: a Video Quality Assessment Toolkit. In *Proc. of ACM International Conference on Multimedia (MM'16), OSSC paper*. 1197–1200.

[26] E. Upenik, M. Rerabek, and T. Ebrahimi. 2017. On the Performance of Objective Metrics for Omnidirectional Visual Content. In *Proc. of IEEE International Conference on Quality of Multimedia Experience (QoMEX'17)*. Erfurt, Germany.

[27] M. Viitanen, A. Koivula, A. Lemmetti, A. Ylä-Outinen, J. Vanne, and T. Hämäläinen. 2016. Kvazaar: Open-Source HEVC/H.265 Encoder. In *Proc. of ACM on Multimedia Conference (MM'16)*. Amsterdam, The Netherlands, 1179–1182.

[28] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg. 2010. Real-Time Panoramic Mapping and Tracking on Mobile Phones. In *Proc. of Virtual Reality Conference (VR'10)*. Waltham, MA, 211–218.

[29] H. Wang, M. Chan, and W. Ooi. 2015. Wireless Multicast for Zoomable Video Streaming. *ACM Transactions on Multimedia Computing, Communications, and Applications* 12, 1 (2015), 5:1–5:23.

[30] M. Yu, H. Lakshman, and B. Girod. 2015. A Framework to Evaluate Omnidirectional Video Coding Schemes. In *Proc. of IEEE International Symposium on Mixed and Augmented Reality (ISMAR'15)*. Fukuoka, Japan, 31–36.

[31] V. Zakharchenko, K. Choi, and J. Park. 2016. Quality Metric for Spherical Panoramic Video. In *Proc. of Optics and Photonics for Information Processing*. San Diego, CA, 99700C:1–99700C:9.

[32] C. Zhou, Z. Li, and Y. Liu. 2017. A Measurement Study of Oculus 360 Degree Video Streaming. In *Proc. of ACM Conference on Multimedia Systems (MMSys'17)*. Taipei, Taiwan.