

Quality Assessment for Adaptive Virtual Reality Video Streaming: A Probabilistic Approach on the User’s Gaze

Jeroen van der Hoof[†], Maria Torres Vega[†], Stefano Petrangeli[§], Tim Wauters[†] and Filip De Turck[†]

[†] IDLab, Department of Information Technology, Ghent University - imec

[§] Adobe Research

E-mail: jeroen.vanderhooft@ugent.be

Abstract—Accurately estimating the users’ perception of adaptive tile-based Virtual Reality (VR) video streaming is fundamental and is often used to determine subsequent streaming decisions. This is due to the fact that adaptive tile-based VR video streaming solutions aim to optimize the bandwidth usage by only streaming the areas (tiles) within the field of view of the user (viewport) at the highest quality, while keeping the remainder of the 360° environment at low quality levels. Thus, understanding the quality perceived by the user in real-time determines the success of the service. Current quality assessment approaches tend to map the quality of the viewport’s center (center tile) to the overall perceived quality. However, while such models enable to measure how long the user spend on each quality layer, they ignore the fact that the user’s eyes also move within the viewport. In this paper, we present a novel video quality assessment metric for adaptive tile-based VR video streaming which takes into account the distribution of eye movement within the viewport, as well as, the tiling scheme. Both elements are combined by means of a probability distribution fit from real users data. The model was evaluated in a well-known VR database and using an emulation environment, benchmarking it against state-of-the-art solutions as well as objective full reference metrics.

I. INTRODUCTION

As the price of Head Mounted Displays (HMD) decreases, the demand of Virtual Reality (VR) video streamed to mobile devices does not cease to increase. Facebook¹ or YouTube² already offer 360° streaming options encoded at medium to high resolutions (maximum 4K) to their clients. In order to provide a fully immersive experience, VR videos demand significantly higher bandwidth, which is often neither available within the current wireless network infrastructures, nor easy to process by light-weight devices such as the smartphones usually installed within a HMD. In addition, while a VR video contains a full 360° panoramic view, only a portion of it, namely the Field of View or viewport is seen by the user in every given instant.

In an attempt to optimize the bandwidth, viewport-aware schemes for VR video streaming based on the HTTP Adaptive Streaming (HAS) paradigm are being explored since the last couple of years [1], [2]. The HAS paradigm proposes to encode the source content at multiple quality representations (bitrates, resolutions). Each of them is segmented into shorter sequences (*i.e.*, segments). To further optimize the bandwidth

usage, recent investigations have been proposed where the quality representations are further split in the spatial domain into tiles [3], which can be individually accessed at streaming time. Approaches such as [4], [5], [6] bring this concept to the challenging VR arena. To this end, the VR video tiles corresponding to the viewport are streamed at the highest quality, while others are either transmitted at lower levels or not at all [6].

Even if viewport-aware techniques lead to bandwidth optimizations, the effects of network performance on VR video streaming still plays an important role on the user’s perception of the services, *i.e.* the user’s Quality of Experience (QoE). Thus, being able to assess the quality perceived by the user in real-time and making the necessary changes if required is fundamental. Traditional HAS has dealt with quality assessment by means of objective QoE models providing an estimation of the user’s perceived quality during the playout of the previous segment(s) based on, among else, the bitrate and resolution of the resulting video [7]. However, the application of the second segmentation dimension (spatial) introduces new factors that are not taken into account on the traditional HAS QoE modelling. In addition, as shown in results obtained by Rai et al., presented in Figure 1, the user’s eyes are rarely fixed to the center of the viewport. It is, instead, possible to observe a peak for angles between 12 and 20 degrees [8]. Including this effect in the modeling of quality could enhance its accuracy when assessing the human perception of the systems.

In this paper we present a novel video quality assessment metric for adaptive tile-based VR video streaming which takes into account both the spatial dimension as well as the distribution of eye movement within the viewport. Both are combined by means of a probability distribution fit from real-users data. We have evaluated our approach for a well-known VR database and using an emulation environment. We have benchmarked our solution against state-of-the-art solutions as well as well-known state-of-the-art objective full reference metrics.

The remainder of this paper is distributed as follows. Section II, provides a brief discussion of the current state-of-the-art on objective metrics used for VR video streaming pin-pointing what distinguishes our approach. Section III, describes in detail our quality assessment model. Section IV, deals with the adaptive tile-based set-up and the components used for the evaluation. Section V, presents the results. Finally, Section VI,

¹www.facebook.com

²www.youtube.com

discusses the conclusions of this paper and provides guidelines for future work.

II. RELATED WORK

One of the biggest challenges of adaptive video streaming (2D, 3D or VR) applications is the accurate and real-time estimation of quality perceived by the users as well as the provisioning of a feedback loop to dynamically influence the quality adaptation, based on this estimation. This means that even if the user perception is best assessed by humans during subjective evaluation sessions or by full reference objective metrics, both approaches are deemed unfit for real-time assessment and are better suited for benchmarking purposes. In state-of-the-art adaptive streaming approaches, the modeling of QoE has to rely on objective information obtained at the client, the server or the network-side. In regular HAS, many factors have shown to affect the perception: the video quality, the occurrence of playout freezes, the video’s startup time, the end-to-end latency in live streaming, etc. [9]. When VR content is considered, new factors can be distinguished.

The application of tiles can degrade the user’s perception of the service in unexpected ways. Frequent switching between different quality representations can result in a lower QoE. This is true not only on a temporal level, such as in regular HAS, but also on a spatial level. Furthermore, the speed with which the quality is adjusted when moving around plays a crucial role as well: if the user has to wait several seconds before the quality is adjusted, the perceived quality will be strongly affected. With regard to the video quality perceived within the viewport, a number of evaluation metrics are being used in related work. Some works report the average Peak Signal to Noise Ratio (PSNR) values for the obtained video [10], [11], or show results in terms of video bitrate [11], [10], [12]. Other works consider the quality of the tile in the center of the viewport only, either to average this quality over all segments, or to measure the time spent on each layer [13]. Although the latter has shown to be an excellent evaluation metric for regular HAS, it is uncertain whether it is directly applicable to VR. The potentially lower video quality of surrounding tiles can have a significant impact on the perceived quality in the center of the viewport. To take this into account, some approaches have considered the video quality for a subset of tiles, where tiles are weighted according to a predefined zone they are located in [14]. But this type of models assumes that the effect of each of the tiles within the viewport is the same independent from its position.

Current state-of-the-art on quality modelling for adaptive VR video streaming does not take into account the user’s gaze location when providing measurements. However, as shown in Figure 1, it is possible to see significant variance on where the user’s gaze is located in each instant. In this situation, assuming that the perceived quality is just the quality of the center tile can lead to inaccuracies on the assessment of perception. In the next section, we propose a quality evaluation metric which takes into account the distribution of the user’s gaze, in order to assess the video quality within the viewport.

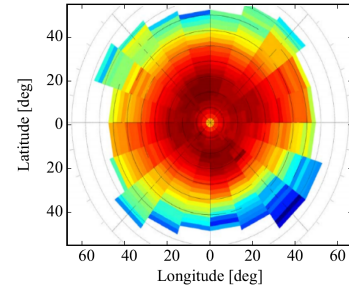


Figure 1: Heatmap of the users’ gaze, relative to the center of the viewport [8]. Blue and red colors indicate a low and high frequency respectively.

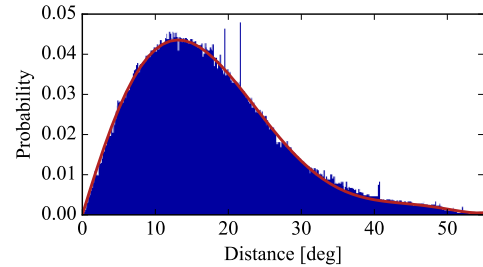


Figure 2: Histogram of the distance between the users’ gaze and the center of the viewport [8]. In red, the best-fitting polynomial function of degree six.

III. A PROBABILISTIC APPROACH ON USER’S GAZE FOR VR VIDEO QUALITY ASSESSMENT

As mentioned in Section II, the related work often focuses on the quality of the viewport center only. While such approaches enable to measure how long the user spends on each quality layer, they ignore the fact that the user’s eyes also move within the viewport. To account for this effect, we introduce a probabilistic approach for quality assessment.

Based on the heatmap shown in Figure 1 and presented by Rai et al. [8], we know that eye movement within the viewport is generally symmetrical. Therefore, to adopt a probabilistic approach, it suffices to start from a density function $f(\delta)$, where δ is the distance between the viewport center and the coordinates of the user’s gaze. In order to understand what the density function needed to look like, we used the results collected by Rai et al [8]. Figure 2 shows a histogram of the distance between the users’ gaze and the viewport center, over all videos and users examined in their work. A polynomial density distribution can be fit to the data taking into account a maximum distance of 55° or 0.96 radians. We assumed a viewport of 110° or 1.92 radians, similar to the VIVE head-mounted display³. Equation 1 presents the polynomial fit where the distance δ is expressed in radians.

$$f(\delta) = 187.6 \delta^6 - 576.4 \delta^5 + 625.2 \delta^4 - 249.0 \delta^3 - 6.8 \delta^2 + 19.4 \delta + 0.0006 \quad (1)$$

Next, the cumulative distribution $F(\delta)$ of the polynomial

³<https://www.vive.com>

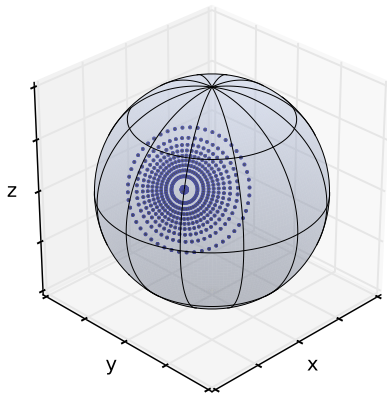


Figure 3: Quality sampling for $n_1 = 10, n_2 = 50$, taking into account the distribution of the eyes' location within the viewport.

density function can be determined. Given the density function's symmetry, $f(\delta, \alpha) = f(\delta)$ and the integral can be set between 0 and 0.96 radians (55°). In the equation, $Q(\delta, \alpha)$ defines the quality of the tile where the point on the sphere defined by the distance δ to the viewport center is located and, the offset α on the corresponding circle. Its value can be set in terms of a relative quality index (quality 1, ..., quality n), in terms of bitrates, or measured with objective metrics such as PSNR or SSIM applied to the different tiles.

$$\int_0^{0.96} \int_0^{2\pi} f(\delta, \alpha) Q(\delta, \alpha) d\delta d\alpha = \int_0^{0.96} \int_0^{2\pi} f(\delta) Q(\delta, \alpha) d\delta d\alpha \quad (2)$$

Since $Q(\delta, \alpha)$ is not continuous, this integral cannot be used in practice. Rather, we define a discrete variation by sampling uniformly on the cumulative distribution at n_1 values uniformly distributed between 0 and 1. For each of the resulting distance values, the corresponding circle of latitude is sampled n_2 times uniformly. Finally, the video quality Q_{vp} within the viewport is defined in Equation 3, where the quality result is weighted according to the distribution of the eye location within the viewport, resulting in a more accurate and fairer assessment. Figure 3 shows an example of the sampled analysis for $n_1 = 10$ and $n_2 = 50$.

$$Q_{vp} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} Q \left(F^{-1} \left(\frac{i}{n_1} \right), \frac{2\pi j}{n_2} \right) \quad (3)$$

IV. EVALUATION SETUP

In order to evaluate the strengths of our proposed model, we setup the adaptive VR video streaming testbed of Figure 4. In it, the VR video tiles and segments are streamed through an emulated network setup based on Mininet⁴, where the client is connected to an HTTP/1.1-enabled Jetty (Java) server⁵.

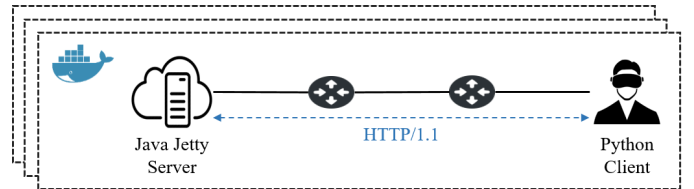
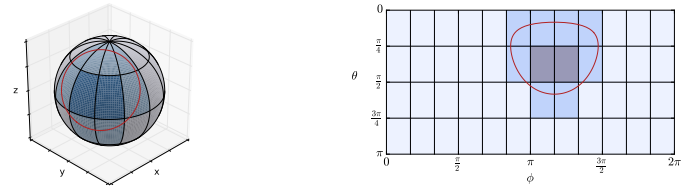


Figure 4: Experimental setup. MiniNet is used to host a virtual network within a Docker container. A Python-based VR player is used to play video streaming sessions from the HTTP/2-enabled Jetty server.



(a) Viewport extension on the VR (b) Equirectangular projection of the viewport.

Figure 5: Viewport definitions: From the 3D extension of the viewport (a) to its projection on the equirectangular plane.

The client is a headless Python-based adaptive streaming application (implementation available online⁶). The emulated network imposes different bandwidth and latency constraints. For this evaluation we have assumed latency zero and six different levels of available network bandwidth, ranging from 2.5 to 15Mb/s (2.5Mb/s, 5Mb/s, 7.5Mb/s, 10Mb/s, 12.5Mb/s and 15Mb/s).

As for the video content, we used the dataset provided by Wu et al. [15], which contains head movement traces for 48 unique users when watching 9 different VR videos. These traces specify the coordinates of the viewport center throughout all video sessions with an average sampling rate of 47 Hz. Out of the original 9, we picked 3 significantly different videos. While Sandwich presents an indoor performance with no extreme motion and complexity, Spotlight is an action movie where the action changes nearly in every sequence. Finally, Surf shows an ensemble of surf clips made using a GoPro camera. Each of the considered videos was encoded using the HM encoder⁷, applying a 4×4 tiling scheme at 4K resolution and 30 FPS. The Group of Pictures (GOP) length was set to 32, which resulted in a segment duration of around 1.067 s. Five quality levels were defined corresponding to CRF factors 15, 20, 25, 30 and 35, resulting in the average bitrates specified in Table I. In addition, in order to assess the degradation inflicted by the encoding process (by the HM encoder) on the different segments and tiles, we performed both SSIM and PSNR on a tile-based scheme using the maximum quality (CRF = 15) as the original content and each of the lower qualities as the impaired content. These quality indexes will further be used for the evaluation and benchmarking of our solution.

⁴<http://mininet.org>

⁵<https://www.eclipse.org/jetty/>

⁶<https://github.com/jvdrhoof/VRClient>

⁷<https://hevc.hhi.fraunhofer.de/HM-dol/>

Video	CRF	Relative Quality	Average SSIM [%]	Average PSNR[dB]	bitrate [Mb/s]
Sandwich	15	5	1	361.2	21.9 ± 6.6
	20	4	98.9	56.35	10.3 ± 3.2
	25	3	98.7	53.75	4.5 ± 1.4
	30	2	97.3	51.5	2.2 ± 0.7
	35	1	95.3	48.43	1.2 ± 0.3
Spotlight	15	5	1	361.2	20.8 ± 13.9
	20	4	93.2	45.01	10.6 ± 8.9
	25	3	92.8	43.03	5.2 ± 5.1
	30	2	92.3	41.02	2.6 ± 2.7
	35	1	91.3	38.9	1.4 ± 1.3
Surf	15	5	1	361.2	26.4 ± 12.7
	20	4	99.1	62.3	16.7 ± 8.7
	25	3	98.6	59.5	9.6 ± 5.4
	30	2	97.3	56.2	4.8 ± 2.8
	35	1	95.5	53.3	2.4 ± 1.4

Table I: Obtained bitrates for the three videos, presenting the relative quality index, the overall relative SSIM, the overall relative PSNR (both relative to the maximum quality CRF=15) and the average value and the standard deviation among all segments.

In addition to the tiling scheme and the segment duration, two optimizations define adaptive tile-based scheme for VR video streaming: the viewport prediction algorithm and the rate adaptation heuristic. For both of them, we chose two optimizations from our previous work [16]. On the one hand, the viewport algorithm estimates the future location by means of a three dimensional extrapolation. When a user has covered a certain distance in a given time interval, the current path is extended unidirectionally to predict the future user trajectory. This approach reflects the user’s movement within the video scene more naturally. Given the volatile behavior of the user, rather than predicting from the whole trajectory, the algorithm extrapolates the viewport prediction only based on the most recent samples. On the other hand, the rate adaptation heuristic decides the qualities of the tiles based on the according to their great arc distance to the center of the viewport. This approach results on a higher overall quality. A more detailed description of both optimizations is provided in [16].

One final aspect of our evaluation regards the definition of the viewport. Traditionally, the viewport has been defined as the equirectangular projection on the 2D video which will be streamed. For adaptive streaming purposes, this has been translated to the tiles allocated within a rectangle. However, this approach is not fully accurate as the transformation of a the viewport corresponds to a circular area on the sphere (Figure 5a). This area, when projected on a 2D plane corresponds to an area defined by the arc distance of the angles in the sphere (Figure 5b). Taking this notion in our work, we considered tiles to be within the viewport if the arc distance between their center and the center of the viewport (given by the HMD) is smaller than half of the viewport size (for a viewport of 110°, this means 55°). In addition to these inner tiles, we also considered the tiles on the edges as belonging to the viewport.

The complete setup is wrapped in a docker container, increasing portability and allowing parallel execution of video streaming sessions. Experiments were carried out on imec’s Virtual Wall⁸, with at most four docker containers running

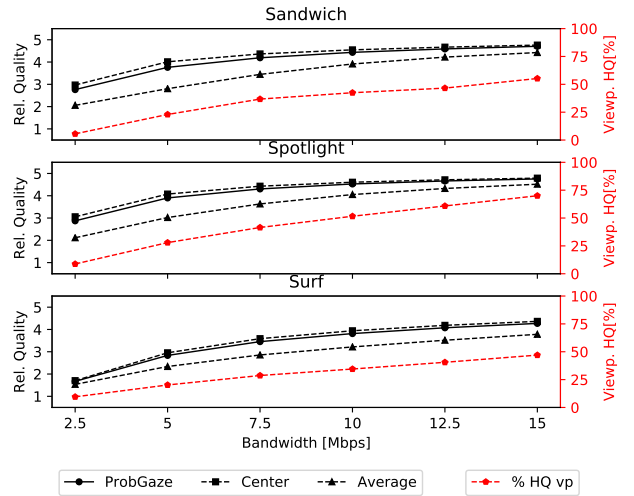


Figure 6: Relative quality assessment for each of the bandwidths and the three relative metrics: ProbGaze, Center and Average.

simultaneously on a hexacore Intel(R) Xeon(R) CPU E5645 @ 2.40GHz with 24 GB of RAM.

V. EVALUATION

This Section presents the evaluation of our metric. First, Section V-A analyses its performance to assess degradations derived from bandwidth constrains and quality misalignments within the viewport. Second, Section V-B explores in detail its user movement’s detection procedure. Finally, Section V-C evaluates its suitability as part of an objective full reference metric.

A. Effects of the Bandwidth Restrictions

First, we aimed to understand if our method were able to assess the quality degradations related to the quality changes typical from adaptive tile-based video streaming. Therefore, we took the obtained dataset (Section IV) and analyzed the quality by means of our metric, for each video, user, and generated segment. For simplification purposes, in the remainder of this Section, our method will be dubbed as ‘ProbGaze’. We benchmarked our solution against two other relative metrics, namely the center tile quality index (‘Center’) and the average quality index of all the tiles within the viewport (‘Average’). These two metrics correspond to the most optimistic and the most pessimistic assumption, respectively. On the one hand, ‘Center’ assumes that the user is only looking at the center tile of the viewport. Thus, any quality changes on adjacent tiles have no effect on its perception. On the other hand, on ‘Average’ the effect of each of the tiles within the viewport is the same, independent from its location.

We calculated the three relative metrics for all the segments belonging to each of the videos and users. Then, we averaged them to obtain one overall value per network bandwidth scenario and video type. Figure 6 presents the results of this first evaluation where each of the markers (circles for ProbGaze, squares for Center and triangles for Average) is one of these average quality index. To give a grasp of the different

⁸<https://doc.ilabt.imec.be/ilabt-documentation/>

trends, the Figure shows fitted curves to the three metrics for each of the videos. Finally, to provide further understandings on the different behaviors, Figure 6 includes on the right y-axes (red) the overall percentage of tiles within the viewport that were set to the highest quality (quality index 5).

As it can be seen, for each of the tested cases, our metric assessment is in-between the other two (the maximum and minimum), but always closer to the values provided by the Center tile approach, denoting the importance of the central tile on the quality assessment. The reason why our metric outputs, in general, a lower quality index comes from the probabilistic approach principle. While the Center approach ignores other tiles surrounding the viewport center, the probabilistic approach ponders and weights the effect of the adjacent tiles when providing a value. At lower bandwidths, the rate adaptation is required to decide on lower qualities on those adjacent tiles. This effect can be seen on the higher percentage of tiles downloaded at the highest quality. As the available bandwidth increases, our metric outputs values closer to the optimistic approach. These results encouraged us to further explore the capabilities of our approach.

B. Effects of the Users' Head Movements

One determinant aspect of the perception of adaptive VR video streaming is the reaction to the movement of the viewport. First of all, systems are required to react to and predict the new positions, and download the new required tiles in time before the user detects degradations. In addition, there is a need for quality metrics able to model the effect of the movement of the user on the experience perception. The purpose of this second analysis was to understand if it were possible for ProbGaze to detect the quality degradations derived from sudden user's movement within the spherical environment.

Thus, for this analysis, rather than looking for the average overall performance, our purpose was to pinpoint at the details and insights of the quality assessment. We evaluated the quality for each of the users during each of the simulated video sessions.

This Section presents the analysis of four 20-segment sections of the streaming sessions of two users while watching the video 'Spotlight' when the network is subjected to a bandwidth restriction of 7.5Mb/s. Out of the 48 users, the two presented were picked because they showed the most extreme behaviors in terms of average head movement speed. User 25 presented a speed of nearly 0.4rad/s, while User 26 barely reached 0.07rad/s.

Figure 7 shows the results of this analysis. After the initialization period, in which the rate adaptation heuristic only requests low qualities, all three algorithms stabilize their qualities to values close to the maximum (Figure 7a). From that moment on, it is possible to detect changes on the viewport (fast movements) just by following the drops on quality suffered by the Center algorithm, *i.e.* when the user moves, the center tile changes, and if that tile was not at the highest quality, a strong drop is detected. Examples of such changes can be found within the eight presented plots, more drastic in the case of User 25 (the fastest user). When those changes occur, the effect on Center is stronger than on the

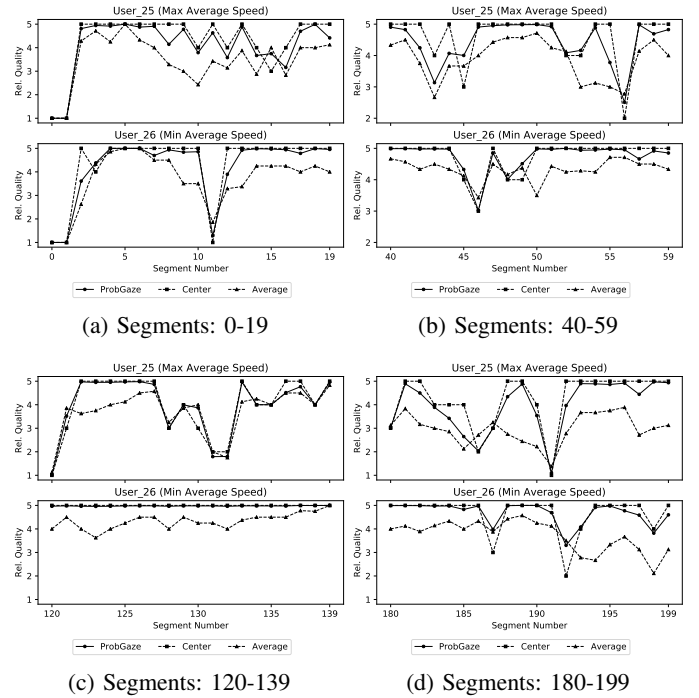


Figure 7: Relative quality measurements for the video Spotlight at 7.5Mb/s and the users that move the most (User 25) and the least (User 26).

other two metrics. As expected, the Average quality metric will smoothen the effect by the defined qualities in the newly defined viewport. ProbGaze, on the other hand, takes into account the weighted effect of the adjacent tiles, providing a more accurate measure of the quality degradation.

C. Towards an Objective Video Quality Metric for VR video streaming

For our final experiment, we wanted to explore the possibilities of integrating our probabilistic approach on state-of-the-art objective quality metrics. Therefore, we picked two well-known state-of-the-art objective frame-based video quality metrics, namely PSNR and SSIM and applied them to the video segments on a tile-by-tile basis, using the highest quality videos (CRF 15) as reference. Once the qualities were obtained, we took a similar approach as in the previous two evaluations and calculated three variants for each of the objective metrics. While ProbGaze-SSIM and ProbGaze-PSNR are the ProbGaze versions of the objective metrics, Center-SSIM, Center-PSNR provide the quality value of the center tile and Average-SSIM, Average-PSNR average the quality values within the tiles belonging to the viewport. As in the first analysis, each of the qualities were averaged per video and network bandwidth scenario for all users.

Figures 8 and 9 show the results of this analysis. For all the cases, our metrics' value is maintained just below the most optimistic approach (Center) but following a very similar and correlated trend.

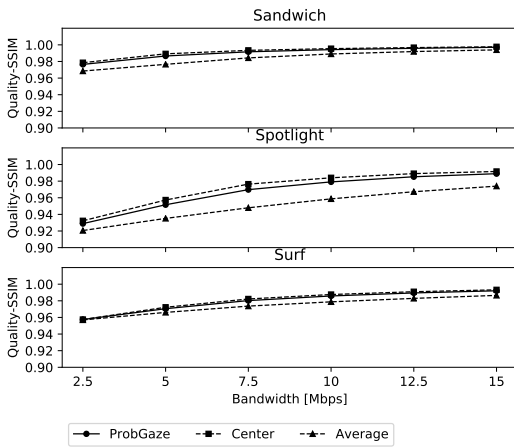


Figure 8: SSIM quality assessment for each of the bandwidths and the three SSIM variants: ProbGaze-SSIM, Center-SSIM and Average-SSIM.

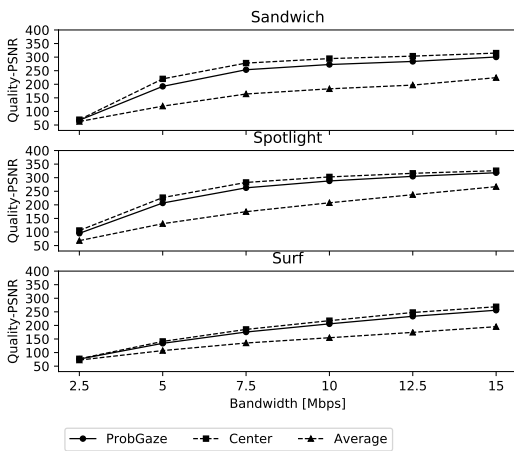


Figure 9: PSNR quality assessment for each of the bandwidths and the three PSNR variants: ProbGaze-PSNR, Center-PSNR and Average-PSNR.

VI. CONCLUSION

In this paper we have presented a novel video quality assessment for adaptive tile-based VR video streaming. By means of a probabilistic approach we model the influence of the eye movement within the viewport. While most of the current approaches provide a measure of quality by means of only the quality of the central tile or an average of all the tiles belonging to the viewport, our method considers the influence of each of the tiles by probabilistic weights to provide a more accurate measure of the perceived quality. This type of methodology opens a new venue of objective evaluation of the perceived quality of VR video services.

ACKNOWLEDGMENT

Jeroen van der Hooft is funded by the Agency for Innovation by Science and Technology in Flanders (VLAIO). Maria Torres Vega is funded by the Research Foundation Flanders (FWO), grant number 12W4819N. This research was performed partially within the project G025615N “Optimized

source coding for multiple terminals in self-organizing networks” from the fund for Scientific Research - Flanders (FWO-V).

REFERENCES

- [1] G. Dimopoulos, I. Leontiadis, P. Barlet-Ros, and K. Papagiannaki, “Measuring video qoe from encrypted traffic,” in *Proceedings of the 2016 Internet Measurement Conference*, ser. IMC '16. New York, NY, USA: ACM, 2016, pp. 513–526. [Online]. Available: <http://doi.acm.org/10.1145/2987443.2987459>
- [2] I. Sodagar, “The mpeg-dash standard for multimedia streaming over the internet,” *IEEE Multimedia*, vol. 18, no. 4, pp. 62–67, 2011.
- [3] 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.
- [4] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck, “An http2-based adaptive streaming framework for 360 virtual reality videos,” in *Proceedings of the 2017 ACM on Multimedia Conference*, ser. MM '17. New York, NY, USA: ACM, 2017, pp. 306–314. [Online]. Available: <http://doi.acm.org/10.1145/3123266.3123453>
- [5] O. A. Niamut, E. Thomas, L. D’Acunto, C. Concolato, F. Denoual, and S. Y. Lim, “Mpeg dash srd: Spatial relationship description,” in *Proceedings of the 7th International Conference on Multimedia Systems*, ser. MMSys '16. New York, NY, USA: ACM, 2016, pp. 5:1–5:8. [Online]. Available: <http://doi.acm.org/10.1145/2910017.2910606>
- [6] F. Qian, L. Ji, B. Han, and V. Gopalakrishnan, “Optimizing 360 video delivery over cellular networks,” in *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, ser. ATC '16. New York, NY, USA: ACM, 2016, pp. 1–6. [Online]. Available: <http://doi.acm.org/10.1145/2980055.2980056>
- [7] S. Petrangeli, J. van der Hooft, T. Wauters, and F. De Turck, “Quality of Experience-Centric Management of Adaptive Video Streaming Services: Status and Challenges,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, November 2017.
- [8] Y. Rai, P. Le Callet, and P. Guillotel, “Which Saliency Weighting for Omni Directional Image Quality Assessment?” in *Proceedings of the International Conference on Quality of Multimedia Experience*, 2017, pp. 1–6.
- [9] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, “A Survey on Quality of Experience of HTTP Adaptive Streaming,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.
- [10] 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 ACM Multimedia Systems Conference*, 2017, pp. 315–323.
- [11] Z. Xu, X. Zhang, K. Zhang, and Z. Guo, “Probabilistic Viewport Adaptive Streaming for 360-Degree Videos,” in *Proceedings of the IEEE International Symposium on Circuits and Systems*, 2018, pp. 1–5.
- [12] H. Ahmadi, O. Eltobgy, and M. Hefeeda, “Adaptive Multicast Streaming of Virtual Reality Content to Mobile Users,” in *Proceedings of the Thematic Workshops of ACM Multimedia*, 2017, pp. 170–178.
- [13] S. Petrangeli, V. Swaminathan, M. Hosseini, and F. De Turck, “An HTTP/2-Based Adaptive Streaming Framework for 360-Degree Virtual Reality Videos,” in *Proceedings of the ACM Multimedia Conference*, 2017, pp. 306–314.
- [14] R. I. T. da Costa Filho, M. C. Luizelli, M. Torres Vega, J. van der Hooft, S. Petrangeli, T. Wauters, F. De Turck, and L. P. Gaspary, “Predicting the Performance of Virtual Reality Video Streaming in Mobile Networks,” in *Proceedings of the ACM Multimedia Systems Conference*, 2018, pp. 270–283.
- [15] C. Wu, Z. Tan, Z. Wang, and S. Yang, “A Dataset for Exploring User Behaviors in VR Spherical Video Streaming,” in *Proceedings of the ACM Multimedia Systems Conference*, 2017, pp. 193–198.
- [16] J. van der Hooft, M. Torres Vega, S. Petrangeli, T. Wauters, and F. de Turck, “Optimizing Adaptive Tile-Based Virtual Reality Video Streaming,” in *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management 2019*, 2019.