https://doi.org/10.2352/ISSN.2470-1173.2019.10.IQSP-220

Are people pixel-peeping 360° videos?

Stephan Fremerey[a], Rachel Huang[b], Steve Göring[a], Alexander Raake[a]
[a] **Audiovisual Technology Group (AVT), Technische Universität Ilmenau, Ilmenau, Germany**
[b] **Huawei Technologies Co. Ltd., Shenzhen, China**
E-Mail: [a] [stephan.fremerey, steve.goering, alexander.raake]@tu-ilmenau.de, [b] rachel.huang@huawei.com

Abstract

In this paper, we compare the influence of a higher-resolution Head-Mounted Display (HMD) like HTC Vive Pro on 360° video QoE to that obtained with a lower-resolution HMD like HTC Vive. Furthermore, we evaluate the difference in perceived quality for entertainment-type 360° content in 4K/6K/8K resolutions at typical high-quality bitrates. In addition, we evaluate which video parts people are focusing on while watching omnidirectional videos. To this aim we conducted three subjective tests. We used HTC Vive in the first and HTC Vive Pro in the other two tests. The results from our tests are showing that the higher resolution of the Vive Pro seems to enable people to more easily judge the quality, shown by a minor deviation between the resulting quality ratings. Furthermore, we found no significant difference between the quality scores for the highest bitrate for 6K and 8K resolution. We also compared the viewing behavior for the same content viewed for the first time with the behavior when the same content is viewed again multiple times. The different representations of the contents were explored similarly, probably due to the fact that participants are finding and comparing specific parts of the 360° video suitable for rating the quality.

Introduction

360° video, also referred to as omnidirectional video (OV), immersive or panoramic video is a new type of media. Within the last few years, HMD devices became affordable for customers enabling them to consume such contents. To provide a good Quality of Experience (QoE) for such videos, a sufficiently high visual quality of the contents in terms of technical aspects like the resolution or framerate is important. Except of using Head-Mounted Displays (HMD), there are a few ways for displaying OVs to users.

For example 360° projection walls can be used [10]. Furthermore, Tang and Fakourfar [15] used tablet PCs to playout 360° videos. It is also possible to use a simple PC with a classical screen and a mouse for navigation within these contents [14].

For this study, we are only focusing on playing OVs using HMDs. As in the future, the resolution of these devices will be increasing, it is even more important to study the effect of a higher screen resolution on the perceived quality, the discrimination power of quality ratings, the meaningfulness of high resolution contents and the head rotation behavior of people watching OVs.

Within the presented study, we investigated how the higher panel resolution of the Vive Pro HMD influences the discrimination power of subjective ratings obtained in quality rating tests. For doing so, we conducted and compared a series of subjective tests using entertaining 360° contents with 20 s duration. Studies comparing the influence of a higher-resolution HMD like HTC Vive Pro on the QoE of 360° videos are currently not available.

In addition, we investigated whether a resolution of 8K really provides a better perceived quality than 6K. This comparison is necessary to get to know, if such high-resolution contents result in a quality improvement when shown on HMDs like the Vive Pro. From another study in our lab, we already know that for classical 2D video for some sequences there are no perceptual differences between UHD and HD content [5].

During all subjective tests, we additionally tracked the head-motion behavior of the participants of the study. The tracking data were analyzed to determine the preferred view ports in terms of respective heat maps and the behavior of people. Further, we investigated whether the exploration behavior is differing between the single quality levels. We found out that, there is no difference regarding user behavior across several quality levels for the same content, which is also important for 360° video production.

This paper is organized as follows. The following section briefly describes the state-of-the-art. In the next section, the setup used for conducting our tests will be presented. In the results section we will analyze the conducted tests in detail. The last section concludes and identifies some future work.

Related work

There are a series of studies investigating the topic of quality evaluation for OVs. The majority of studies conducted subjective tests using OVs with a resolution of 4K or smaller, while some of them also compared their results with the ones from objective quality metrics. There are also studies focusing on the quality evaluation for images. Within this section, the mentioned work will be explained more detailed.

For example, in [12], Schatz et al. examined the influence of stalling in the context of streaming 360° media and compared it with traditional TV media. For doing so, they conducted a subjective test using one omnidirectional and one traditional video with a duration of 60 s impaired by different stalling patterns. The subjects had to rate the Processed Video Sequences (PVS) using a 5-point Absolute Category Rating (ACR) scale (cf. [9]) and the presence in the virtual environment using a subset of questions from the Igroup Presence Questionnaire (IPQ) (cf. [11]). In the study [12], it is concluded that stalling has an impact on the QoE of OVs in a very similar way compared to traditional videos. In addition, they found out that subjects were more captivated while watching 360° media than for viewing traditional video material.

Furthermore, Xu et al. [18] analyzed the viewing behavior of subjects. They conducted a subjective test using the HTC Vive and recorded the head rotations of participants. In a first test, they presented 48 not distorted OV sequences having a duration ranging from 20 to 60 s and a resolution from 2880 × 1440 to 7680 × 3840 pixels to 40 subjects. From the evaluated head rotation data they conclude that the viewing directions of subjects are highly consistent across the different participants. Most of the subjects focused the center part of the video, whether for some videos other parts were more interesting. In the second test of this study [18], they presented 12 reference and 36 distorted OV sequences with a resolution of 4096 × 2048 and a duration of 12 s to 48 participants, while three different bitrate settings were used. They found out that the correlation between their two developed methods called O-DMOS (Difference Mean Opinion Score) and V-DMOS, whether O-DMOS refers to the overall and V-DMOS to the regional quality reduction within the 360° contents, is very high, i.e., a Spearman rank correlation coefficient of 0.96 was ascertained.

In another study of Zhang et al. [21], 50 impaired PVS obtained from 10 different 360° source videos were shown to 30 subjects. Test subjects rated the video quality using the 5-point ACR-HR (ACR-Hidden Reference) method. It was concluded that an optimal display resolution of the HMD is reducing unexpected quality changes evoked by the sampling of the device, making the assessment of these type of videos more reliable.

In [22], Zhou et al. presented 12 omnidirectional images to participants using different resolutions (4K, 2K, 1080p, 720p) with the HTC Vive HMD. The different processed images were presented in a randomized order whether subjects had to rate them using the Single Stimulus Continuous Quality Evaluation (SSCQE) method proposed in [8]. The authors found out that the developed parametric model is reflecting the perceptual quality of 360° images in a good way, i.e., a Pearson Correlation (PC) coefficient larger than 0.94 was achieved. The results show, that there is nearly no perceptual quality differences for resolutions higher than 2K. So it could be concluded that 4K does not provide a significantly higher subjective quality rating.

There are also objective metrics, like WS-PSNR (Weighted Spherical Peak Signal-to-Noise Ratio) and others that can be applied on 360° videos. For example, Tran et al. [17] compared results of objective quality metrics for 360° videos with results from a subjective test. They used objective metrics from the 360Lib tool (cf. Hanhart et al. [6]) that are based on the PSNR or variations of it adapted to OV. Within the test, 3 OVs of 30 s duration with 20 encoding settings were presented to 18 participants using Samsung's Galaxy S6 smartphone inserted in a Gear VR headset, while the resolution of the PVS was ranging from 1280 × 720 to 3840 × 1920. They are concluding, that the results of the objective quality metrics are well correlated with the results from the subjective quality test.

Other studies in the area of QoE evaluation for panoramic videos were done by Tran et al. [16], Zhang et al. [20] and Yang et al. [19], whether Yang et al. proposed a no reference quality assessment method for 360° videos based on using 3D convolutional neural networks.

To sum up, there are a lot of studies available for 360° video or image quality analysis, however none of the studies considers a quality evaluation for resolutions of OVs beyond 4K. We were not able to find studies comparing the QoE for OV shown on higher-resolution HMDs like the HTC Vive Pro or Samsung Odyssey[1]. Current HMDs like the HTC Vive Pro or Samsung Odyssey are able to playout higher resolution OVs, e.g., 4K, 6K and 8K. Furthermore it is not clear and not yet analyzed, if such higher resolutions have any perceptual benefit. Moreover, current literature is not considering to what extent a higher screen resolution can have an influence on user behavior during viewing OVs. In addition, we also found out that it is not covered whether different quality representations of the same 360° content have an impact on the exploration behavior or not. Within the presented study, we will cover all identified aspects.

[1] https://bit.ly/2A4idEO

Experimental Setup

In order to evaluate and compare the quality ratings and the exploration behavior for 360° videos, we conducted a series of three subjective tests. For evaluating the effect of the HMD's panel resolution on a) the perceived video quality and b) the user behavior, we conducted two tests with the same test setup using the HTC Vive and Vive Pro. In order to evaluate the quality of 360° videos on Vive Pro for 4K, 6K and 8K resolutions, we conducted another third test. In the following, the selection and preparation of test sequences, test software and equipment and the test method are described in detail.

In all tests, we pre-screened subjects using Ishihara and Snellen charts (20/25). During the tests, the head rotations of participants were recorded using a self-developed framework (cf. [4]), that is publicly available[2]. For all tests, after each PVS the participants were asked to rate the perceived video quality using the 5-point Absolute Category Rating (ACR) scale (cf. [9]). After 16 PVS, the subjects were asked to fill in a Simulator Sickness Questionnaire (SSQ), and afterwards each participant had a break for around 10 minutes. For enabling the continuous wearing of the HMD, the rating scale was displayed in the HMD. The participants mentioned their scores verbally, while they were noted down by the test supervisor.

Selection and Preparation of Test Sequences

In the first and second test, we used the same set of test sequences having a minimum resolution of 3840 × 1920 pixels and a framerate of 30 fps to minimize effects due to different SRC and Hypothetical Reference Circuit (HRC) settings. In the third test, a different set of SRCs having a minimum resolution of 7680 × 3840 pixels was used, partially also matching with SRCs used in the first test. Due to the lack of uncompressed OV SRCs, we chose to download exciting contents from ARTE [1] and YouTube, because we wanted to keep the subjects entertained during the tests. Within Table 1 all SRCs used are are listed with their IDs, names, starting timestamps for cutting the 20 s long sequences and their links. In Figures 1 and 2, Spatial SI/TI values per SRC were calculated for both tests using our publicly available SI/TI tool[3]. It is visible from the plots, that the SRCs reflect a broad range of spatial and temporal complexity. We decided to not display too many contents with high temporal information to minimize typical symptoms evoked by simulator sickness, as it is visible from one of our previous studies [13].

For preparing the SRCs for the subjective tests, as a first step,

[2] https://github.com/Telecommunication-Telemedia-Assessment/AVTrack360

[3] https://github.com/Telecommunication-Telemedia-Assessment/SITI

we decoded them using *ffmpeg*[4] 4.0 on Linux and *ffvhuff* as lossless video codec with a color depth of 8 bit, while *flac* was chosen as audio codec. Afterwards, they were encoded with *libx265* encoder using a typical 2-pass encoding setting. In Table 2, the respective used resolution and bitrate settings per HRC are given. For directly showing the 5-point ACR scale in the HMD and due to playout reasons of the used 360° player, the 2-pass encoded segments were decoded to lossless format (*ffvhuff*) and concatenated with a 5 s mid gray screen and a 10 s ACR scale. As a last step, encoding the concatenated videos using *libx265* and a *crf* (Constant Rate Factor) of 1 ensures that the hardware acceleration of the NVIDIA graphic card is used during playout. We compared the source video and the CRF 1 encoded versions using objective metrics, where we found out that there is nearly no difference. For ensuring a smooth playout of the audio, it was re-encoded at a fixed high bitrate (256k) using *aac*.

Figure 1. *SI/TI values of SRCs from test 1 and 2*

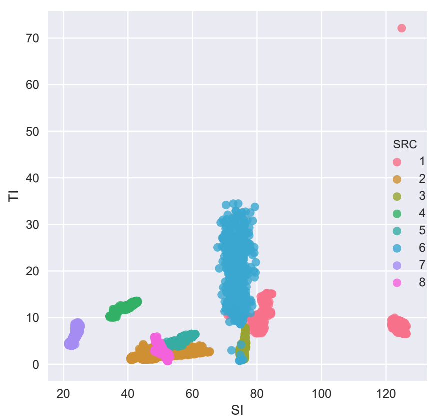
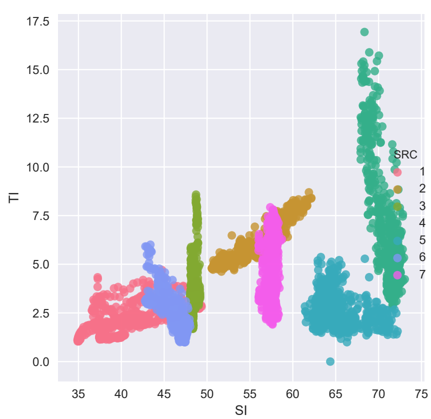


Figure 2. *SI/TI values of SRCs from test 3*



In the first two tests, 2 resolutions, 4 bitrates per resolution and 8 SRCs resulted in 64 different PVSs. Only 63 PVSs were presented to participants in the third test, because only 3 resolutions, 3 bitrates per resolution and 7 SRCs were considered, while partially other SRCs were used than in the first two tests.

Test Software and Equipment

Within [4], we developed AVTrack360, a software which is enabling an automated, randomized playback of 360° videos while the head rotation data are recorded in parallel. The head movements are captured using Pitch/Yaw and Roll data, whether pitch defines the up- and downwards movement of the head, yaw the side wards movements and roll tilting the head. Currently, the framework is supporting HTC Vive (Pro) and Oculus Rift. For playback, we used the Whirligig player[5] 4.2 because of its good audiovisual playback quality. For more information on the concept of the framework, we refer to the respective study [4].

In the conducted subjective tests, participants watched the video while sitting on a rotating chair. For ensuring a smooth playout of the audiovisual stimuli, the HMD was connected to a desktop computer running Windows 10 with a current Intel Core i7 processor, 32 GB RAM, one 1 TB M.2 SSD and a NVIDIA GTX 1080 graphics card.

Table 1: Used source contents in tests

ID	Name	Timestamp	Link
Test 1 and 2			
1	Concert	00:01:58	https://bit.ly/2GQjDWo
2	Fireworks	00:02:38	http://youtu.be/_J2e8HpT2To
3	Child	00:01:46	https://bit.ly/2qMLxLx
4	Etihad	00:01:41	N.a.
5	AngelFalls	00:00:42	http://youtu.be/8rUwdtERUOM
6	Dance	00:00:32	N.a.
7	Icebreaker	00:01:20	https://bit.ly/2G2BGHO
8	Street	00:00:35	N.a.
Test 3			
1	Fireworks	00:02:38	http://youtu.be/_J2e8HpT2To
2	AngelFalls	00:00:42	http://youtu.be/8rUwdtERUOM
3	Taipeh1	00:00:10	http://youtu.be/9hSwjGSWAa4
4	NewYork	00:01:37	http://youtu.be/2Lq86MKesG4
5	Hiking	00:01:00	http://youtu.be/vU4f9xBVo5o
6	Street	00:00:35	N.a.
7	Taipeh2	00:00:27	http://youtu.be/1waLjuPbj64

Table 2: Used conditions in tests

Condition	Resolution	Bitrate [Kbps]
Test 1 and 2		
Q1	1920 × 1080	500
Q2	1920 × 1080	1000
Q3	1920 × 1080	3500
Q4	1920 × 1080	7000
Q5	3840 × 2160	1000
Q6	3840 × 2160	2000
Q7	3840 × 2160	6000
Q8	3840 × 2160	12000
Test 3		
Q1	3840 × 1920	500
Q2	3840 × 1920	2000
Q3	3840 × 1920	6000
Q4	5760 × 2880	1000
Q5	5760 × 2880	4500
Q6	5760 × 2880	13500
Q7	7680 × 3840	2000
Q8	7680 × 3840	8000
Q9	7680 × 3840	24000

Test Method

To exclude any further influence factors, test 1 and 2 had the same test design. The only difference is the used HMD, whether we used HTC Vive in the first and HTC Vive Pro in the second and third test. At first, two subjects participating at a test were pre-screened for color vision and visual acuity using Ishihara and Snellen charts (20/20). After filling in a consent form, the test supervisor explained the test procedure to the participants. After-

[4]http://ffmpeg.org

[5]http://www.whirligig.xyz

wards, they had to fill in a SSQ. After mounting the HMD correctly on the head, in the initial session each subject had to watch 4 training contents first, representing the typical quality levels occurring in the real test. The SRCs used for training were different ones than used in the test. While watching the training contents in a randomized order, participants also had the possibility to adjust the Interpupillary Distance (IPD) of the HMD, which was initially set to 62mm for female and 65mm for male subjects (cf. Dodgson [3]). The 64 PVSs (63 in the third test) were divided into 4 sessions with 16 OVs per session, whether the playlists were randomized for each subject. After watching each PVS, the participant was asked to rate the perceived video quality, whether the score should be called verbally to the test supervisor, who noted down the score using AVTrack360. To avoid an influence of simulator sickness symptoms, potentially having an effect on quality ratings, after every session the participant had a break of 10 minutes. After each session, the participant filled another SSQ. With that procedure we were able to save time during the tests and conduct two tests at one time slot.

The participants, mostly students, were recruited from our university. In the first test, 27 subjects (14 female, 13 male, average/median age: 28/26), in the second test 28 participants (12 female, 16 male, average/median age: 26/24) and in the third test 27 subjects (13 female, 14 male, average/median age: 28/27) participated in the tests.

Results

In the following section, we will explain the results of subjective quality evaluation and analyze the head rotation data. At first, we will investigate to what extent the higher panel resolution of the HTC Vive Pro has an influence on the quality ratings. As a next step, we analyze whether higher resolutions than 6K are having any perceptual benefit. Afterwards, we will focus on evaluating the exploration behavior between a) the two HMDs and b) between the single quality levels watched by the subjects.

Video Quality Evaluation

Within this part, we want to a) evaluate how the HMD's screen resolution influences the quality ratings of participants and b) whether OVs with 8K resolution really provides a considerably better quality than 4K or 6K. To check the reliability of the participants, we performed an outlier detection on the given quality ratings of the subjects. We computed the Pearson Correlation Coefficient (PCC) between the scores of each user to the MOS of each condition, whether subjects were defined as outlier if the PCC is lower than 0.7. As a result, in the first test we had 6, in the second test 3 and in the third test 4 outliers, which were not considered in the following evaluation.

At first, it should be investigated to what extent the higher screen resolution of the Vive Pro is influencing the reliability of the quality ratings compared to the HTC Vive HMD used in the first test. It needs to be noticed that there were 50% fewer outliers in the second test compared to the first one. To get more insights, we computed the Standard deviation of Opinion Scores (SOS) analysis using the SOS hypothesis proposed in the respective study (cf. Hoßfeld, Schatz, and Egger [7]):

$$SOS(x)^2 = -ax^2 + 6ax - 5a \quad (1)$$

Figures 3 and 4 show the results for the first two tests, whether for test 1, $a \approx 0.246$ and for test 2 $a \approx 0.218$ was computed. Based on the computed a values and considering the plots it can be concluded that the agreement between the subjects on a specific rating per PVS is higher for the test using the Vive Pro compared to the Vive.

In Figure 5 the SOS analysis of the third test is shown. We calculated an a value of $a \approx 0.235$, whether the plot is indicating that the test is nearly as complex as the first test. This may be due to the fact that in this test 9 different conditions with 3 resolutions were used. Hence, it was more difficult for the participants to decide for a specific quality rating. From Figure 4 it can be concluded that using a higher-resolution HMD like Vive Pro helps subjects to more consistently rate the quality of 360° videos. This can be explained by the less visible screen-door effect (cf. Denison and D'Zmura [2]) of the Vive Pro, enabling the subjects to better distinguish between distortions evoked by the video and the screen of the HMD.

Figure 3. *SOS analysis test 1*

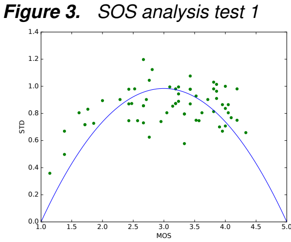


Figure 4. *SOS analysis test 2*

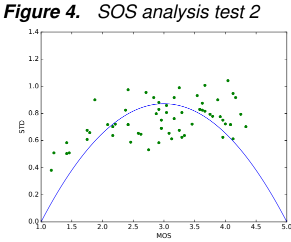
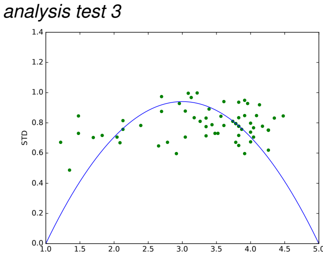


Figure 5. *SOS analysis test 3*

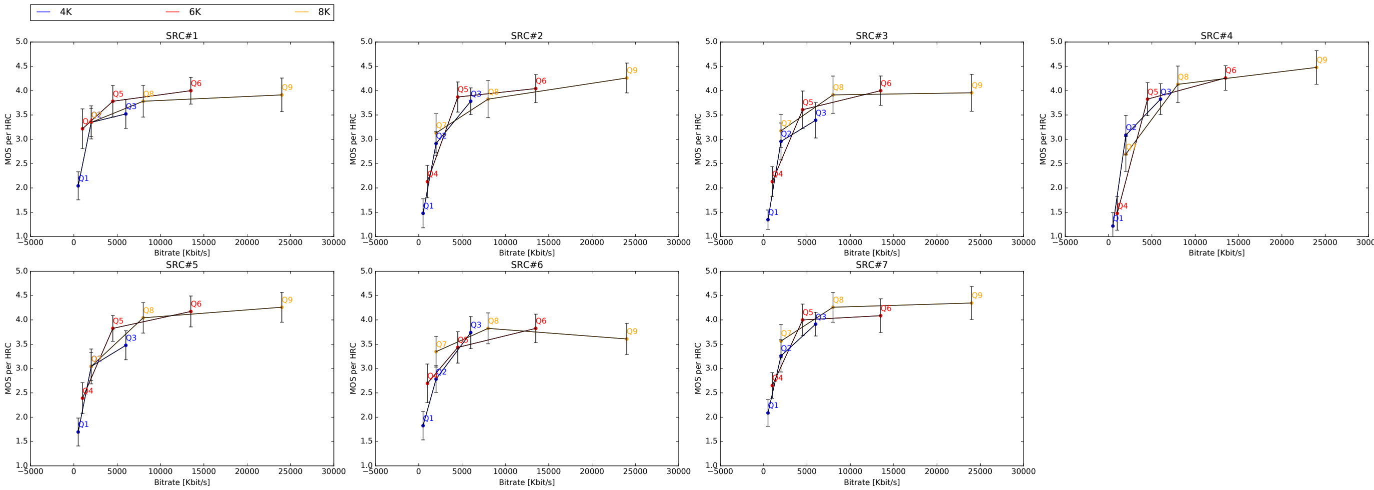


In the following, we want to investigate, if 8K provides a better quality compared to 6K or 4K at representative bitrates. In Figure 6, the MOS for all presented contents and conditions of test 3 is shown. In addition to the MOSs we computed 95% Confidence Intervals (CIs) according to the respective ITU recommendation (cf. [8]).

Considering the quality ratings, it can be concluded that for the highest bitrate of 6K and 8K resolution the perceived quality is almost the same. The MOS for the highest bitrate of 8K and 6K resolution mostly is between 4.0 and 4.5. For some SRCs (e.g. 1 and 6), 8K even provides a worse quality compared to 6K, but the differences are not significant. Nevertheless, 6K is always providing a considerably better quality than 4K resolution, mostly an ≈ 0.5 higher MOS can be observed for the highest bitrate, independent of the used content.

It could be assumed, that the HTC Vive Pro is not able to properly display the 8K content because of the limited resolution of the built-in panel. Most of the currently released HMDs (including HTC Vive and Vive Pro) are showing a Field of View

***Figure 6.** MOS for all contents in test 3*

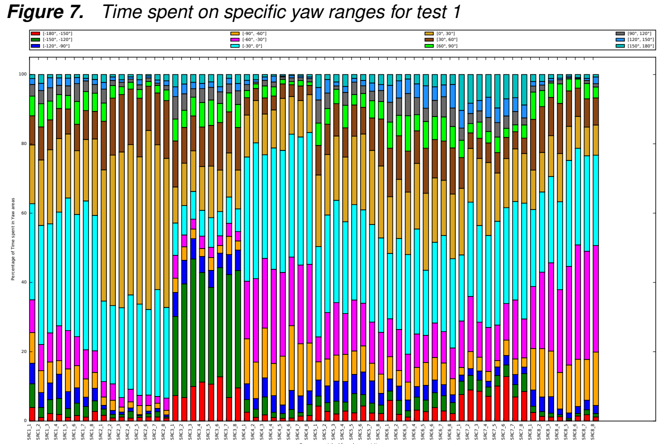


(FoV) of 110°. As from the literature, there are no clear statements on the vertical and horizontal FoV, we therefore assume a roughly spherical FoV. This corresponds to $\approx 30\%$ of the sideways 360° and $\approx 60\%$ of the upper and lower 180° shown in an OV. With a typical content resolution of 7680×3840, the percentage wise resolution of a specific part of the 360° video shown in the HMD, also referred to as viewport, would be around 2300×2300. The combined screen resolution of the HTC Vive is 2160×1200 pixels[6], whether the one of the HTC Vive Pro is 2880×1600 pixels[7]. Considering the total amount of pixels from the example 8K video ($\approx$ 5290000) and the HTC Vive Pro's screen (4608000), it could be assumed that an OV with 8K resolution could not be properly shown by the HTC Vive Pro's panel. At this point, it should be mentioned that this calculation is only a theoretical consideration and needs to be verified using respective measurement tools. Another reason why 8K seems to not have a considerable quality advantage compared to 6K could be that all of the contents in test 3 were pre-encoded by YouTube. Visually, we perceived no quality difference between the encoding with the highest bitrate and 8K (Q9, cf. Table 2) and the content from YouTube.

Behavioral Analysis

Beside the quality analysis, we are further interested in user behavior for 360° videos. Within this part, we want to analyze whether the exploration behavior was different between the HMDs and in-between the single PVS of the specific contents. Figures 7-9 show the respective yaw areas where the subjects spent time watching during the complete video. On the top of the shown plots we have the specific color-coded yaw ranges defined, whether we decided to use to take quantization steps of 30°, hence approximately $\frac{1}{4}$ of the HMD's FoV. On the y-axis, the percentage of time spent in the quantized yaw ranges is shown, whether on the x-axis the respective SRC watched for subsequent times in different quality levels is presented. Whether in tests 1 and 2, the same content was watched 8 times, referring to 8 quality levels, in test 3 the same contents were watched 9 times, referring to 9 quality levels.

***Figure 7.** Time spent on specific yaw ranges for test 1*



All graphs indicate that for subsequent times of watching the same contents, participants watched the different quality levels similarly. At this point we want to emphasize that the playlists were randomized for all subjects. It can be concluded that people seem to "pixel-peep" the video. This also can be ascertained by the "grouping" of different contents, visible from similar colored bars belonging to one content. By watching the same content several times, participants adapt to view particular parts of the 360° videos that are suitable for the task of rating the quality. This is e.g. visible in SRC 3 of test 1 and 2, shown in Figures 7-8, where $\approx 30\%$ of the time people were looking at an area of the video (green colored bars), where a river and a child was shown, suitable for the task of rating the quality. Figures 7-8 also show that the higher resolution of the Vive Pro doesn't have a considerable influence on the exploration behavior in comparison to the HTC Vive.

When comparing the areas focused over the time of watching a video, participants show similar behavior between the two devices. The reason for this could be because of nearly the same characteristics as Field Of View (FOV) and the same lenses used

[6] https://www.vive.com/us/product/vive-virtual-reality-system
[7] https://www.vive.com/us/product/vive-pro

in both HMDs. Another interesting observation is, that the areas focused by a different selection of subjects in test 1 and 2 are very similar, which is another evidence that people seem to "pixel-peep" the OVs.

Figure 8. *Time spent on specific yaw ranges for test 2*

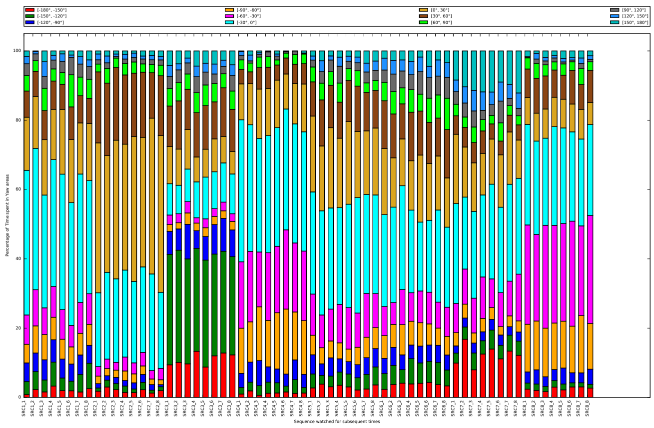
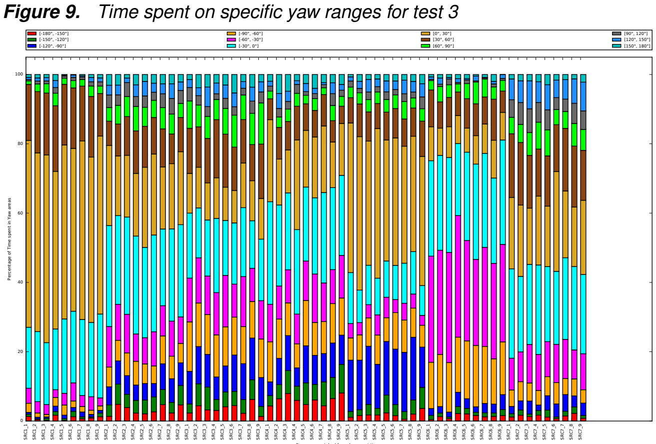


Figure 9. *Time spent on specific yaw ranges for test 3*



Conclusions

We analyzed state-of-the-art studies for QoE evaluation of 360° videos and found out that it currently is not investigated to what extent higher resolutions of the HMD's panel or the content have an impact on the perceived video quality. Furthermore, it is not clear, whether this has an influence on the exploration behavior.

For a detailed analysis regarding video quality perception and analyzing the head rotation behavior, we conducted three different subjective tests including 4K, 6K and 8K resolutions with representative bitrates using the HTC Vive and Vive Pro HMD. From our studies, we conclude that the higher display resolution of the HTC Vive Pro is helping participants to better evaluate the quality of 360° videos. In addition, we found out that there is nearly no perceivable difference between 6K and 8K resolution OVs using the HTC Vive Pro HMD.

While analyzing the head rotation of the participants, it can be concluded that people seem to "pixel-peep" 360° videos, focusing on the parts suitable for rating the video quality. Furthermore, from the results it became clear that the higher resolution of the HTC Vive Pro does not have a visible influence on the exploration behavior.

In the future, we want to conduct a study presenting high-resolution OVs using uncompressed SRCs. We also want to focus on establishing a link between user behavior, quality and simulator sickness.

References

[1] ARTE G.E.I.E. *ARTE360 VR*. 2018.

[2] M. S. Dennison and M. D'Zmura. "Cybersickness without the wobble: Experimental results speak against postural instability theory". In: *Applied ergonomics* 58 (2017), pp. 215–223.

[3] N. A. Dodgson. "Variation and extrema of human interpupillary distance". In: *Stereoscopic Displays and Virtual Reality Systems XI*. Vol. 5291. International Society for Optics and Photonics. 2004, pp. 36–47.

[4] S. Fremerey et al. "AVtrack360: an open dataset and software recording people's head rotations watching 360° videos on an HMD". In: *Proceedings of the 9th ACM Multimedia Systems Conference*. ACM. 2018, pp. 403–408.

[5] S. Göring et al. "Analyze And Predict the Perceptibility of UHD Video Contents". In: *Electronic Imaging, Human Vision Electronic Imaging* (2019). to appear.

[6] P. Hanhart et al. "360-Degree Video Quality Evaluation". In: *2018 Picture Coding Symposium (PCS)*. IEEE. 2018, pp. 328–332.

[7] T. Hoßfeld, R. Schatz, and S. Egger. "SOS: The MOS is not enough!" In: *QoMEX*. 2011, pp. 131–136.

[8] International Telecommunication Union. "ITU-R BT.500-13". In: *Methodology for the subjective assessment of the quality of television pictures* (2012).

[9] International Telecommunication Union. "ITU-T P. 910". In: *Subjective video quality assessment methods for multimedia applications* (2008).

[10] M. McGinity et al. "AVIE: a versatile multi-user stereo 360 interactive VR theatre". In: *Proceedings of the 2007 workshop on Emerging displays technologies: images and beyond: the future of displays and interacton*. ACM. 2007, p. 2.

[11] H. Regenbrecht and T. Schubert. "Real and illusory interactions enhance presence in virtual environments". In: *Presence: Teleoperators & Virtual Environments* 11.4 (2002), pp. 425–434.

[12] R. Schatz et al. "Towards subjective quality of experience assessment for omnidirectional video streaming". In: *Proc. 9th Int. Conf. Qual. Multimedia Exp.(QoMEX)*. 2017, pp. 1–6.

[13] A. Singla et al. "Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays". In: *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on*. IEEE. 2017, pp. 1–6.

[14] Y.-C. Su, D. Jayaraman, and K. Grauman. "Pano2Vid: Automatic Cinematography for Watching 360° Videos". In: *Asian Conference on Computer Vision*. Springer. 2016, pp. 154–171.

[15] A. Tang and O. Fakourfar. "Watching 360 videos together". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. 2017, pp. 4501–4506.

[16] H. T. Tran et al. "A subjective study on QoE of 360 video for VR communication". In: *Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on*. IEEE. 2017, pp. 1–6.

[17] H. T. Tran et al. "An evaluation of quality metrics for 360 videos". In: *Ubiquitous and Future Networks (ICUFN), 2017 Ninth International Conference on*. IEEE. 2017, pp. 7–11.

[18] M. Xu et al. "A subjective visual quality assessment method of panoramic videos". In: *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 2017, pp. 517–522.

[19] J. Yang et al. "3D panoramic virtual reality video quality assessment based on 3D convolutional neural networks". In: *IEEE Access* 6 (2018), pp. 38669–38682.

[20] B. Zhang et al. "Subjective and objective quality assessment of panoramic videos in virtual reality environments". In: *Multimedia & Expo Workshops (ICMEW), 2017 IEEE International Conference on*. IEEE. 2017, pp. 163–168.

[21] Y. Zhang et al. "Subjective Panoramic Video Quality Assessment Database for Coding Applications". In: *IEEE Transactions on Broadcasting* (2018).

[22] R. Zhou et al. "Modeling the impact of spatial resolutions on perceptual quality of immersive image/video". In: *3D Imaging (IC3D), 2016 International Conference on*. IEEE. 2016, pp. 1–6.