Gaze Data for Quality Assessment of Foveated Video

Oliver Wiedemann
oliver.wiedemann@uni-konstanz.de
University of Konstanz
Germany

Dietmar Saupe
dietmar.saupe@uni-konstanz.de
University of Konstanz
Germany

Figure 1: Foveation centered around the cheetah's head.

ABSTRACT

This paper presents current methodologies and challenges in the context of subjective quality assessment with a focus on adaptively encoded video streams.

CCS CONCEPTS

• **Computing methodologies → Image compression**.

KEYWORDS

adaptive video compression, quality assessment, eye tracking

ACM Reference Format:
Oliver Wiedemann and Dietmar Saupe. 2020. Gaze Data for Quality Assessment of Foveated Video. In *Symposium on Eye Tracking Research and Applications (ETRA '20Adjunct), June 2–5, 2020, Stuttgart, Germany*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3379157.3391656

1 INTRODUCTION

The demand for real-time video streaming, increasing display resolutions, and the emergence of consumer-grade eye tracking open up a branch of research on adaptive video compression that's not only of interest in the lab, but expected to be practically relevant in the foreseeable future. Enhancing foveated codecs thematically broaches subfields of both computer science and psychology.

ETRA '20 Adjunct, June 2–5, 2020, Stuttgart, Germany

ACM ISBN 978-1-4503-7135-3/20/06…$15.00
https://doi.org/10.1145/3379157.3391656

2 QUALITY IN MULTIMEDIA

Perceived visual quality is traditionally equated with signal fidelity metrics such as the PSNR, even though the observer and critic of a medium commonly is a human. The intricacies of the visual and cognitive systems lead to insufficient correlations of the PSNR with ground truth opinion scores. Quality metrics considering these aspects [Lin and Kuo 2011] and the creation of data sets through subjective studies constitute an active field of research.

3 ROI CODING AND FOVEATION

A primary consideration in compression is balancing the bitrate with visual quality. Region-of-interest (ROI) coding allows to spatially steer this trade-off. This can reduce the overall bitrate while retaining a constant level of perceptual quality [Hosu et al. 2016; Wang and Bovik 2006].

Static approaches use eye-tracking data, or a suitable approximation thereof, for video pre-encoding. They are limited, having to consider unknown and unexpected gaze paths. However, for real time applications, it is feasible to perform this adaptation live, according to the user's gaze that is captured by eye tracking and communicated back to the video encoder.

Foveation is a special case of ROI coding named after the retina's most sensitive centre, though the nomenclature is somewhat ambiguous. The prototypical application scenario consists of a unique observer, who expects a high quality video stream under strict latency and bandwidth constraints.

Existing video content can be pre-encoded in a high and a reduced quality version, which are then spatially cropped, streamed and reassembled, as presented in [Arndt and Antons 2016]. Our work focuses on the alternative scenario, which also includes the necessity to encode raw video in real time. A prime example is cloud gaming, for which the performance of a foveated streaming framework was evaluated in [Illahi et al. 2020].

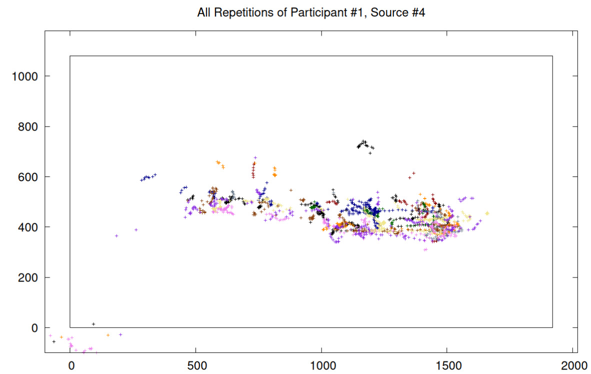

Figure 2: Expected behaviour.

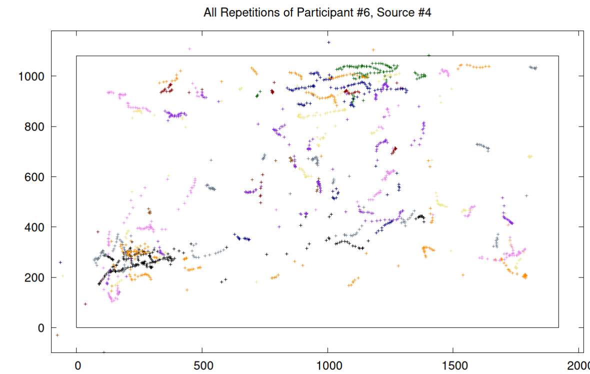

Figure 3: Explorative participant, eager to spot distortions.

At last, we want to point out a parallel development in the VR community, where foveated rendering is gaining traction to reduce the immense computational requirements.

4 APPLICATIONS OF GAZE DATA

The fixation data used in this section is a byproduct of the study presented in [Wiedemann et al. 2020]. In short, participants were presented a foveated video that was encoded according to their gaze. The bitrate of the periphery was decreased over time. Participants were asked to report distortions by pressing a button as soon as they became visible with the goal of identifying the point of just noticeable distortion for each source video.

For the above research eye tracking data has only been used in the encoder. Here, we propose to also make use of these data for the following natural applications.

User Filtering. Figure 2 depicts a gaze pattern that is consistent with the presumable region of interest of the back and forth moving cheetah, shown in Figure 1. Another participant's gaze for the same source video is displayed in Figure 3. Here, larger saccades occur much more frequently, and they often terminate in background regions, leading to fixation groups in arguably uninteresting regions. We attribute this to an ambitious participant, who is eagerly trying to spot distortions in the outer regions, analogous to the behaviour presented in [Alers et al. 2011]. Depending on the assessment task, this may be undesirable, when a more *natural* user behaviour is expected. To our knowledge, eye tracking has not been used for participant post-filtering in image or video quality databases.

Prediction of Just Noticeable Distortions. In foveated coding, one can make assumptions about the location of visible distortions, as the quality decreases radially around the current fixation point. Gaze paths in between runs of the same participant as well as between participants were surprisingly different and did not show significant, yet simple patterns, that allowed to establish a connection between gaze and distortion visibility. The size of our data collection is currently a limiting factor that prevents the application of, e.g., machine learning based analysis methods.

Towards Eye Tracking in Crowdsourcing Studies. In quality assessment, studies are often carried out by crowd sourcing in order to gather numerous opinions on a large number of media items. In this context, the dependence on lab-studies is a downside of eye tracking. However, as foveated coding arguably does not require pixel-level precision, it seems promising to carry out further testing through approximative eye tracking, e.g., on the basis of [Xu et al. 2015]. This is directly applicable, e.g., in the scenario of webcam based video telephony and entails all the practically relevant obstacles, which are easily overlooked in lab experiments.

REFERENCES

Hani Alers, Lennart Bos, and Ingrid Heynderickx. 2011. How the task of evaluating image quality influences viewing behavior. In *2011 Third International Workshop on Quality of Multimedia Experience*. IEEE, 167–172.

Sebastian Arndt and Jan-Niklas Antons. 2016. Enhancing video streaming using real-time gaze tracking. In *in Proc. ISCA/DEGA Workshop on Perceptual Quality of Systems*.

Vlad Hosu, Franz Hahn, Oliver Wiedemann, Sung-Hwan Jung, and Dietmar Saupe. 2016. Saliency-driven image coding improves overall perceived JPEG quality. In *2016 Picture Coding Symposium (PCS)*. IEEE, 1–5.

Gazi Karam Illahi, Thomas Van Gemert, Matti Siekkinen, Enrico Masala, Antti Oulasvirta, and Antti Ylä-Jääski. 2020. Cloud Gaming with Foveated Video Encoding. *ACM Trans. Multimedia Comput. Commun. Appl.* 16, 1, Article Article 7 (Feb. 2020), 24 pages. https://doi.org/10.1145/3369110

Weisi Lin and C-C Jay Kuo. 2011. Perceptual visual quality metrics: A survey. *Journal of visual communication and image representation* 22, 4 (2011), 297–312.

Zhou Wang and Alan C Bovik. 2006. Foveated image and video coding. *Digital Video, Image Quality and Perceptual Coding* (2006), 431–457.

Oliver Wiedemann, Vlad Hosu, Hanhe Lin, and Dietmar Saupe. 2020. Foveated Video Coding for Real-Time Streaming Applications
Accepted at the *12th International Conference on Quality of Multimedia Experience (QoMEX), IEEE.*

Pingmei Xu, Krista A Ehinger, Yinda Zhang, Adam Finkelstein, Sanjeev R Kulkarni, and Jianxiong Xiao. 2015. Turkergaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv preprint arXiv:1504.06755* (2015).

ACKNOWLEDGMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 251654672 – TRR161 (Project A05).