Beyond Foveal Rendering: Smart Eye-tracking Enabled Networking (SEEN)

Konrad Tollmar, Pietro Lungaro, Alfredo Fanghella Valero, Ashutosh Mittal
KTH - The Royal Institute of Technology
Mobile Service Lab, Department of Communication Systems
SE-16440 Kista, Sweden
konrad|pietro|ajfv|amittal@kth.se

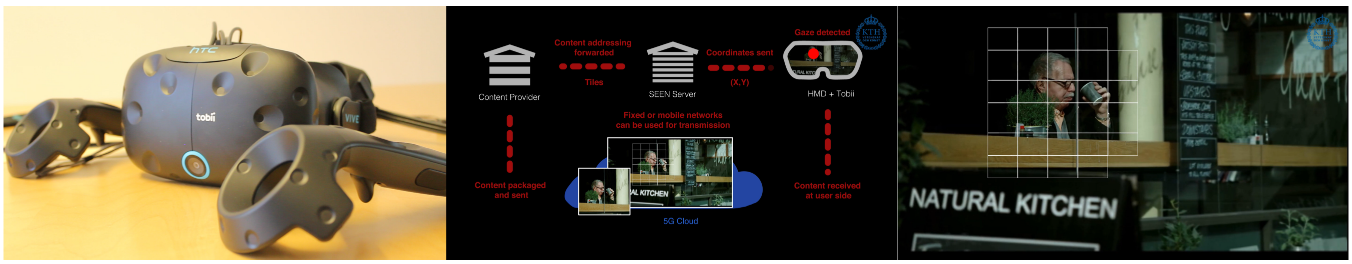

Figure 1: From left to right, a) the new developer version of HTC Vive with integrated Tobii eye-tracking, b) signaling and content packaging and delivery between the various systems units c) visualization of the tile based approach to implement foveated streaming.

ABSTRACT

Smart Eye-tracking Enabled Networking (SEEN) is a novel end-to-end framework using real-time eye-gaze information beyond state-of-the-art solutions. Our approach can effectively combine the computational savings of foveal rendering with the bandwidth savings required to enable future mobile VR content provision.

CCS CONCEPTS

• **Hardware** → **Emerging interfaces**; • **Networks** → *Network protocols*; • **Computing methodologies** → *Rendering*;

KEYWORDS

Eye-tracking, Mobile systems, Foveated Content delivery, Foveal rendering, User-experience, QoE

ACM Reference format:
Konrad Tollmar, Pietro Lungaro, Alfredo Fanghella Valero, Ashutosh Mittal. 2017. Beyond Foveal Rendering: Smart Eye-tracking Enabled Networking (SEEN). In *Proceedings of SIGGRAPH '17 Talks, Los Angeles, CA, USA, July 30 - August 03, 2017,* 2 pages.
https://doi.org/10.1145/3084363.3085163

1 INTRODUCTION

While foveal rendering [Boluda et al. 1996] can optimize computational resources at the client side [Guenter et al. 2012], it requires that all content is already locally available. The real challenge for future real-time systems and applications is to effectively transport massive amounts of information from where the content is created or stored on end-user devices. Thus, the basic philosophy of SEEN is that if bits are not going to be rendered on users screen, they should not be transported either. To make this possible one of SEENs cornerstones is the definition of a novel feedback loop between user devices and content caches, implemented via connected eye trackers. Our solution effectively combines eye-tracking with novel streaming protocols for 5G networks.

Initial investigations have shown that SEEN could at least cut in half the amount of bandwidth currently required by viewport-based streaming while delivering the same Quality of Experience (QoE) to the end users [Lungaro and Tollmar 2016]. These dramatic bandwidth savings, coupled with the improved carrier capabilities of 5G, are the prerequisite to enable future mobile VR.

The development and validation of the SEEN technology require a multidisciplinary approach, and it is the result of joint RD efforts in cooperation with our industrial partners at Tobii and Ericsson. Our research work explores the design space, constraints and trade-offs of the end-to-end foveated content delivery system. Further, ongoing development activities include the definition of novel network components that have the potential to revolutionize current services and enable new exciting and immersive interaction modalities. One of these is the Foveal Cloud Server, a network component designed to collect, store and process eye-gaze data from the different users. Information from the foveal server makes it possible to build advanced user profiles and content consumption models, to support new modalities for content personalization and interactivity. Further, understanding how different users visually

SIGGRAPH '17 Talks, July 30 - August 03, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5008-2/17/07.
https://doi.org/10.1145/3084363.3085163

consume specific multimedia objects will also enable novel storytelling options and content optimization opportunities for content providers.

2 RELATED WORK

The current state-of-the-art in content delivery optimization for VR is represented by viewport-aware methods [Corbillon et al. 2016; Kuzyakov and Pio 2016]. These use viewing angle information from the HMD to reduce the quality of 360 videos in the areas outside the users view. While this approach has already shown bandwidth cost savings of about 70% as compared to standard 360 streaming [El-Ganainy and Hefeeda 2016], only a limited portion of the content within a view is actually consumed by the users. The basic idea for SEEN is to complement viewport based methods by adding foveated content provision in each view.

Related work has also shown clear benefits in QoE-centric networks, i.e. optimizing and pre-caching mobile video-streaming based on QoE. However, it is not trivial to measure QoE and to map QoE metrics onto network Key Performance Indicators (KPIs). Nevertheless, using QoE metrics for optimization in a complex parameter space has now a particularly strong focus in the definition of new mobile networks, such as 5G.

3 SEEN ARCHITECTURE AND TESTBED

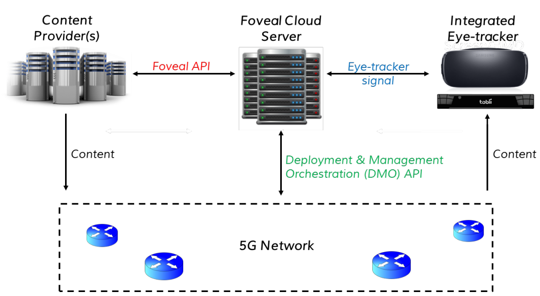

Figure 2: Network architecture for supporting Foveated content provision and interfaces (APIs).

Part of the RD activities for the development of SEEN have focused on the development of a prototype for an eye-tracker integrated with an HMD. The latter is connected to a computer which is also used to transfer the real time eye-gaze information to the foveal server. This is in charge of coordinating the foveated content provision activities. In particular, it can also store eye-gaze data locally for building user profiles and performing eye-gaze predictions. These can be used to predict upcoming saccadic movements and to identify likely Points of Interest (POIs) for future fixations. The identification of personalized POIs is invaluable to reduce the time to render high quality information after saccadic movements, since the system can speculatively activate several areas of a frame with high quality before knowing the coordinates of upcoming fixations.

The foveal cloud server can also monitor network conditions in real time, and when severe changes in performances are detected, it can react by adapting key foveal parameters. For example, it can lower the quality in the foveal region or increase the quality of the background, or increase or decrease the size of the foveal region. This flexibility enables a broad range of services and applications. Two main APIs have been designed and implemented for SEEN, and these are illustrated in Figure 2. The Foveal API is an interface between the Foveal Server and Cloud components and content providers and eye-trackers enabled VR headsets. This API supports optimizing foveated service provision for specific content types and wanted experiences. Instead the Deployment and Management Orchestration (DMO) API is defined at the interface between the Foveal Cloud and the network platform to enable services and adapt them to different network conditions.

4 FOVEATED STREAMING

To implement foveated content provision for 360 videos streaming, a video is spatially divided into a series of tiles, and each tile encoded into different qualities. The High-Efficiency Video Coding (HEVC) standard introduced the possibility of tiling specific videos. This allows splitting content into independent regions, with a bitstream that can be processed with a single decoder. The idea is to send in high quality only the tiles in the users RoI. In our approach, the low-quality background can be sent as a full video on a separate stream. However, it can also be implemented by setting the low resolution on the tiles outside the high-quality region. The data from the eye-tracker is used to identify which tiles are in the users field of view and to retrieve them in high resolution. In particular, the number of tiles selected around the gaze location depends on system parameters that can be tuned in function of specific user profile/settings and information extracted by processing the video content itself. The quality levels selected for both the foveal region and the background can also be adapted to varying network conditions, in a similar manner to standard adaptive streaming implementations. The information concerning communication performances can be provided by the network utilizing the DMO API

5 ACKNOWLEDGMENTS

This work is supported by Vinnova 2016-01854 and done by KTH, Tobii Technology and Ericsson Research.

REFERENCES

J. A. Boluda, F. Pardo, T. Kayser, J. J. Perez, and J. Pelechano. 1996. A new foveated space-variant camera for robotic applications. In *Proceedings of Third International Conference on Electronics, Circuits, and Systems* (1996-10), Vol. 2. 680–683 vol.2. https://doi.org/10.1109/ICECS.1996.584453

Xavier Corbillon, Alisa Devlic, Gwendal Simon, and Jacob Chakareski. 2016. Viewport-Adaptive Navigable 360-Degree Video Delivery. (2016). arXiv:1609.08042

Tarek El-Ganainy and Mohamed Hefeeda. 2016. Streaming Virtual Reality Content. (2016). arXiv:1612.08350 http://arxiv.org/abs/1612.08350

Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D Graphics. 31, 6 (2012), 164:1–164:10. https://doi.org/10.1145/2366145.2366183

Evgeny Kuzyakov and David Pio. 2016. Next-generation video encoding techniques for 360 video and VR. (2016). https://code.facebook.com/posts/1126354007399553/next-generation-video-encoding-techniques-for-360-video-and-vr/

Pietro Lungaro and Konrad Tollmar. 2016. Eye-gaze Based Service Provision and QoE Optimization. In *PQS 2016 5th ISCA/DEGA Workshop on Perceptual Quality of Systems* (2016). 6–10. https://doi.org/10.21437/PQS.2016-2

Simon Parkin. 2016. Nvidia Eye-tracking technology could make virtual reality much clearer. (2016). https://www.technologyreview.com/s/601941/nvidias-eye-tracking-tech-could-revolutionize-virtual-reality/