Gazi Karam Illahi, Gazi; Siekkinen, Matti; Kämäräinen, Teemu; Ylä-Jääski, Antti

On the Interplay of Foveated Rendering and Video Encoding

Published in:
Proceedings - VRST 2020

DOI:
10.1145/3385956.3422126

Published: 01/11/2020

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
On the Interplay of Foveated Rendering and Video Encoding

Gazi Karam Illahi
Aalto University, Finland
gazi.illahi@aalto.fi

Matti Siekkinen
Aalto University, Finland
matti.siekkinen@aalto.fi

Teemu Kämäräinen
Aalto University, Finland
teemu.kamarainen@aalto.fi

Antti Ylä-Jääski
Aalto University, Finland
antti.yla-jaaski@aalto.fi

FR NVE
NR NVE
FR FVE
NR FVE

0
10
20
30
40
50
KB/Frame
Translation
Rotation
mean
Slow
Fast

FR NVENR NVE FR FVE NR FVE FR NVENR NVE FR FVE NR FVE

0
10
20
30
40
50 KB/Frame
Translation
Rotation
mean
Slow
Fast

Figure 1: Left to right: (1) An overview of a Foveally Rendered scene with three shading rates and (2) a detailed sample from a frame rendered with different shading rates and QP values. The shading rates used in (1) are 1 shading pass per 1 pixel (within green ellipse), per 2x2 pixels (between green and yellow ellipses), and per 4x4 samples (beyond yellow ellipse). In (2) the same shading rates are used from right to left and encoded with (top to bottom) QPs: 0, 28, 38. QP=0 corresponds to lossless encoding, QP=28 is a typical value used for encoding and QP=38 corresponds to the QP of the lowest quality region in our FVE scheme.

ABSTRACT
Humans have sharp central vision but low peripheral visual acuity. Prior work has taken advantage of this phenomenon in two ways: foveated rendering (FR) reduces the computational workload of rendering by producing lower visual quality for peripheral regions and foveated video encoding (FVE) reduces the bitrate of streamed video through heavier compression of peripheral regions. Remote rendering systems require both rendering and video encoding and the two techniques can be combined to reduce both computing and bandwidth consumption. We report early results from such a combination with remote VR rendering. The results highlight that FR causes large bitrate overhead when combined with normal video encoding but combining it with FVE can mitigate it.

Matti Siekkinen and Teemu Kämäräinen are also with University of Helsinki, Finland.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

VRST '20, November 1–4, 2020, Virtual Event, Canada
© 2020 Copyright held by the owner/author(s).
https://doi.org/10.1145/3385956.3422126

CCS CONCEPTS
• Computing methodologies → Virtual reality; Rendering; Non-photorealistic rendering; • Computer systems organization → Real-time system architecture; • Networks → Cloud computing.

KEYWORDS
virtual reality, foveated rendering, video encoding, cloud rendering

ACM Reference Format:

1 INTRODUCTION
Interactive applications aiming for high quality visual experience, such as AAA video games and Virtual Reality (VR) applications, require powerful graphics hardware for real time rendering. To avoid the need for local PC having a dedicated graphics card to render graphics for these applications, remote rendering systems for gaming [2, 7, 9, 22] and VR [3, 11, 13, 14] have emerged. They offload (most) rendering tasks from client device to a remote server and stream the rendered graphics in real time as encoded video to
We ran experiments with different combinations of the techniques: without reducing perceived quality. Samples with different shading variable rate shading [19]. This can be done by, e.g., varying the quantization allows setting different shading rates for different regions within a

VRST ‘20, November 1–4, 2020, Virtual Event, Canada Illahi, et al.

of Nvidia VRworks suite of APIs [17] and the Vive plugin [6]. It allows setting different shading rates for different regions within a frame and varying them across frames. It can improve performance without reducing perceived quality. Samples with different shading

rates are shown in Figure 1. We used three level shading where high quality (1 shading pass per pixel) region was fixed to a radius of 1/8th of frame width, the transition quality (1 pass per 4 pixels) region to a radius of 1/6th of frame width and the rest was set to low quality (1 pass per 16 pixels).

To capture and encode rendered frames into video, we use a modification of the 360 Capture SDK [4]. Frame rate was set to 60 fps. FVE in the form of foveated quantization is implemented by adjusting the underlying h.264 encoder’s quantization offset $QO$ for each macroblock of a video frame. The method is the same as in [8]. We set the standard deviation of the 2-D Gaussian used to calculate the $QO$ to 1/8th of the frame width, and the maximum possible $QO$ ($QO_{max}$) to 10. In all experiments, the encoding scheme used was Nvidia’s low latency preset and the rate control mode was constant $QP$ with $QP = 28$ as the baseline unless otherwise mentioned.

3 RESULTS

The top-right plot in Fig 1 shows the resulting frame sizes that directly reflect the bandwidth demand. Interestingly, FR with NVE produces up to 30% larger frames on average compared to NR and NVE. The reason is that, even though there is less total visual information in a frame rendered with FR, the dissimilarity between two successive frames rendered with FR increases compared to NR. This means that inter-frame compression is less effective and the size of the predicted frames increases. We obtain similar but more pronounced results with experiments where only gaze shifts within a static scene with FR nearly doubling the bandwidth demand compared to normal rendering when FVE is not applied. Applying FVE together with FR remedies the situation. However, in all experiments we observe that NR+FVE yields smaller frames than FR+FVE, which hints that it may be possible to design a scheme that produces even smaller frames than the two independent methods combined.

The bottom-right plot in Fig 1 shows how structural similarity index measure (SSIM) and frame size behave with different parameter settings. The results were calculated using a single static scene rendered using NR, FR and encoded using FVE, NR with NVE being the reference. As expected, SSIM decreases with increasing $QO_{max}$. However, comparing the shapes of the SSIM and frame size curves reveals an interesting tradeoff between them: most of the savings in frame size can be obtained with small sacrifice in quality (e.g., using $QO_{max} = 10$). The results agree with those reported in [8] where the authors used a similar FVE scheme.

4 CONCLUSION

This paper reports early results from combining foveated rendering and video encoding in remote rendering systems. The results suggest that foveated rendering combined with normal video encoding may dramatically increase bandwidth consumption but applying foveation also in video encoding mitigates the problem. As future work, we plan to conduct user studies to better understand the impact of the two techniques on visual experience and to explore whether frame sizes can be further optimized with a tailored combination of FR and FVE.

ACKNOWLEDGMENTS

This work is supported by the Academy of Finland (grant 332306).
REFERENCES


